



INSTITUTIONEN FÖR PEDAGOGIK
OCH SPECIALPEDAGOGIK

Att bedöma språklig kompetens

Rapporter från projektet
Nationella prov i främmande språk



Redaktör: Gudrun Erickson

RIPS: Rapporter från Institutionen för pedagogik och
specialpedagogik, nr 16

© Hans Albinsson, Ulrika Arvidsson, Lena Asp, Patrik Axelson,
Ann Bjerklund Larsson, Jerker Brorsson, Lena Börjesson, Gudrun Erickson,
Ingela Finndahl, Lena Hedenbratt, Iani Hoff, Carina Jochens,
AnnaKarin Lindqvist, John Löwenadler, Sofia Nilsson, Eva Olsson,
Frank Perrotte, Håkan Schönberg, Åsa Sebestyén, 2018
ISBN 978-91-86857-24-0 (pdf)

<http://hdl.handle.net/2077/57683>

RIPS: Rapporter från Institutionen för pedagogik och specialpedagogik
RIPS är en digital skriftserie, startad 2011, som ges ut av institutionen för
pedagogik och specialpedagogik vid Göteborgs universitet och publiceras
som Open Access via GUPEA (Göteborgs universitets publikationer –
elektroniskt arkiv).

Redaktion: Girma Berhanu och Monica Reichenberg

Abstract

Title: Assessing Language Competence
Editor: Gudrun Erickson
Language: Swedish with an English summary
ISBN: 978-91-86857-24-0 (pdf)
Keywords: language competence, assessment, validity, national tests, collaboration

The current report is a collection of texts on the assessment of language competence at the national level in the Swedish educational system. More specifically, it deals with the development of materials for English, French, German and Spanish undertaken at the University of Gothenburg, Department of Education and Special Education, on behalf of the Swedish National Agency for Education. There is a long tradition of national assessment in Sweden, with varying aims over time due to political and policy-related decisions. However, there is also a strong tradition to trust teachers' judgments regarding their students' development of competences. Hence, since long, the national assessment materials are seen as a complement to teachers' continuous observations, albeit meant to provide substantial support for the assignment of grades, which is also the responsibility of teachers. Since grades are used, to a large extent, for admission to higher education, this is indeed a demanding task.

The report comprises eleven texts on various topics within the national language assessment context, written by members of the NAFS project (the Swedish acronym for 'national tests of foreign languages'). There are teacher educators as well as researchers in the group of test developers and authors of the different texts, all of them, however, with a common background as language teachers in compulsory school and/or at upper secondary level. Furthermore, there is a number of international contacts and networks surrounding the project, providing plenty of opportunities for inspiration and cooperation.

In the initial background to and summary of the volume, some short information is given on the history of national assessment in Sweden, including some of the crucial issues related to the aims of the system. In a separate chapter, basic principles for assessment in general, and language assessment in particular are discussed briefly, also seen in the light of the specific and multifaceted function of the national system (Erickson).

The section connected closely to the current assessment materials starts off with a chapter on the development of a diagnostic material to be used continuously from the start of English language instruction to the point when grades are awarded for the first time, in school year six (Bjerklund Larsson & Jochens). This is followed by texts related to various aspects of test development: one on the function, and use, of multiple choice items in tasks aimed at assessing receptive language competence (Schönberg & Lindqvist), and three describing and discussing the role of test-taker feedback in the development and validation of materials for French, emphasizing aspects of authenticity (Finndahl & Perrotte); Spanish, in particular focusing on attitudes expressed by male and female students, (Hedenbratt & Axelson), and French, German and Spanish, regarding tasks aimed to test reading comprehension (Sebestyén & Albinsson). Next, there is a chapter focusing on gender differences in English language proficiency as shown in the results for tests of English at the upper secondary level (Börjesson & Nilsson). Further, one text analyses the introduction and function of tasks aimed to test usage in an assessment material for the highest course of English in upper secondary school (Löwenadler & Hoff), and another, aspects of vocabulary in the assessment of English at different proficiency levels (Olsson). The second last chapter describes and discusses the handling and use of national tests at the local school level (Arvidsson, Asp & Brorsson). The volume concludes with a text looking forward to possible effects of the digitalization currently undertaken within the Swedish national assessment system (Lindqvist, Arvidsson & Nilsson).

Innehåll

ABSTRACT	3
INNEHÅLL.....	5
BAKGRUND OCH SAMMANFATTNING.....	7
BEDÖMNING AV SPRÅKLIG KOMPETENS	15
<i>GUDRUN ERICKSON</i>	
BEDÖMNING I ENGELSKA FÖR YNGRE ÅLDRAR.....	41
<i>ANN BJERKLUND LARSSON & CARINA JOCHENS</i>	
ÄR DET FARLIGT ATT KÖRA MC?	51
<i>HÅKAN SCHÖNBERG & ANNAKARIN LINDQVIST</i>	
TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK.....	67
WAS PIENSAN LES ÉLÈVES? – TEST-TAKER FEEDBACK I SAMBAND MED UTPRÖVNINGAR I TYSKA, SPANSKA OCH FRANSKA	69
<i>ÅSA SEBESTYÉN & HANS ALBINSSON</i>	
”DET VAR FAKTISKT KUL ATT PRATA OM VARDAGLIGA SAKER SOM MAN KAN PRATA OM I FRANKRIKE TYP”	79
<i>INGELA FINNDAHL & FRANK PERROTTE</i>	
MOTIVERAD FÖR UPPGIFTEN? – TEST-TAKER FEEDBACK UR ETT GENUSPERSPEKTIV	91
<i>LENA HEDENBRATT & PATRIK AXELSON</i>	
KÖNSSKILLNADER I ENGELSK SPRÅKFÄRDIGHET	107
<i>LENA BÖRJESSON & SOFLA NILSSON</i>	
USAGE I KURSPROVET FÖR ENGELSKA 7	119
<i>JOHN LÖWENADLER & LANI HOFF</i>	
VOKABULÄR OCH BEDÖMNING AV SKRIFTLIG FÖRMÅGA	139
<i>EVA OLSSON</i>	
PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA	167
<i>ULRIKA ARVIDSSON, LENA ASP, JERKER BRORSSON</i>	
SPRÅKPROV PÅ DATOR – VAD HÄNDER DÅ?.....	183
<i>ANNAKARIN LINDQVIST, ULRIKA ARVIDSSON, SOFLA NILSSON</i>	

Bakgrund och sammanfattning

Denna rapport handlar om bedömning av språklig kompetens, närmare bestämt om olika aspekter av de nationella material i engelska och moderna språk som tillhandahålls i den svenska skolan. Utvecklingen av materialen sker, på uppdrag av Skolverket, vid Institutionen för pedagogik och specialpedagogik på Göteborgs universitet inom projektet NAFS (Nationella prov i främmande språk).

Bakgrund: nationella prov

Bedömning på nationell nivå har funnits under lång tid i det svenska skolsystemet, fram till och med 1960-talet inte minst i form av examensprov för real- och studentexamen men även som centralt tillhandahållna prov med syfte att stödja lärares betygssättning. Under det normrelaterade, relativa betygssystemets tid innebar detta att standardprov för grundskolan och centrala prov för gymnasiet tillhandahölls. Syftet med dessa var att provresultaten skulle ge upplysning om klassers nivå i relation till riksgenomsnittet, men för enskilda elever skulle de inte betraktas som viktigare än andra provresultat. I början av detta system angavs fasta procenttal för respektive betygssteg, från vilket man som lärare inte fick avvika på något betydande sätt utan att motivera varför. Detta kom dock att frångås i grundskolan, där det under de senare åren enbart fastslogs att betygssteget 3 i normalfallet skulle vara det dominerande (Gustafsson & Yang Hansen, 2009; Marklund, 1987).

Införandet 1994 av nya läroplaner för grund- och gymnasieskolan, Lpo 94 och Lpf 94, medförde bland annat en övergång till ett mål- och kunskapsrelaterat betygssystem. Detta innebar att de nationella prov som tillhandahölls skulle ge ett mått på enskilda elevers kunskaper i relation till verbaliserade betygskriterier, inte som tidigare en indikation på grupperns nivå. Detta var givetvis en mycket stor förändring på flera plan, såväl individuellt som pedagogiskt och strukturellt, och innebar stora utmaningar för olika berörda kategorier av användare (se t.ex. Tholin, 2006). I samband med reformen utökades det nationella provsystemet kraftigt och kom att innefatta både fler typer av material och fler ämnen. Expansionen fortsatte och accentuerades i samband med införandet av nya läro- och

ATT BEDÖMA SPRÅKLIG KOMPETENS

kursplaner i 2011 års reform och i anslutning till detta tidigare betyg och en ny betygsskala.

Diskussioner om bedömning, prov och betyg är aktuella på olika plan och med olika aktörer och intressenter. Under de senaste tjugofem åren, d.v.s. sedan det målrelaterade systemet infördes i Sverige, har frågor om implementering respektive likvärdighet haft en framträdande roll och, relaterat till detta, definitionen av de nationella provens syfte, antalet syften och innehållet i dessa. Sammanfattningsvis kan konstateras att utvecklingen gått från fler till färre explicita syften och att provens roll för individuell rättvisa och likvärdighet successivt fått en allt större betydelse. Vidare har lärare sedan länge etablerade och mångfacetterade roll i bedömning och betygssättning fokuserats, dvs. att de kontinuerligt ska utvärdera sina egna elevers kunskapsmässiga utveckling, bedöma deras nationella prov samt ansvara för den sammanfattande, individuella betygssättningen.

Under det tidigare skedet av det målrelaterade systemet, från mitten av 1990-talet och ett drygt tiotal år framåt, hade de nationella proven explicit formativa och implementerande syften: proven skulle bidra aktivt till elevers lärande och medvetenhet om sitt lärande och lärare skulle genom materialen få hjälp att tolka, genomföra i undervisning samt utvärdera styrdokumentens kunskaps- och ämnessyn. Vidare skulle provresultaten medverka till likvärdighet och rättvisa samt dessutom ge möjlighet att studera kunskapsutveckling över tid. Flera studier, bland annat av Skolverket (2007; 2009) visade dock att proven och resultaten användes på ett påtagligt variabelt sätt, såväl vad gällde bedömningen av elevsvar som den rådgivande funktion de avsågs ha vad gäller elevernas slutbetyg. Detta, tillsammans med Skolinspektionens studier av bedömersamstämmighet (Gustafsson & Erickson, 2013; Skolinspektionen, 2012), medförde att provens formativa och implementerande syften successivt tonades ner och att rättvisa- och likvärdighetsaspekter kom att betonas alltmer. Detta låg också i linje med internationella iakttagelser och förslag vad gällde det nationella provsystemet (Gustafsson, Cliffordsson & Erickson, 2014; Nusche m.fl., 2011).

De nationella proven har sedan länge mycket god acceptans bland lärare vad gäller innehåll, utformning och upplevt stöd för betygssättning (Skolverket, 2014). Dock har i tilltagande grad frågor om arbetsbörda uppmärksamats, liksom oklarheter vad gäller provens stödjande roll för betygssättning och därav följande variabilitet, såväl inom som mellan ämnen (Erickson, 2017).

Till följd av ovan nämnda iakttagelser och diskussioner beslutade regeringen 2015 om en statlig utredning av de nationella proven med syfte att ”åstadkomma ett effektivare och tillförlitligare nationellt sammanhållet system för

bedömning, uppföljning och utvärdering av elevers kunskaper med minskad administrativ börda för lärare” (Regeringskansliet, Dir. 2015:36). Utredningen rapporterades 2016 och presenterade ett antal förslag om utvecklingen av det nationella provsystemet (SOU 2016:25). Det mest grundläggande av dessa avsåg en uppdelning av systemet i tre delar med en renodling av respektive syfte: nationella prov, nationellt bedömningsstöd och nationell kunskapsutvärdering. För de nationella proven innebar detta en reducering till ett föreslaget syfte, nämligen stöd för individuell betygssättning, samt en förtydligad relation mellan nationella provresultat och betyg. Bland de övriga förslagen fanns t.ex. ett minskat antal nationella prov, en översyn av tidsomfattning för proven, en stegvis digitalisering av nationella prov samt ökad likvärdighet i bedömningen.

En fråga som diskuterats av Skolverket sedan början av 2000-talet avser behovet av ett gemensamt ramverk för de olika ämnena i det nationella provsystemet. Vikten av ett sådant betonades även av OECD (Nusche m.fl., 2011) och resulterade i ett uppdrag att utveckla ett förslag till systemramverk, som skulle ligga till grund för de olika ämnenas specifika provramverk. Behovet av ett sådant nämndes även i SOU 2016:25. Det framtagna förslaget (Erickson, Gustafsson & Nyström, 2017) resulterade i en första version av ett systemramverk (Skolverket, 2017), som ska utgöra grundval för de olika ämnenas och provens konkretiserade och motiverande beskrivningar av utvecklingsarbetet för respektive material. Den bärande tanken i ramverket är enligt Skolverket (2017, s. 4) ”att säkerställa högsta möjliga kvalitet i de nationella proven, och högsta möjliga trovärdighet i användning och konsekvenser av resultat på nationella prov i förhållande till de syften som proven har”. Detta ska åstadkommas inte minst genom successiv kvalitetssäkring, med målet att öka såväl reliabilitet som stabilitet i systemet.

Bakgrund: nationella prov i engelska och bedömningsstöd i moderna språk

De nationella proven och bedömningsstöden utvecklas, på uppdrag av Skolverket, vid olika universitet och högskolor i landet. Beslutet att lägga denna verksamhet utanför den centrala skolmyndigheten togs tidigt på 1980-talet och motiverades av frågor om kompetens och legitimitet. Uppdragen innefattar i nuläget obligatoriska nationella prov i svenska/svenska som andraspråk, engelska och matematik samt i grundskolan även prov i natur- och samhällsorientering.

ATT BEDÖMA SPRÅKLIG KOMPETENS

rande ämnen (åk 9). Till denna kategori hör också proven i svenska för invandrare (Sfi), som avser vuxna elever. Vid sidan av dessa prov tillhandahålls i ett vidgat antal ämnen bedömningsstöd, som från den nationella nivån är ett erbjudande (se vidare Skolverkets hemsida), men som kan göras obligatoriska lokalt. Bland dessa återfinns de material i moderna språk, franska, spanska och tyska, som utvecklas parallellt med och enligt samma principer som de nationella proven i engelska.

Vad gäller proven i det som traditionellt, och till synes något besynnerligt i vår globaliserade värld, refereras till som 'främmande språk', dvs. de traditionella skolspråken utöver modersmål, svenska och svenska som andraspråk, ligger ansvaret sedan länge hos den språkpedagogiska forskningsmiljö vid Göteborgs universitet, Institutionen för pedagogik och specialpedagogik, där språklig kompetens förenas med ämnesdidaktisk kunskap och tillgång till expertis såväl i pedagogisk mätning och bedömning som specialpedagogik. Materialen utvecklas på basis av forskning och beprövad erfarenhet och enligt principer relaterade både till processer och produkter (www.nafs.gu.se). Dessa principer diskuteras i den första texten i denna rapport (Erickson), där även den kunskaps- och språksyn behandlas som präglar de svenska styrdokumenterna i språk och som följaktligen ligger till grund för utformningen av de nationella proven och bedömningsstöden. Provmaterialen diskuteras utifrån ett antal aspekter relaterade såväl till grundläggande teoretiska ställningstaganden och kurs- och ämnesplaner som till elevers och lärares uppfattningar och förslag, vilka fyller en central funktion i utvecklingen och den successiva kvalitetssäkringen av materialen.

Texterna i den här föreliggande rapporten bygger till allra största del på data insamlade och erfarenheter gjorda inom ramen för det reguljära utvecklingsarbetet. De är skrivna av medarbetare i det nationella provprojektet, samtliga i grunden lärare med gedigen erfarenhet av undervisning i grundskolan, gymnasiet och/eller vuxenutbildning. Vissa delar sitt arbete i provprojektet med undervisning i skolan, somliga är engagerade i lärarutbildning och några även i forskning inom andra projekt. De olika artiklarna ger inblickar och reflektioner kring proven, som förhoppningsvis kan vara intressanta och användbara både i fortbildnings- och utbildningsmässiga sammanhang. Nedan följer en kort presentation av de olika texterna i den ordning de förekommer i rapporten.

Texter i rapporten

Rapporten inleds med en text av övergripande och principiellt slag, där grundläggande begrepp inom bedömningsområdet diskuteras och relateras till arbetet med nationella prov i språk, från syfte till innehåll, val av utvecklingsmetoder och användning av resultat. Texten, som har titeln *Bedömning av språklig kompetens*, är skriven av Gudrun Erickson, professor i pedagogik med inriktning mot språk och bedömning, tidigare projektledare och numera vetenskaplig ledare för Nafs-projektet, som också är rapportens redaktör.

Den andra artikeln i rapporten har rubriken *Bedömning i engelska för yngre åldrar* och beskriver utvecklingen av ett diagnostiskt och formativt material i engelska avsett för de första årens engelska, fram till det nationella provet i årskurs 6. Såväl syften som process och produkt fokuseras, liksom elevers och lärares synpunkter på materialet. Texten är skriven av Ann Bjerklund Larsson och Carina Jochens, båda provutvecklare och erfarna lärare för yngre elever i engelska, den senare även ansvarig för det nationella provet i åk 6.

Texten *Är det farligt att köra MC?* handlar om uppgiftstyper och svarsformer i poängbaserade, receptiva uppgifter. Efter en bakgrund om flervalssystemets egenskaper, möjligheter och risker, presenteras en studie baserad på en utprovning av uppgifter för gymnasiets prov i kursen engelska 5, där en uppgift med tre respektive fyra svarsalternativ analyseras och diskuteras. Textens författare är Håkan Schönberg och AnnaKarin Lindqvist, båda provutvecklare, den senare även ansvarig för provet i engelska 5 samt projektledare för Nafs.

I tre separata texter om utvecklingen av bedömningsstöden i moderna språk beskrivs och diskuteras elevers bidrag under utprovningsskedet till de slutliga proven. Avsnittet inleds med ett kort, gemensamt avsnitt om bakgrunden till arbetet. Därefter följer en text med rubriken *Was piensan les élèves? – Test-taker feedback i samband med utprovningar i tyska, spanska och franska*, som rapporterar och diskuterar en jämförande studie av elevenkäter och resultat från tre olika utprovningar av läsförståelse. Texten är skriven av Åsa Sebestyén och Hans Albinsson, nuvarande respektive tidigare ansvarig för utvecklingen av de tyska provmaterialen. Den andra texten, *”Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ”*, behandlar elevers synpunkter på och upplevelse av utprovningar i reception samt produktion och interaktion i franska. Den är författad av Ingela Finndahl och Frank Perrotte, provutvecklare i franska, den förra doktorand i språkdidaktik, den senare provansvarig samt biträdande projektledare för Nafs. Avslutningsvis analyseras attityder, utprovningens resultat

och självskattningar från elevenkäter i samband med utprovningar av spanska receptiva provuppgifter i texten *Motiverad för uppgiften? – Test-taker feedback ur ett genusperspektiv*. Texten är författad av Lena Hedenbratt, provansvarig, och Patrik Axelson, båda provutvecklare och verksamma i utvecklingen av bedömningsstöden i spanska. Avsnittet avslutas med en gemensam referenslista för de tre texterna.

Skillnader mellan manliga och kvinnliga elevers provresultat analyseras i artikeln *Könskillnader i engelsk språkfärdighet*. I denna text fokuseras i ett långsiktigt perspektiv resultat främst i gymnasieskolans högre kurser, engelska 6 och engelska 7 – det förra på basis av nationella prov, det senare grundat på insända resultat från ett kursprov inom ramen för nationella bedömningsstöd. Artikeln är skriven av Lena Börjesson, tidigare provansvarig och Sofia Nilsson, nuvarande ansvarig för det nationella provet i engelska för kurs 6.

Kursprovet för engelska kurs 7 innehåller en uppgiftstyp med fokus på *usage*, det vill säga språklig precision i form av ord och fraser samt grammatiska strukturer. Uppgiftstypen syftar till att pröva förmågor såväl inom den receptiva som den produktiva domänen. I artikeln, *Usage i kursprovet för engelska 7*, redovisas analyser av dessa uppgifters egenskaper och relation till andra frågeformat. Texten är skriven av John Löwenadler och Iani Hoff, båda provutvecklare, den förre universitetslektor i ämnesdidaktik med inriktning mot språk, den senare ansvarig för utvecklingen av materialet för engelska 7.

I artikeln *Vokabulär och bedömning av skriftlig förmåga* jämförs elevtexter skrivna i de nationella proven i engelska i grundskolans årskurs 6 och 9 samt i gymnasieskolans kurser engelska 5 och engelska 6, som bedömts uppnå kraven för betygsnivåerna E och A avseende skriftlig förmåga. I studien undersöks såväl gemensamma drag som skillnader i texternas vokabulär mellan årskurser och betygssteg. Texten är författad av Eva Olsson, provutvecklare och universitetslektor i pedagogik med inriktning mot språk.

De nationella provens mottagande och användning på skolor står i fokus för artikeln *Proven i praktiken – om nationella prov i engelska*. I texten diskuteras olika aspekter av förberedelser inför proven för elever och lärare, samt genomförande och efterarbete i form av bedömning och rapportering. Texten är skriven av tre provutvecklare med vardera 20–25 års erfarenhet av undervisning i engelska i grundskolans senare år och på gymnasiet: Ulrika Arvidsson, Lena Asp och Jerker Brorsson.

Den sista texten i rapporten blickar framåt och har titeln *Språkprov på dator – vad händer då?* Den inleds med en belysning av olika validitetsaspekter i relation

till aktuell forskning och tar därefter upp två exempel på studier gjorda inom Nafs-projektet – prövning av läsförståelse respektive skriftlig produktion och interaktion som pappers- eller datorbaserade prov. Artikeln är författad av tre provutvecklare: AnnaKarin Lindqvist, även projektledare, Ulrika Arvidsson samt Sofia Nilsson, provansvarig för engelska 6.

Sammantaget ger de olika artiklarna en bild av det mångfacetterade utvecklingsarbete som ligger bakom de nationella proven i engelska och moderna språk. Texterna kan givetvis läsas var för sig, men det inledande avsnittet med sin bakgrund och sammanfattning liksom den första texten, som ger en begreppslig bakgrund och förankring, gör förhoppningsvis såväl kontexten som konklusionerna tydligare.

Referenser

- Erickson, G. (2017). Experiences with Standards and Criteria in Sweden. I Blömeke, S. & Gustafsson, J-E., *Standard Setting in Education. The Nordic Countries in an International Perspective* (s. 123–142). Cham, Switzerland: Springer International Publishing AG.
- Erickson, G., Gustafsson, J-E. & Nyström, P. (2017). *Förslag till systemramverk för nationella prov*. Rapport till Skolverket 30 juni 2017.
- Gustafsson, J-E., Cliffordson, C. & Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan: problem och möjligheter*. Stockholm: SNS förlag. Hämtad 12 juli 2018 från <https://www.sns.se/aktuellt/likvardig-kunskapsbedomning-i-och-av-den-svenska-skolan-problem-och-mojligheter/>
- Gustafsson, J-E. & Erickson, G. (2013). To trust or not to trust – teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, 25(1), 69–87.
- Gustafsson, J-E. & Yang Hansen, K. (2009). Resultatförändringar i svensk grundskola. I Skolverket, *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer*, (Kap. 3, s. 39–83). Stockholm: Skolverket. Hämtad 12 juli 2018 från <https://www.skolverket.se/publikationer?id=2260>
- Marklund, S. (1987). *Skolsverige. Del 5: Läroplaner*. Stockholm: Liber.
- Nusche, D., Halász, G., Looney, J., Santiago, P. & Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education – Sweden*. OECD: Paris. Hämtad 12 juli 2018 från <https://www.oecd.org/sweden/47169533.pdf>
- Regeringskansliet (2015). *Kommittédirektiv: Översyn av de nationella proven för grund- och gymnasieskolan, 2015:36*. Hämtad 12 juli 2018 från

ATT BEDÖMA SPRÅKLIG KOMPETENS

- <https://www.regeringen.se/rattsliga-dokument/kommittedirektiv/2015/04/dir.-201536/>
- Skolinspektionen (2012). *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*. Hämtad 12 juli 2018 från <http://www.skolinspektionen.se> > Publikationer
- Skolverket (2007). *Provbetyg – Slutbetyg – Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolans årskurs 9, 1998–2006*. Sammanfattning av rapport 300. Stockholm: Skolverket.
- Skolverket (2009). *Likvärdig betygssättning i gymnasieskolan? En analys av sambandet mellan nationella prov och kursbetyg*. Rapport 338. Stockholm: Skolverket. Hämtad 12 juli 2018 från <https://www.skolverket.se/publikationer?id=2286>
- Skolverket (2014). *Så tycker lärarna om de nationella proven 2013*. Rapport 404. Stockholm: Skolverket. Hämtad 12 juli 2018 från <https://www.skolverket.se/publikationer?id=3218>
- Skolverket (2017). *Skolverkets systemramverk för nationella prov*. Stockholm: Skolverket. Hämtad 12 juli 2018 från <https://www.skolverket.se/publikationer?id=3890>
- Statens offentliga utredning, SOU 2016:25 (2016). *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. Hämtad 12 juli 2018 från <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2016/03/sou-201625/>
- Tholin, J. (2006). *Att kunna klara sig i okänd natur*. Doktorsavhandling. Göteborg: Göteborgs universitet. Hämtad 12 juli 2018 från <https://gupea.ub.gu.se/handle/2077/16892?locale=sv>

Bedömning av språklig kompetens

Guðrun Erickson

Bedömning är en naturlig del i lärandet. Redan som små barn värderar vi våra färdigheter och framsteg, inom oss själva men också påverkade av kommentarer från jämnåriga eller vuxna. När så skolan börjar institutionaliseras värderingen successivt och det blir allt vanligare med utifrån kommande synpunkter och omdömen av mer eller mindre formellt slag. I samband med detta marginaliseras inte sällan den egna bedömningen av lärandet.

Denna text syftar till att belysa bedömningens olika funktioner i lärande och undervisning, med fokus på frågor som behöver ställas och begrepp som ligger bakom analyser av kvaliteten i processer, uppgifter, resultat och slutsatser. De grundläggande principerna är generella för all slags kunskapsbedömning, men exemplifieringarna som görs handlar om språk. Med anknytning till den rapport som denna text inleder, ägnas också en betydande del av texten åt de nationella material för bedömning i språk som på Skolverkets uppdrag utvecklas inom det s.k. Nafs-projektet (Nationella prov i främmande språk) vid Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet. Förhoppningsvis kan texten i sin helhet också relateras till bedömningar av skilda slag, inte minst till sådana som görs kontinuerligt i undervisningen.

Definitioner och uppfattningar

Bedömning definieras på olika sätt, beroende på användningsområde och användare. Exempel på synonymer i lexikon sträcker sig från *uppskattning*, *värdering* och *granskning* till *omdöme*, *dom*, *kritik* och *betygsättning*. Orden i sig visar på ett intressant sätt både bredd i uppfattningar och värderingsrelaterad laddning. Det är också vanligt med en glidning mellan begreppen *bedömning* och *utvärdering*, liksom mellan *bedömning* och *prov*. Utan att gå närmare in på orden i sig, kan konstateras att begreppet *utvärdering* i regel avser en mera omfattande och systemrelaterad aktivitet med bredd i syften, respondenter och metodanvändning, medan *prov* tvärtom kan ses som en speciell typ av bedömning som oftast genomförs vid en definierad tidpunkt och omfattar flera individer. I Skolverkets terminologi används begreppet 'nationella prov' för obligatoriska prov, i regel på fasta dagar, till exempel i engelska,

medan 'bedömningsstöd' också innefattar material i andra ämnen, till exempel franska, spanska och tyska, där användningen beslutas lokalt. Det kan dock konstateras att ordet bedömning oftast används på ett brett och övergripande sätt och innefattar både kontinuerlig bedömning av olika slag och prov med skilda syften.

Det torde råda tämlig konsensus kring att bedömning är en del av den pedagogiska och didaktiska helheten, dvs. att den hör samman med lärande och undervisning. Detta synsätt manifesteras också i det inflytelserika språkdokumentet *Common European Framework of Reference for Languages: Learning, teaching, assessment* (Council of Europe, 2001), tillgänglig på 40 språk, bland annat på svenska med titeln Gemensam europeisk referensram för språk, GERS (Europarådet/Skolverket, 2009). Vidare är det uppenbart att skola och utbildning i allmänhet och bedömning och betyg i synnerhet engagerar många och att åsikterna går vitt isär. Detta bottnar inte sällan i egna erfarenheter och har givetvis ofta även principiella och politiska grunder. Följaktligen kan man identifiera ett antal uppfattningar om bedömning, som går från det mest positiva och hoppfulla till det mest negativa och pessimistiska (Erickson & Gustafsson, 2017). I centrum för argumenten finns dock så gott som alltid användningen i vid mening av bedömning och dess resultat. Vi får anledning att återkomma just till användning senare i texten, i samband med diskussioner kring giltighet och kvalitet.

Bedömningens dubbla funktion

Bedömning är, som sagt, en naturlig komponent i lärande och undervisning. Som elev är man förhoppningsvis intresserad av att lärandet går framåt och att man når så långt som möjligt. Lärare resonerar rimligen på samma sätt. Om man i sin profession är fokuserad på att elever lär sig så bra som bara möjligt är av det man själv uppfattar som positivt och viktigt, vore det högst besynnerligt om man inte också var intresserad av att ta och få reda på hur det fungerat och fungerar och hur det skulle kunna förbättras ytterligare, även vad gäller den egna undervisningen. Detta är just vad god bedömning åstadkommer. Bedömningen är till för att stödja och öka lärandet, både för elever och lärare. *The Assessment Reform Group* (ARG), en grupp forskare som var verksam i Storbritannien från sent 1980-tal till 2010, beskriver i sina forskningsbaserade principer kring bedömningens funktion för lärande på följande sätt:

BEDÖMNING AV SPRÅKLIG KOMPETENS

“Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002).

Detta innebär följaktligen ett gemensamt sökande och tolkande av evidens som kan användas både av elever och lärare för att kunna göra tre centrala saker: (1) kartlägga hur elevernas individuella kunskapsprofiler ser ut, vad gäller styrkor och svagheter, självförtroende, motivation och preferenser; (2) klargöra så tydligt som möjligt vart det hela är på väg, dvs. vilka målen är och (3) tillsammans tänka ut hur den bästa möjliga vägen till dessa mål – och förhoppningsvis bortom dem – kan utformas, både vad gäller elevers och lärares arbete.

Denna pedagogiskt inriktade bedömning kallas ibland formativ bedömning (Scriven, 1967), ibland Bedömning för lärande (ARG, 2002; Black och Wiliam, 1998). Det mest karakteristiska är, som sagt, den pedagogiska funktionen samt att den försiggår kontinuerligt under lärandets gång. Bedömningen har emellertid ytterligare en basfunktion, nämligen att bidra till en rättvis och likvärdig bedömning av elevernas kunskaper, även detta för att öka kvaliteten, såväl individuellt som på systemnivå. Denna bedömning görs i regel i och för grupper, dvs. innefattar fler elever, och använder mera standardiserade metoder än den kontinuerliga bedömningen. Den refereras stundtals till som summativ bedömning (Scriven, 1967) eller Bedömning av lärande (Harlen, 2007) och försiggår både i anslutning till mindre projekt eller undervisningssekvenser i klassrummet eller som större summeringar av kunskapsläget, stundtals på nationell och internationell nivå. Den kanske mest uppenbara summativa bedömningen är lärares betygssättning efter en termin eller en kurs, då ett antal iakttagelser av skilda slag summeras till ett enda resultat, i Sverige i nuläget stött av kunskapskrav och presenterat i form av en av bokstäverna A till F.

Formativ och summativ bedömning står inte i något motsatsförhållande till varandra; tvärtom hör de ihop och utgör olika aspekter av samma fenomen (Gardner, 2012; Harlen, 2012; Taras, 2005). Syftena kan variera liksom metoderna, men båda har sin naturliga och självklara roll i lärande och undervisning. Detta gäller inte minst i ett land som Sverige, där lärare har en tämligen unik roll i och med ansvaret för betygssättning av enskilda elevers kunskaper. Det som också bör fastslås är att all kunskapsbedömning i grunden vilar på samma principer. Dessa kan formuleras på olika sätt, men det torde råda enighet om att de i vid bemärkelse inkluderar aspekter av transparens, giltighet, tillförlitlighet och respekt. Det handlar följaktligen om att skapa så stor tydlighet som

möjligt både kring vad som ska läras och följaktligen bedömas, med avseende såväl på innehåll som metod, liksom att se till att bedömningen är giltig visavi styrdokumentet och att resultaten går att lita på så långt möjligt är. Respekt har att göra med de övergripande etiska aspekter som måste prägla såväl synen på individer som på konsekvenser av de sätt på vilka bedömningsresultat används. Detta för osökt över till några av de grundläggande begrepp som är viktiga i planering, genomförande och analys av kunskapsbedömningar.

Grundläggande begrepp

Alla akademiska utbildningar, inte minst lärarutbildningar, torde – och bör självklart – i någon form beröra de två grundläggande begreppen *validitet* och *reliabilitet*. Dessa har i allra högsta grad att göra med kvalitet och gäller både forskning och utveckling på ett övergripande plan och frågor om kunskapsbedömning. *Validitet*, eller giltighet, kan sägas utgöra en grundbult i kvalitetssäkringen av bedömning. Enligt den traditionella definitionen handlar detta om att bedöma det som verkligen ska bedömas, dvs. det så kallade ’konstruktet’, som definierats t.ex. i en kursplan. Begreppet validitet har emellertid kommit att få en utvidgad betydelse, där slutsatser, användning och konsekvenser är i fokus. I synnerhet två företeelser brukar anges som allvarliga hot mot validitet, nämligen att man i sina bedömningar har med för lite av den kunskap som avses respektive tvärtom, att man tar med sådant som inte hör dit. Dessa två hot, som refereras till som *construct under-representation* och *construct irrelevant variance* (Messick, 1996), är ytterst påtagliga och manifesteras både i mera vardagliga bedömningssituationer och i stora internationella studier. Ett tydligt exempel på underrepresentation är att muntlig språkfärdighet inte ingick i den europeiska språkstudien 2011 (European Commission, 2012), vilket givetvis hade konsekvenser för resultaten. Vidare kan konstateras att om elevers förståelse av det de läst och hört enbart ska visas genom långa, skriftliga svar, finns risk att man på ett irrelevant sätt bedömer skrivande i stället för förståelse.

En viktig aspekt av och förutsättning för validitet är *reliabilitet*, eller tillförlitlighet, dvs. att resultaten är så fria som möjligt från slump och därmed går att lita på som indikatorer på en persons förmåga inom ett visst område. Detta kräver att uppgifter ska vara relevanta, klara och entydiga i sin utformning, men det handlar också om att bedömare är rimligt överens i sina bedömningar. Strävan efter samstämmighet är givetvis central och behöver understödjas på alla tänkbara sätt men har ibland också lett till oönskade effekter. Ett exempel

BEDÖMNING AV SPRÅKLIG KOMPETENS

är att man i vissa bedömningsområden helt undviker uppgifter med produktiva svar och skapar prov med bara flervalsfrågor, vilket framstår som klart besynnerligt givet en kommunikativ språksyn, där personlig och kreativ användning av språket, såväl muntligt som skriftligt – på egen hand och tillsammans med andra – är central.

Begrepp som validitet och reliabilitet uppfattas ofta som tekniska och teoretiska, med relevans för akademiska analyser av skilda slag men föga brukbara i konkreta pedagogiska sammanhang. Detta är att snäva in betydelsen av något som i själva verket fyller en högst användbar och konkret funktion, inte minst i planering och analys av arbetet med bedömning. En manifestation av detta finns i de *Guidelines for Good Practice in Language Testing and Assessment*, som tagits fram inom *European Association for Language Testing and Assessment* (EALTA). Detta dokument kan laddas ner fritt från associationens hemsida och innehåller dels en inledning kring det som karakteriserar god bedömning på ett generellt plan, dels riktade avsnitt till olika kategorier av användare, inte minst lärare och lärarutbildare. Riktlinjerna, som finns tillgängliga på 35 språk, har visat sig fungera väl och används runt om i Europa och övriga delar av världen som stöd för utveckling av god praxis kring språklig bedömning, såväl i klassrum som vid lärarutbildningsinstitutioner och centra för utveckling av storskaliga prov.

Begreppet validitet betraktas ofta som övergripande, som ett slags kvalitetsparaply som inrymmer andra väsentliga aspekter inom bedömningsområdet, inte minst reliabilitet. Detta visar sig bl.a. i att reliabilitet stundtals refereras till som *scoring validity* (Weir, 2005). Likaså inordnas i regel etiska aspekter i validitetsbegreppet, inte minst vad gäller användningen av de resultat som genereras av olika bedömningsformer. Detta betonas inte minst av Samuel Messick, som inspirerat till begreppet *consequential validity*, genom att framhålla bredden i begreppet: “Broadly speaking, then, validity is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use” (Messick, 1989, s. 13). Begreppet ’konsekvenser’ används här i vid bemärkelse och innefattar såväl individuella som pedagogiska och strukturella företeelser, liksom den påverkan prov och bedömning kan ha både på elevers syn på sitt kunnande och sin förmåga och på lärares undervisning, det senare refererat till som *washback effects* (Messick, 1996).

Slutligen bör också bedömningars praktikalitet eller ’genomförbarhet’ (på engelska *practicality* eller *feasibility*) nämnas som en aspekt av validitet. Uppgifter och prov av olika slag måste vara rimliga till sin typ och tidsåtgång för att kunna

genomföras och, inte minst viktigt, kunna generera resultat som bidrar till en rättvisande bild av det man vill bedöma.

Grundläggande frågor

Likaväl som det finns grundläggande principer för bedömning, kan vissa generella basfrågor med fördel användas både i planering och genomförande av bedömning i undervisning och i samband med storskalig bedömning, t.ex. nationella prov (Erickson, 2011; Takala, Erickson, Figueras & Gustafsson, 2016). Frågeställningarna berör centrala aspekter av bedömning och hjälper till att successivt kontrollera att såväl processer som produkter håller god kvalitet. Detta innebär inte minst att bedömningen i ett helhetsperspektiv lever upp till de två grundläggande funktionerna att stödja både lärande och likvärdighet. Frågan *varför* handlar om syften, *vad* om den kunskap som ska bedömas, *hur* om olika tillvägagångssätt vid konstruktion av uppgifter av skilda slag, liksom vid analys av resultat; *när* berör tidpunkter, frekvens och sekvensering, *vem* aktörer, och slutligen fokuserar frågan *och...?* på användning och effekter. Alla dessa frågor har med kvalitetssäkring att göra, och inte minst den sista hänger tydligt samman med validitet så som begreppet beskrivits ovan. Nedan berörs de olika frågorna med exempel från språkämnen i de svenska kurs- och ämnesplanerna och med visst fokus på de nationella proven och bedömningsstöden. I anslutning till detta ges också hänvisningar till de kapitel som följer i denna rapport. Det är viktigt att säga att dessa texter ofta berör flera av de grundläggande frågorna, även om de här nämns i anslutning till en särskild av dessa.

Frågan om syfte

Frågan om bedömningars syfte är central och måste behandlas tidigt, eftersom svaret påverkar både utformning av uppgifter och genomförande av själva bedömningen. En grundregel är att man bör undvika att blanda för många syften, eftersom detta kan innebära att inget av det man vill åstadkomma riktigt hinns med eller fokuseras tillräckligt. Skolverket (2011, s. 7) väljer att definiera fem vida syftedimensioner relaterade till skolans uppdrag, nämligen att bedömning kan avse att *kartlägga kunskaper*, *värdera kunskaper*, *återkoppla lärande*, *synliggöra praktiska kunskaper* samt *utvärdera undervisning*. Dessa funktioner har att göra med den övergripande intentionen, utifrån vilken praktiska beslut måste fattas. De är därmed ett slags riktninggivare som behöver konkretiseras för olika behov

och situationer. Det kan handla om allt från kartläggning och diagnos till påverkan av attityder vad gäller olika aspekter av ämnet till summerande bedömningar som stöd för betygssättning. Ett exempel vad gäller språk kan vara att man genom att tydliggöra bedömningsaspekten kan få eleverna att bli [ännu] mera medvetna om vikten av att använda sitt språk, kanske inte minst vad gäller muntlighet. Om bedömningen av denna färdighet inte är synlig, finns annars risken att eleverna uppfattar den som mindre viktig från lärarens synpunkt sett.

I den här aktuella rapporten finns en artikel ([Bjerklund Larsson & Jochens](#)), som handlar om utvecklingen av ett nationellt diagnostiskt material i engelska för yngre elever, vad som utmärker denna typ av syfte och bedömning samt hur utvecklingsarbetet gått till.

Frågan om innehåll/'konstrukt'

Frågan om vad bedömningen ska försöka fånga handlar om den definition av kunskap som ligger bakom beskrivningen av respektive ämne i kurs- och ämnesplaner. Detta refereras allt oftare till som 'konstruktet', dvs. med en försvenskning av det engelska begreppet *construct*. I fallet engelska och moderna språk bygger detta på den kommunikativa och funktionella språksyn som präglat svenska styrdokument i språk sedan 1980 och framåt och som kan hänföras till Hymes' (1971) banbrytande arbete och på preciseringar gjorda av många efterföljare, t.ex. Canale och Swain (1980). Kommunikativ kompetens ses i detta sammanhang som en övergripande kompetens, som inrymmer ett antal delkompetenser, i GERS (2009) identifierade som en lingvistisk, en sociolingvistisk och en pragmatisk kompetens. Kärnpunkten handlar om att använda språket i meningsfull kommunikation med andra kring ett – så långt möjligt är – autentiskt innehåll. Detta kallas i GERS för en handlingsorienterad språksyn, som bygger på de tre grundkompetenserna och vars mål är utpräglat funktionellt. Tolkningen av detta vållar ibland problem, i synnerhet vad gäller synen på mål och medel – vad som behövs för att språket ska fungera och utvecklas. Det finns flera sätt att förklara detta, inte minst för elever, t.ex. med hjälp av metaforen om husbygget, där byggstenarna, murbruket och ytskiktet är självklara, men där huset med sin funktionalitet och trivsel är det uppenbara målet. Med andra ord: man kan förstås varken lära sig eller vidareutveckla sitt språk utan till exempel ord och fraser, grammatik och syntax, stavning, uttal och prosodi, men det är det fungerande språket i sin helhet, det man både kan och vill använda

för kommunikation, som är i fokus (Erickson, 2016). Detta är också det som beskrivs både i de svenska styrdokumenterna och i GERS, som har en nära knytning till varandra, och som därmed utgör svaret på frågan *vad*. Detta innebär också att det är användningen av språket i vid bemärkelse som ska utvecklas, undervisas och bedömas. Byggstenarna är givetvis av central betydelse, men de är inte mål vare sig i sig eller var för sig. Detta innebär också att de nationella proven har fokus på funktionella kompetenser – receptiva, produktiva och interaktiva – och att inom ramen för detta såväl lingvistiska som sociolingvistiska och pragmatiska företeelser ingår.

I rapporten finns i synnerhet två texter med fokus på olika aspekter av konstruktet: [Löwenadler & Hoff](#) undersöker den för de nationella proven nya uppgiftstypen *usage* i kursprovet för engelska 7, vilket har att göra med tydlighet och precision i språket. Temat för [Olssons](#) artikel är vokabulär på olika nivåer inom ramen för skriftlig produktion och hur detta relaterar till betygssättning.

Frågan om metoder

Hur-frågan är omfattande och innefattar många inbördes olika aspekter av bedömning: utformning av uppgifter, bedömningsprinciper och anvisningar, beslut om analyser och gränser för olika kvalitativa nivåer och, inte minst väsentligt, utveckling av sätt att presentera resultat, som kan åstadkomma såväl återkoppling som framåtkoppling (jfr Sadler, 1989: *feedback* och *feed forward*). I detta ligger kontinuerliga analyser av bakomliggande syften och styrdokument, alltså av vad man avser bedöma och varför, av resultat, för alla elever och för undergrupper, samt av användarnas synpunkter och reaktioner. Allt detta relaterar till frågor om validitet och reliabilitet, där t.ex. sätten eleverna förväntas svara på och agera bör varieras. Detta rimmar väl med det som erfarenhet och forskning tydligt visar, nämligen att människor lär sig på distinkt olika sätt både vad gäller metoder och tid. I ljuset av detta förefaller det därför högst rimligt att de också får visa sina kunskaper på olika sätt. Vad som är mest adekvat varierar dock delvis med syfte och situation; till exempel är uppgifter med flervalsoalternativ mest lämpade för storskaliga prov som de svenska nationella proven och bedömningsstöden, där gedigna och nödvändiga utprovningar är del av utvecklingsarbetet, och frågan har också att göra med målgruppens ålder och mognad. Det kan också nämnas att variation av bedömningsmetoder är något som tycks uppskattas av elever, vilket bland annat visats i en stor, internationell studie av

BEDÖMNING AV SPRÅKLIG KOMPETENS

elevs och lärares uppfattningar om prov och bedömning i språk (Erickson & Gustafsson, 2005).

En ytterligare aspekt av hur-frågan är hur man skapar så stor stabilitet och tillförlitlighet i bedömningen som möjligt och därmed bidrar både till giltighet och tillförlitlighet. I detta ligger, förutom vikten av stora utprövningsvolymerna och ankaruppgifter för att kunna jämföra grupperns resultat, även frågan om holistisk respektive analytisk bedömning. Detta handlar om helhetsbedömningar eller bedömningar av olika aspekter var för sig, med eller utan slutlig sammanslagning, är att föredra. Forskning ger inga entydiga belegg för att något av sätten i sig skulle vara överlägset andra; i stället torde det vara starkt beroende av syftet med den aktuella bedömningen. Detta bör också relateras till frågan om matriser, vars syfte är att förtydliga innehåll och kvalitativa nivåer för eleverna genom att bryta ned styrdokumentens formuleringar i mindre och tydligare delar som kan granskas, analyseras och bedömas var för sig. Även i detta fall bör behov och beslut analyseras utifrån syfte och tänkta effekter, där den avsedda tydligheten får vägas mot risken för fragmentisering av den kunskap som ska bedömas. I linje med detta varnar Stobart (2012, s. 238) för att *“make the objectives so detailed that they encourage impoverished learning”*, och Torrance (2007), som har myntat begreppet *criteria compliance*, pekar på riskerna med en situation där *“assessment procedures and practices come completely to dominate the learning experience, and ‘criteria compliance’ come to replace ‘learning’”* (s. 282). Den goda ambitionen att förtydliga får följaktligen aldrig innebära att det lätt mätbara tar överhand eller att stoff förvandlas till stoft, som inte går att rekonstruera till den helhet som i slutändan ska bedömas på ett tillförlitligt och likvärdigt sätt (Erickson & Gustafsson, 2017). Kunskapsbedömning är heller inte additiv i bemärkelsen att summan av delarna är identisk med helheten; det finns alltid mellanliggande och interagerande faktorer som inte låter sig fångas i enkelt avprickbara beskrivningar. Det är just för att komma åt detta sammanbindande i språket, att kunna kombinera det specifika med det generella, som lärares övergripande analys och professionella värdering intar en central roll i all kunskapsbedömning.

Det bör vidare nämnas att frågan om det språk som används i bedömningarna också faller under avdelningen ’hur?’. Även här måste syfte och kontext analyseras. I detta finns aspekter av görbarhet/praktikalitet samt likvärdighet. I fallet nationella prov och bedömningsstöd, vars syfte bland annat är att skapa ett standardiserat underlag för lärares användning av resultat, har sedan mycket länge en enspråkig princip tillämpats. Detta innebär att enbart målspråket används, såväl i instruktioner som förväntade svar. För att kvalitén i detta ska

optimeras, spelar utprövningar även av instruktioner en viktig roll för att undanröja risken för att anvisningarna försvårar eller hindrar eleverna från att förstå vad som efterfrågas i uppgiften. Erfarenheterna av detta sätt att använda målspråket är mycket positiva.

Två texter i rapporten fokuserar på hur-frågan från två olika utgångspunkter: [Schönberg & Lindqvist](#) undersöker och diskuterar frågan om flervalssuppgifter och ger exempel från en utprövning i kursen engelska 5 av samma text men med olika antal alternativa svar. [Arvidsson, Asp & Brorsson](#) beskriver och resonerar kring de nationella provens mottagande och användning på skolor och berör såväl förarbete och genomförande som efterarbete och rapportering.

Tidpunkter och frekvenser

Frågan om tid gäller såväl när och hur ofta bedömningar görs som hur lång tid som ägnas åt bedömningsaktiviteter av skilda slag. I detta är det viktigt att vara tydlig kring vilket syfte olika bedömningar har, inte minst för att klargöra för elever att de inte är utsatta för ständig granskning under sitt lärande (Erickson, 2016). Detta kan lätt leda till ängslan och ovilja att använda språket, vilket är rakt emot den språksyn som genomsyrar styrdokumentet. Här är också det egna, aktiva språkbruket av stor betydelse, till exempel hur man förklarar och distingerar mellan begrepp som observation, kommunikation kring kunskap, feedback och bedömning. Betydelsen av att skapa rimlig tydlighet kring när bedömning sker och vad som bedöms berörs även i GERS (Europarådet/Skolverket, 2009) där det i samband med diskussion om kontinuerlig bedömning varnas för att *”göra livet till ett enda oändligt prov för eleven och till en byråkratisk mar-dröm för läraren”* (s. 183). Det måste dock samtidigt sägas att det naturligtvis är viktigt att inte reducera bedömningstillfällena till ett minimum, där väldigt mycket står på spel vid enstaka provtillfällen. Allt detta understryker vikten av att ha en öppen och kontinuerlig dialog både med elever och kolleger kring frågor om lärande, undervisning och bedömning och att tillsammans söka efter goda balanser och lösningar.

Frågan om aktörer

Frågan om vilka kategorier av personer som medverkar i bedömningar är viktig av flera skäl och har såväl med lärande som likvärdighet att göra. Det handlar

BEDÖMNING AV SPRÅKLIG KOMPETENS

både om den viktiga subjekt-/objektfrågan och om ansvarsfördelning, samverkan och respekt. Det är i detta sammanhang viktigt att komma ihåg att man som lärare inte är – och heller inte bör vara – ensam i sitt bedömningsarbete. Om detta skulle vara fallet, blir bördan mycket snabbt väldigt tung. Givetvis har lärare ett formellt ansvar till exempel för betygssättning, men det innebär inte att man har mycket hjälp i arbetet fram till detta både av kolleger och elever. För att använda ett uttryck myntat av ett mycket litet barn, måste man ”bäras åt” (Erickson, 2009a) för att optimera både processen och produkten, dvs. arbetet och resultatet. Eleverna själva har insikter om sina kunskaper och sin egen kunskapsutveckling av ett slag som lärare behöver och kan få ta del av genom att, som en del av undervisningen, på kloka sätt använda sig av själv- och kamratbedömning. Det är dock synnerligen viktigt att eleverna successivt fått vänja sig vid sättet att arbeta och att klimatet dem emellan är positivt och öppet; annars kan oavsedda, negativa effekter uppstå. Vad gäller detta aktiva medarbetande finns material att hämta från flera håll, bland annat från nationella bedömningsstöd och från de olika europeiska språkportfolios (ESP), som är framtagna för att konkretisera och operationalisera den handlingsorienterade språksyn som präglar både den gemensamma europeiska referensramen för språk (GERS) och de svenska kurs- och ämnesplanerna. Det finns tre språkportfolios av detta slag i Sverige, två för grundskolan (6–11 och 12–16 år) och en för den gymnasiala nivån (16+). Grundskolans material är nedladdningsbart från Skolverkets hemsida, portfolion för gymnasiet från Fortbildningsavdelningen vid Uppsala universitet.

Samverkan mellan kolleger kring olika undervisningsrelaterade frågor har ett odiskutabelt värde, inte minst vad gäller bedömning. Detta handlar bland annat om att utveckla en samsyn kring styrdokumentens språksyn både vad gäller innehåll och nivå och att finna arbetsformer där sambedömning, medbedömning och diskussion av resultat utgör en naturlig del. Likaså bör externa material och influenser nämnas, till exempel nationella prov och bedömningsstöd, vars avsikt är att bidra med ett vidgat perspektiv och med den *second opinion*, som ofta nämns som ett starkt bidrag till kvalitet.

Som del av frågan om ’vem’ bör också den grundläggande subjekt-/objektfrågan nämnas, nämligen den som definierar bedömningens inriktning. Här är det tyvärr alltför vanligt att på olika nivåer använda sätt att uttrycka sig, som ger vid handen att objektet för bedömning skulle vara individer, inte deras kunskaper. Det sägs alltför ofta att man ’bedömer elever’, när det givetvis handlar

ATT BEDÖMA SPRÅKLIG KOMPETENS

om att bedöma elevers kunskaper och kunskapsutveckling. Det finns all anledning att vakta på sina ord, eller som en av språkbedömningens mest kända forskare, Lyle Bachman, 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” (Bachman, 1990, s. 20).

I rapporten finns tre texter med fokus på något som sedan lång tid präglar arbetet i det nationella provprojektet, nämligen elevers aktiva medverkan i utvecklingen av material. De aktuella texterna utgår från utvecklingen av bedömningsstöden i moderna språk. Efter en gemensam inledning finns en del som fokuserar på elevers synpunkter på läsförståelse i franska, tyska och spanska (Sebestyén & Albinsson), en som behandlar elevers i franska attityder till autenticitet (Finndahl & Perrotte) samt en text som diskuterar genusaspekter på basis av åsikter uttryckta av elever som medverkat i utprovning av receptiva uppgifter i spanska (Hedenbratt & Axelsson).

Frågan om användning och konsekvenser

Den sista av frågorna som bör ställas vad gäller utformning av bedömningar och prov gäller användning och konsekvenser av de resultat som framkommer. Detta för osökt tankarna till det gamla engelska ordspråket från 1300-talet, nämligen att ”the proof of the pudding is in the eating”. Inget kan alltså riktigt rubriceras som lyckat om det inte har prövats och använts. Kvalitén i bedömning ligger följaktligen i användningen och effekterna av användningen, vilket anknyter till den expanderade och konsekvensrelaterade syn på validitet som tidigare nämnts (Messick, 1989). I linje med detta betonar Stobart (2012) att all bedömning måste kunna visa upp sin giltighet eller validitet. Vad gäller formativ bedömning eller bedömning för lärande innebär detta till exempel att det måste kunna visas i vad mån bedömningen bidragit till ökat lärande, innan man uttalar sig om arbetets funktionalitet och framgång.

Och-frågan har emellertid ett ännu bredare fokus än den centrala frågan om elevers lärande. Den innefattar även det kunskapsbidrag bedömningens resultat ger för lärarens vidare planering av undervisning och vidare bedömning, såväl på individuell som mera kollektiv nivå. Dessutom ligger i frågan om användning och konsekvenser av bedömningar även en strukturell aspekt. Shohamy (2001) framhåller att bedömning inte är en isolerad företeelse utan integrerad i sociala,

utbildningsmässiga och politiska kontexter. Hon anlägger vidare ett etiskt perspektiv på frågan och varnar för direkt och indirekt, negativ maktutövning med hjälp av prov och bedömning. I stället betonas vikten av utveckling av ett kritiskt perspektiv, som fokuserar på ansvar och rättigheter för olika inblandade parter. Detta perspektiv är givetvis relevant också i planeringar av internationella studier och nationella bedömningssystem. Olika möjliga användningar och konsekvenser måste tänkas in redan från början vad gäller eventuella jämförelser över tid och mellan länder, och i det nationella fallet inte minst vad gäller frågor om måluppfyllelse och likvärdighet.

Rapporten innehåller två texter som relaterar till användnings- och konsekvensfrågan. Börjesson & Nilsson analyserar utfallet på prov i engelska på högre gymnasial nivå ur ett genusperspektiv och relaterar det till lärares kursbetyg. Slutligen blickar Lindqvist, Arvidsson & Nilsson framåt och rapporterar, på basis av resonemang om begreppsvaliditet, två studier där elever genomfört prov på papper respektive dator.

De nationella proven och bedömningsstöden i språk

Ovanstående resonemang om validitet vad gäller användning och konsekvenser leder fram till de svenska nationella proven och bedömningsstöden i språk. Dessa utgör en del av det nationella bedömningssystemet, som beskrivs kortfattat i avsnittet 'Bakgrund och sammanfattning' i denna rapport. Utvecklingen av materialen relaterar till de begrepp och frågor som diskuterats ovan och kommenteras ytterligare nedan. Detta görs kortfattat och i anslutning till de gemensamma principer som sedan många år ligger på Nafs-projektets hemsida (www.nafs.gu.se) och som också finns översatta till engelska, franska, spanska och tyska. På hemsidan finns vidare en kort text på engelska (Erickson, 2017) med information om de svenska nationella provmaterialen, avsedd för en internationell publik. Texten uppdateras successivt, då förändringar av uppdraget görs till följd av beslut på den politiska nivån. I det följande ges korta kommentarer till projektets gemensamma principer.

Gemensamma principer för projektet Nationella prov i främmande språk

Provmaterialen har delvis olika syften men bygger på gemensamma principer.

- Proven utgör en konkretisering av respektive läro- och kurs-/ämnesplans kommunikativa och funktionella språksyn. I samtliga material visas i s.k. testspecifikationer relationen mellan de olika uppgifterna och kurs-/ämnesplanen. Här framgår också vilka mål som inte prövas.

Kommentar: Provens innehållsliga inriktning liksom den kommunikativt, funktionella språksyn som utgör fundamentet för materialen kommenteras ovan i anslutning till frågan 'Vad'. Vad gäller testspecifikationer kan sägas att de existerar i två former, nämligen dels en extern, som är den som visas i proven, dels en intern, som på ett tämligen detaljerat sätt beskriver uppgifter, principer för sekvensering, data från utprovningar, riktvärden för olika mättekniska egenskaper etc. Båda typerna av specifikationer ingår i de provramverk som finns i anslutning till de olika nationella proven och som bygger på Skolverkets gemensamma ramverk för samtliga nationella prov (Skolverket, 2017).

- Proven skall vara ett stöd för lärare i bedömningen av elevernas språkfärdighet. De utgör därmed ett komplement till den kontinuerliga bedömning som görs i undervisningen.

Kommentar: Formuleringen kring provens syfte och relation till lärarens sammantagna bedömning och betygssättning av individuella elevers kunskaper har varit densamma sedan införandet av de nationella proven på 1990-talet. Någon indikation om vad som kan anses vara 'lagom stöd' har inte givits. Detta har i tilltagande grad upplevts som besvärande, i synnerhet eftersom en betydande variabilitet har kunnat noteras vad gäller användningen av provresultaten. Frågan berördes i den utredning av de nationella proven som redovisas i SOU 2016:25 och som inkluderar ett förslag att relationen mellan provbetyg och slut-/kursbetyg ska tydliggöras och att resultaten på nationella prov ska ges större uppmärksamhet än hittills. Dock betonas att provbetyget fortfarande ska kombineras med lärarens övriga iakttagelser vid betygssättning. Detta har senare lett fram till ett beslut som redovisas i det avslutande avsnittet i detta kapitel.

- Proven skall bidra till likvärdig bedömning och betygssättning i landet.

BEDÖMNING AV SPRÅKLIG KOMPETENS

Kommentar: För att de nationella proven ska kunna bidra till likvärdighet i bedömning och betygssättning behöver både utvecklingsprocessen, själva proven och genomförandet standardiseras. Detta innebär bland annat att testspecifikationer följs, att utprovningar av uppgifter görs i stora, slumpmässigt valda grupper i landet, att människor med olika typer av erfarenhet och expertis deltar både i utvecklings- och analysarbetet samt att en successiv kvalitetsgranskning görs, alltså inte bara efter det att proven givits. Detta arbetssätt har sin begreppsliga grund i resonemang om validitet och reliabilitet och syftar även till att stärka provens stabilitet över tid. Analyser visar att dels att bedömningen i proven har hög reliabilitet (Erickson, 2009b), dels att lärare i stor utsträckning anser att resultaten stämmer överens med egna bedömningar. Vidare uppvisar resultaten i engelska i stor utsträckning dels att det sammantagna provbetyget stämmer överens med lärares betygssättning, dels att lägre respektive högre slutbetyg än provbetyg är i relativ balans (se Skolverkets statistik på <https://www.skolverket.se/skolutveckling/statistik>).

- Samtliga material innehåller delar som på olika sätt prövar elevernas förmåga att förstå vad de hör och läser (reception) samt att själva uttrycka sig i tal och skrift (interaktion, produktion). Strävan är vidare att proven i så stor utsträckning som möjligt skall knyta an till de kulturmål som finns fastlagda i kursplanerna. Detta sker t.ex. genom text- och uppgiftsval samt genom användning av olika varianter av talat språk i avlyssningsavsnitten. Vidare finns i, och i anslutning till, vissa av materialen delar som fokuserar reflektion, själv- och kamratbedömning.

Kommentar: Frågor om bedömningars relation till kurs- och ämnesplaner (så kallad *construct coverage*) har delvis berörts ovan i anslutning till frågan om 'vad', dvs. avsett innehåll. De nationella proven, som ska vara en återspeglning av styrdokumentet, reflekterar därmed en kunskaps- och ämnessyn med kommunikation i fokus, grundat i det skifte från ett mera strukturellt sätt att betrakta språk till ett väsentligen mera funktionellt, som präglade senare delen av 1900-talet. Samma synsätt genomsyrar den gemensamma europeiska referensramen, GERS, som gradvis fått allt större inflytande, även i det svenska systemet. I enlighet med de förmågor som beskrivs som mål i styrdokumentet prövas även integrerat i de receptiva, produktiva och interaktiva kompetenserna förmågan till anpassning av språket till syfte, mottagare och situationer, strategier för användning av språket samt kulturell och interkulturell kompetens. Konkretiserat

ATT BEDÖMA SPRÅKLIG KOMPETENS

innebär detta tre delprov med olika fokus i de nationella provmaterialen: ett muntligt (produktivt och interaktivt), ett receptivt med två delar, läsa respektive lyssna, samt ett som prövar elevernas skriftliga, kommunikativa förmåga.

- Ambitionen är att såväl material som uppgifter i proven skall vara så autentiska som möjligt.

Kommentar: Autenticitet anknyter till den funktionella språkliga kompetens som beskrivs i språkens styrdokument och ses därmed som en aspekt av validitet. Vidare kan nämnas att autenticitet stundtals definieras i två kategorier, nämligen *the authenticity of the text* respektive *the authenticity of the task* (Carroll, 1980), dvs. med fokus på graden av verklighetsanknytning i det material som används i olika uppgifter respektive den anknytning till verkliga situationer och äkta språk-användning som uppgifterna syftar till att efterlikna och pröva. Det finns dock anledning att påpeka att både institutionaliserad undervisning och bedömning i någon mån är fiktiva till sin karaktär, även om den goda och språksynsanknutna ambitionen givetvis är att göra dem så verklighetstroga som möjligt.

- Instruktioner till eleverna ges på respektive målspråk, och eleverna förväntas också använda enbart målspråket.

Kommentar: Beslutet att enbart använda målspråket i de nationella proven har även det att göra med ett övergripande validitetsperspektiv, inte minst relaterat till den standardisering som krävs för att proven ska fylla sitt syfte. Att blanda in majoritetsspråket svenska vore inte acceptabelt visavi den stora andel elever för vilket detta inte är ett förstaspråk, och att, i logisk parallellitet med detta, producera anvisningar och bedöma svar på alla de modersmål som finns representerade i den svenska grund- och gymnasieskolan faller på sin egen orimlighet. Som tidigare nämnts, förutsätter dock anvisningar på målspråket gedigna utprovningar, vilka lett fram till en situation där enspråkigheten tycks fungera bra, såväl vad gäller tydlighet och begriplighet som acceptans bland elever och lärare.

- Eleverna ges i proven möjlighet att visa vad de kan, inte primärt vad de inte kan; detta innebär bl.a. att minuspoäng inte ges samt att materialens olika delar är av mycket skiftande slag, både vad gäller innehåll och svarsform, så att så många elever som möjligt skall komma till sin rätt.

BEDÖMNING AV SPRÅKLIG KOMPETENS

Kommentar: Det har funnits, och finns kanske delvis fortfarande, en tradition i bedömning att fokusera på sådant eleverna inte kan, detta säkerligen med den goda avsikten att på så sätt få ett gediget underlag för åtgärder med syfte att rikta undervisning och lärande mot områden som behöver förbättras. I detta är det dock lätt att glömma bort dels att förfaringssättet riskerar att sända tveksamma signaler till eleverna, dels att underlaget därmed kan komma att vara bristfälligt vad gäller bedömningen och betygssättningen av det eleverna faktiskt *kan* och *kan göra* med sitt språk. Detta är allvarligt, eftersom kurs- och ämnesplanernas formuleringar entydigt handlar om vad eleverna kan använda sitt språk till. För att möjliggöra för så många som möjligt att visa så mycket som möjligt av vad de kan är därför bredd och variation av stor vikt vad gäller såväl innehåll och genre som svårighetsgrad och svarsform i de uppgifter som ingår i nationella prov. Elevernas respons ges till övervägande del genom aktiv språk-användning: helt i de muntliga respektive skriftliga delproven, till ungefär hälften i det receptiva, där ett betydande antal uppgifter dock även ska besvaras genom val bland flera alternativa svar. Orsaken till användning av flervalsuppgifter är främst att det ses som högst rimligt att eleverna bereds tillfälle att visa sin förmåga att förstå utan att behöva aktivt producera/skriva sina svar. Flervalsformatet bidrar vidare till en tämligen säker och standardiserad bedömning, utan den variabilitet som åvilar uppgifter med produktiva svarsformat – vilka dock givetvis också är högst motiverade i relation till styrdokumentens språksyn. (För vidare diskussion om flervalsuppgifter, se Schönbergs & Lindqvists kapitel i denna rapport.)

- [Ambitionen är att det skall finnas en svårighetsmässig progression såväl inom som mellan de olika uppgifterna i proven.](#)

Kommentar: De stora utprövningar som görs av allt material tänkt för användning i proven ger tillsammans med ankaruppgifter möjlighet att med tillfredsställande precision avgöra olika uppgifters svårighetsgrad. Detta är till god hjälp vid komposition och sekvensering av proven, där ambitionen är att det ska finnas en progression i svårighetsgrad både inom och mellan uppgifter. Orsaken till detta är att undvika att elever krokmar eller fastnar på sin väg genom respektive delprov och i värsta fall ger upp. Om detta skulle ske, har provet misslyckats med en av sina viktigaste uppgifter, nämligen att ge varje elev chansen att visa så mycket som möjligt av vad han/hon verkligen kan göra med sitt språk. Det är emellertid viktigt att i detta sammanhang undvika en förenklad tolkning av

ATT BEDÖMA SPRÅKLIG KOMPETENS

vad 'svårt' innebär och inse att begreppet kan definieras på åtminstone tre sätt, nämligen genom analyser av (1) komplexitet i relation till det konstrukt som avses; (2) vad utprövningar visar i form av t.ex. medelvärden och spridningsmått på uppgifterna och (3) vad elever själva anser vara svårt (Erickson & Åberg-Bengtsson, 2012). Det senare spelar stor roll vid sammansättningen av delprov, eftersom det har påverkan på hur elever tar sig igenom uppgifterna. Att lägga en uppgift som kanske har hög lösningsfrekvens (medelvärdet uttryckt i procent) men som uppfattas som svår av eleverna först är därmed inte en särskilt bra idé. Detta gäller stundtals uppgifter i läsförståelse av så kallad matchningstyp (korta texter ska kombineras med utsagor av skilda slag), där elever ofta överskattar den faktiska svårighetsgraden, dvs. hur väl de faktiskt presterar. (För vidare diskussion om elevers upplevelse av matchningsuppgifter, se Erickson & Åberg-Bengtsson, 2012 samt Sebestyéns & Albinssons kapitel i denna rapport.)

- Elevernas vilja och förmåga att förstå och att förmedla ett innehåll fokuseras vid bedömningen; förtjänster analyseras före brister; vid analys av fel distingeras mellan fel som (eventuellt) stör och fel som förstör, dvs. mellan fel som i olika hög grad påverkar kommunikationen.

Kommentar: Såväl komposition som bedömning av de nationella proven reflekterar den kommunikativa och funktionella språksyn som präglar styrdokumentet i språk. I detta ligger att med systematik kommentera det eleverna visar att de kan, detta dock utan att negligera fel av olika slag. Vad gäller det senare är begriplighet i centrum, inte minst genom att en distinktion görs mellan störande och förstörande fel vilket är relevant på alla nivåer, kanske extra mycket så på de lägre. Det finns i detta naturligtvis också möjlighet att bedöma aspekter av anpassning och strategier. Vad gäller analys av begriplighet och stil är talare av målspråket en stor tillgång, men det är viktigt att komma ihåg att dessa kan representera en stor inbördes variation vad gäller uppfattningar om vad som går att förstå respektive är lämpligt språkbruk. På grund av detta är ambitionen att konsultera flera personer vad gäller bedömningen av graden av allvar i olika typer av språkliga fel.

- Utvecklingen av provmaterialen sker i samarbete med aktiva lärare i respektive språk och med personer som har respektive språk som

BEDÖMNING AV SPRÅKLIG KOMPETENS

modersmål. I samband med omfattande utprövningar ges såväl medverkande lärare som elever tillfälle att kommentera de olika uppgifterna.

Kommentar: Utvecklingen av ett nationellt prov eller bedömningsstöd tar sammantaget ungefär två år och innehåller definierade steg av planering, konstruktion, utprövning, analys och beslut. Ett kollaborativt arbetssätt tillämpas genomgående i processen. I detta spelar kontinuerlig samverkan med olika kategorier av sakkunniga en väsentlig roll. Exempel på samverkanspartners är lärare, lärarutbildare, pedagoger med specialpedagogisk kompetens, forskare inom relevanta discipliner (språk, didaktik, metod m.fl.), pedagoger med målspråket som förstaspråk samt, inte minst betydelsefullt, stora grupper av elever i förproven aktuella åldrar. Ett centralt element i utvecklingsarbetet är utprövningar med många elever, initialt i mindre format med åtföljande justeringar och revideringar, senare i slumpmässigt valda grupper i landet, med korrigerande för eventuella snedheter vad gäller t.ex. geografiska och demografiska faktorer. Riktvärdet för de stora utprövningarna är minst 400 elever per uppgift. I samband med dessa omgångar genomför eleverna de olika uppgifterna under så provliknande former som möjligt samt besvarar enkäter om sina synpunkter på relevans, svårighetsgrad, upplevd prestation etc. (Erickson, 1998; Erickson & Åberg-Bengtsson, 2012; Lodeiro, 2012). Elever har följaktligen påverkan på alla de områden och frågor som fokuseras i den här föreliggande rapporten, vilket blir särskilt tydligt i texterna om utvecklingen av bedömningsstöd i moderna språk (Sebestyén & Albinsson, Finndahl & Perrotte och Hedenbratt & Axelson). Enkäter ges även till lärarna, som ombeds kommentera delvis samma frågor som eleverna, men även sådana som anknyter till uppgifternas relation till styrdokumentet för språken. I utprövningarna ingår även s.k. ankaruppgifter, som används för att möjliggöra tolkning av resultat i relation till medverkande elevgrupper och därmed skapa ett stabilt stöd för komposition av skarpa prov.

Kontinuerlig kvalitetssäkring

I enlighet med de principer som nämnts tidigare i texten, kvalitetssäkras de nationella proven och bedömningsstöden successivt, dels under utvecklingsprocessen, dels efter det att respektive prov använts i 'skarpt läge'. Det senare gäller dock i huvudsak de nationella proven i engelska, som är obligatoriska och där SCB gör en totalinsamling av resultat. Vad gäller bedömningsstöden i

ATT BEDÖMA SPRÅKLIG KOMPETENS

franska, spanska och tyska görs inga motsvarande centrala insamlingar, eftersom materialen inte omfattas av något obligatorium. Dock uppmanas lärare att rapportera in resultat till Nafs-projektet, vilket möjliggör vissa analyser, om än av data där representativitet inte kan tas för given.

De analyser som genomförs är av såväl mera kvalitativ som kvantitativ typ. I stora drag fokuseras tre aspekter, var för sig och i relation till varandra:

- materialens relation till syfte och konstrukt, så som det beskrivs i styrdokumentet och som forskning kring olika ingående företeelser visat på;
- elevers resultat på de olika uppgifterna, studerade för hela gruppen provtagare samt för undergrupper av elever, till exempel flickor och pojkar; (vad gäller det senare, se Börjessons & Nilssons kapitel i denna rapport); i detta ligger analyser av allt från innehållsliga till etiska och mättekniska egenskaper;
- användares reaktioner, uppfattningar och förslag i samband med användning av proven. I detta är elever och lärare mycket viktiga, men även andra grupper finns med (se ovan).

Resultat av de olika kvalitetsanalyserna görs tillgängliga, inte minst i den kontinuerliga rapportering som görs på provprojektets hemsida (www.nafs.gu.se) av de nationella proven i engelska. Denna innehåller dels data från själva proven i form av betygsandelar för respektive delprov, dels rådata från de lärarenkäter som medföljer proven. Dessutom finns rapporter med resonemang och diskussion kring olika resultat. Även Skolverket publicerar regelbundet rapporter om de olika proven, med huvudsakligt fokus på uppgiftsresultat och provbetyg.

Sammantaget kan konstateras att de nationella proven i engelska har god validitet och reliabilitet, vilket visats i såväl nationella som internationella studier (Erickson, 2009b; Verhelst, 2013). Vidare framgår av Skolverkets kontinuerliga analyser av utfallet på de nationella proven att resultaten i engelska uppvisar god stabilitet över år¹. Elevers och lärares reaktioner är överlag mycket positiva, även om givetvis kritiska synpunkter också framkommer. Dessa avser dock oftast villkoren runt proven och den arbetsbörda de medför, snarare än själva proven. Även bedömningsstöden i franska, spanska och tyska uppvisar hög kvalitet och ges positiva omdömen, även om underlaget för analyser här är mera begränsat.

¹ www.skolverket.se > Statistik

Ovanstående positiva bild innebär givetvis inte att utvecklingsarbetet kring proven och bedömningsstöden mattas av, utan strävan är att hela tiden optimera materialen utifrån de förutsättningar som finns, detta med avseende både på mättekniska aspekter och brukarnas upplevelser och uppfattningar.

Sammanfattning och framåtblick

Bedömning och prov kan ha olika syften och fylla olika funktioner, och de bör i så stor utsträckning som möjligt anpassas till det och till dem de är avsedda för. I denna text har ett antal grundläggande principer och frågor berörts som alla har att göra med att öka kvaliteten i de processer och produkter som används för att, med så stor träffsäkerhet och respekt som möjligt, ta reda på vad och hur mycket människor kan och vad de kan göra med hjälp av sina kunskaper.

Det svenska nationella provsystemet genomgår för närvarande stora förändringar, som understryker behovet av en tydlig vetenskaplig, empirisk och etisk kompass, både på det övergripande planet och i utvecklingen av enskilda prov. Den troligen mest synliga av dessa förändringar är den digitalisering av systemet som är avsedd att vara genomförd år 2022 och som föregås av en omfattande, nationell försöksverksamhet mellan 2018 och 2021. Vissa förändringar införs dock redan tidigt, till exempel för proven i engelska för årskurs 9 och gymnasiet kurser 5 och 6, där delprovet med fokus på skrivande från och med höstterminen 2018 ska genomföras på dator. Förarbetet till detta har inneburit utprovningar och analyser och av allt från effekter på validitet till tekniska och praktiska aspekter samt, inte minst, brukarnas användning och upplevelser av systemet. Detta beskrivs i Lindqvists, Arvidssons och Nilssons kapitel i denna rapport och har tidigare diskuterats av Lindqvist (2012).

Bland andra förändringar som redan beslutats och genomförts kan nämnas att de nationella proven från och med höstterminen 2018 kommer att ha ett huvudsakligt syfte, nämligen att befrämja en rättvis och likvärdig betygssättning av enskilda elevers kunskaper, samt att provresultatet ska ”särskilt beaktas” vid betygssättningen. Det senare innebär inte att några explicita regler för viktning ges utan kan snarare ses som en förstärkning och förklaring av det som hittills formulerats enbart som stöd. Skolverket förklarar den nya bestämmelsen på följande sätt och kommer att ytterligare förtydliga frågan i ett Allmänt råd under hösten 2018:

ATT BEDÖMA SPRÅKLIG KOMPETENS

”Formuleringen ”särskilt beakta” innebär att provresultatet inte helt ska styra betyget, utan att provet ska vara ett stöd vid betygssättningen. Resultatet på ett nationellt prov kan alltså inte vara det enda underlaget för betygssättningen. Att resultat på ett nationellt prov ska ha en särskild betydelse innebär att det har en större betydelse än andra enskilda underlag vid lärarens allsidiga utvärdering av elevens kunskaper vid betygssättningen”².

Vidare är ett system för extern bedömning av de nationella proven under utveckling och kommer att prövas under de kommande årens försöksverksamhet. Slutligen bör nämnas att antalet obligatoriska nationella prov i gymnasieskolan har minskat, och sedan början av 2018 innefattar enbart den högsta kursen på respektive program. Detta innebär för engelskans del att provet för kursen engelska 5 enbart är obligatoriskt på yrkesprogram men kan användas även på andra program om rektor fattat beslut angående detta.

Det svenska nationella prov- och bedömningssystemet är omfattande, såväl vad gäller antal ämnen som typer av material. Förändringar genomförs successivt, baserade på empiriska iakttagelser av skilda slag och/eller politiska beslut. Vad gäller verksamheten med material i engelska, franska, tyska och spanska är de grundläggande kvalitetsprinciper som diskuterats i det här aktuella texten av stor vikt för arbetet. I de följande kapitlen konkretiseras och diskuteras delar av detta arbete.

Referenser

- Assessment Reform Group (2002). *Assessment for Learning: 10 principles. Research-based principles to guide classroom practice*. Hämtad 16 juli 2018 från <https://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing and Assessment*. Oxford: Oxford University Press.
- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148. Hämtad 16 juli 2018 från https://www.michigan.gov/documents/mde/Inside_The_Black_Box_-_Black_and_William_601104_7.pdf
- Canale, M., Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1–47.
- Carroll, B. J. (1980). *Testing communicative performance*. London: Pergamon.

² www.skolverket.se > Nationella prov

BEDÖMNING AV SPRÅKLIG KOMPETENS

- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Erickson, G. (1998). Test-taker feedback – Elevers synpunkter som del av arbetet med nationella prov i engelska. I B. Ljung och A. Pettersson (red.), *Perspektiv på bedömning av kunskap* (s. 79–100). Lärarhögskolan i Stockholm, Institutionen för pedagogik.
- Erickson, G. (2009a). ”Att bäras åt” – Om den goda bedömningens flerfaldighet och ömsesidighet. I Tornberg, U. et al. *Språkdiraktiska perspektiv. Om lärande och undervisning i främmande språk* (s. 159–173). Stockholm: Liber.
- Erickson, G. (2009b). *Nationella prov i engelska – En studie av bedömersamstämmighet*. Hämtad 16 juli 2018 från www.nafs.gu.se/publikationer
- Erickson, G. (2011). Handle with care. Om referensramen och bedömning av språklig kompetens. I Söderberg, C. (red.), *Språklärarens stora blå? En samling texter om Gemensam europeisk referensram för språk* (s. 31–37). Uppsala: Uppsala universitet, Fortbildningsavdelningen för skolans internationalisering.
- Erickson, G. (2016). Bedömningens dubbla funktion – för lärande och likvärdighet. *Communicare – et fagdidaktisk tidskrift fra Fremmedspråksenteret*, 2016, 4–9. Hämtad 16 juli 2018 från <https://www.fremmedspraksenteret.no/communicare>
- Erickson, G. (2017). *National Assessment of Foreign Languages in Sweden*. Hämtad 16 juli 2018 från <https://nafs.gu.se/information>
- Erickson, G. & Gustafsson, J-E. (2005). *Some European Students' and Teachers' Views on Language Testing and Assessment. A report on a questionnaire survey*. European Association of Language Testing and Assessment. Hämtad 16 juli 2018 från <http://www.ealta.eu.org/resources.htm>
- Erickson, G. & Gustafsson, J-E. (2017). Bedömningens dubbla funktion. I Lundgren, U. P; Säljö, R. & Liberg, C. (red.), *Lärande, skola, bildning* (4:e utg., s. 551–581). Stockholm: Natur & Kultur.
- Erickson, G. & Åberg-Bengtsson, L. (2012). A Collaborative Approach to National Test Development. In Tsagari, D. & Csépes, I. (Eds.), *Collaboration in Language Testing and Assessment* (p. 93–108). Frankfurt am Main: Peter Lang.
- Europarådet/Skolverket (2009). *Gemensam europeisk referensram för språk: lärande, undervisning, bedömning*. Stockholm: Skolverket. Hämtad 16 juli 2018 från <https://www.skolverket.se/publikationer?id=2144>
- European Commission (2012). *First European Survey on Language Competences. (Executive Summary)*. Hämtad 16 juli 2018 från http://ec.europa.eu/languages/library/studies/executive-summary-eslc_en.pdf

- European Association for Language Testing and Assessment/EALTA (2006). *EALTA Guidelines for Good Practice in Language Testing and Assessment*. Hämtad 16 juli 2018 från www.ealta.eu.org/guidelines.htm
- Gardner, J. (Ed.) (2012). *Assessment and Learning* (Second edition). London: Sage.
- Harlen, W. (2007). *Assessment of Learning*. London: Sage Publications Ltd.
- Harlen, W. (2012). On the Relationship Between Assessment for Formative and Summative Purposes. I Gardner, J. (red.), *Assessment and Learning* (second edition), s. 87–102. London: Sage Publications Ltd.
- Hymes, D.H. (1972). On communicative competence. In J.B. Pride & J. Holmes (Eds.), *Sociolinguistics*, 269–293. London: Penguin.
- Lindqvist, AK. (2012). *Datorbaserad bedömning i språk – forskningsläge, erfarenheter och implikationer*. Masteruppsats i ämnesdidaktik. Göteborg: Göteborgs universitet.
- Lodeiro, J. (2012). *Bedömning av muntlig reception*. En studie av utveckling av nationellt provmaterial i spanska. Masteruppsats i ämnesdidaktik. Göteborg: Göteborgs universitet.
- Messick, S. A. (1996). Validity and Washback in Language Testing. *Language Testing* 13, 241–256.
- Messick, S. A. (1989). Validity. I R. L. Linn (Ed.), *Educational Measurement* (3rd ed, pp. 13–103). New York: American Council on Education/Macmillan.
- Sadler, R. (1989). Formative Assessment and the design of instructional systems, *Instructional Science*, 18, 119–144.
- Scriven, M. (1967). The methodology of evaluation. I R. W. Tyler, R. M. Gagné, & M. Scriven (red.), *Perspectives of curriculum evaluation*, 39–83. Chicago, IL: Rand McNally.
- Shohamy, E. (2001). *The Power of Tests. A critical perspective on the uses of language tests*. Harlow, Essex: Pearson Education Limited.
- Skolverket (2011). *Kunskapsbedömning i skolan – praxis, begrepp, problem och möjligheter*. Stockholm: Skolverket. Hämtad 16 juli 2018 från <https://www.skolverket.se/publikationer?id=2660>
- Skolverket (2017). *Skolverkets systemramverk för nationella prov*. Stockholm: Skolverket. Hämtad 16 juli 2018 från <https://www.skolverket.se/publikationer?id=3890>
- Stobart, G. (2012). Validity in Formative Assessment. I Gardner, J. (Ed), *Assessment and Learning* (s. 233–242). London: Sage.
- SOU (2016:25). *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. Stockholm: Utbildningsdepartementet. Hämtad 16 juli 2018 från <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2016/03/sou-201625/>
- Takala, S., Erickson, G., Figueras, N. & Gustafsson, J-E. (2016). Future Prospects and Challenges in Language Assessments. I Tsagari, D. &

BEDÖMNING AV SPRÅKLIG KOMPETENS

- Banerjee, J. (eds.), *Contemporary Second Language Assessment, Contemporary Applied Linguistics* (s. 299–315). London: 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.
- Verhelst, N. (2013). *Comparison of national tests across time: some psychometric and statistical issues*. Intern rapport till Skolverket.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Bedömning i engelska för yngre åldrar

Ann Bjerklund Larsson & Carina Jochens

Den bedömning som brukar kallas ”formativ kunskapsbedömning”, eller ibland ”bedömning för lärande”, är en bedömning som äger rum under lärandets gång. En sådan bedömning kan göras både av lärare och av eleven själv och kan ha flera syften, som att få en uppfattning om hur långt eleven kommit i sin kunskapsutveckling och identifiera eventuella luckor i elevens lärande. Den kan också vara ett stöd för den fortsatta utvecklingen av kunskaper och planeringen av hur kunskapsinhämtandet ska ske.

I det följande presenteras utvecklingen av ett nationellt formativt material för bedömning i engelska för yngre åldrar. Inledningsvis ges en bakgrund till varför materialet kom till samt dess syfte. Därefter beskrivs arbetet med att ta fram materialet och innehållet presenteras övergripande. En något mer detaljerad beskrivning av de delar av materialet som berör produktion och interaktion samt reception ges därefter. Som avslutning visas hur materialet är en del av det system av nationella prov- och bedömningsstöd i engelska som finns i Sverige.

Bakgrund

2011 infördes en ny läroplan för grundskolan, vilket bl.a. innebar att det första nationella avstämningsstillfället av kunskapsnivån i engelska kom att skjutas fram ett år, från åk 5 till åk 6. Lärare som undervisar yngre åldrar fick kursplanens långsiktiga mål, centralt innehåll för åk 1–3 och 4–6 samt kunskapskrav för årskurs 6 att förhålla sig till vid bedömning av elevers språkfärdighet i engelska. Som stöd vid denna bedömning finns sedan vårterminen 2012 ämnesprov för årskurs 6 (Äp 6).

Eftersom den nya kursplanen inte kom att innehålla kunskapskrav i engelska för åk 3 och det därmed inte heller blev aktuellt med något ämnesprov för den årskursen, uppstod ett behov av ett bedömningsstöd i engelska för de yngre åldrarna. Vid Göteborgs universitet, Institutionen för pedagogik och specialpedagogik, inom projektet Nationella prov i främmande språk, utvecklades därför på Skolverkets uppdrag ett ***Nationellt diagnosmaterial i engelska för årskurs 1–6***. 2009 påbörjades arbetet med materialet och en första version i

ATT BEDÖMA SPRÅKLIG KOMPETENS

pappersformat levererades 2011 till alla skolor med åk 1–6. Ett par år senare utökades materialet med ett antal uppgifter och hela materialet finns idag på Skolverkets Bedömningsportal. (<https://bp.skolverket.se/>)

Syfte

Diagnosmaterialets främsta syfte var ursprungligen att

- vara underlag och stöd för lärares kontinuerliga bedömning av elevers språkliga kompetens i engelska.

Det skulle även kunna

- tydliggöra kursplanens kommunikativa och funktionella språksyn
- vara ett stöd vid tolkning av kursplanens innehåll
- åskådliggöra olika kunskapsnivåer
- vara ett stöd för elevers lärande
- bidra till lärares kompetensutveckling.

Detta har senare sammanfattats i följande tre punkter:

- att ge stöd för analys och bedömningen av elevens språkliga förmåga
- att tydliggöra elevens språkutveckling
- att bidra till att stödja elevens lärande

Materialet ska kunna integreras i lärande- och undervisningsprocessen och användas fritt under åk 1–6, för att ge en bred och varierad bild av elevernas språkutveckling samt underlag för fortsatt planering. I materialet presenteras bedömning som en naturlig del av lärande och undervisning, och det ska också ge eleverna möjlighet att visa vad de kan och utvecklas vidare. För läraren kan bedömningen även bli ett sätt att utvärdera sin egen undervisning.

Hur materialet arbetats fram

Grunden för arbetet med diagnosmaterialet har varit Lgr11 med kursplanen i engelska (Skolverket, 2011) och GERS, Gemensam europeisk referensram för språk (Andered, 2001; Europarådet/Skolverket, 2009). Materialet bygger på forskning och beprövad erfarenhet, och under utvecklingen av materialet har kontinuerliga möten med en referensgrupp hållits. Även specialpedagogisk och didaktisk expertis har tagit del i arbetet. Hur materialet har utformats bygger

också på de grundläggande principer som projektet Nationella prov i främmande språk har, och som ligger till grund för alla bedömningsmaterial som konstrueras inom projektet (<https://nafs.gu.se/information>). En sådan princip är att alla elevmaterial är enspråkiga. Detta gäller också för diagnosmaterialet. Eftersom eleverna endast förväntas använda engelska i de olika uppgifterna, skulle det bli motsägelsefullt om t.ex. instruktioner var på svenska. Vidare vore det besynnerligt att lägga in ett implicit krav på svenskkunskaper för att kunna visa sina kunskaper i engelska. Detta är en självklarhet i proven för de äldre eleverna, men hur skulle instruktioner på engelska fungera för de yngsta eleverna i årskurs 1 och 2? För att få svar på den frågan behövdes noggranna utprövningar för att ta reda på hur eleverna kunde instrueras på bästa sätt. Det visade sig att engelska på en mycket enkel nivå i kombination med bilder och kroppsspråk fungerade alldeles utmärkt. För de lärare som så önskar, ges förslag till instruktioner på engelska för varje uppgift, i samband med att genomförandet av uppgiften beskrivs. Enda undantaget från denna princip är kamratbedömningen, som förutom på engelska också finns på svenska.

Konstruktion

Ett omfattande konstruktionsarbete har gjorts och initialt fanns många faktorer och frågeställningar att ta hänsyn till. Först och främst var det viktigt att definiera vad som faktiskt skulle prövas och vilka svårighetsgrader uppgifterna skulle ha för att fylla de syften som Skolverket presenterat. En annan fråga var hur omfattande materialet skulle bli. Med kursplanen i engelska och GERS som stöd samt i dialog med Skolverket påbörjade en grupp provutvecklare tillsammans med aktiva lärare i engelska det konkreta arbetet med att konstruera uppgifter. Vid konstruktionsarbete av det här slaget är det viktigt att tänka på variation i såväl innehåll som svarsformat och att uppgifterna har en svårighetsmässig progression, samt är så valida och reliabla som möjligt. För att optimera uppgifternas kvalitet, gäller det som konstruktör bland annat att se till att innehållet är sådant att inte några grupper av elever får fördelar framför andra p.g.a. förkunskaper, kulturaspekter, kön osv. Vidare är det viktigt att de uppgifter som konstrueras verkligen prövar det som faktiskt ska prövas.

För att uppfylla materialets syften behövde inte bara elevuppgifter konstrueras utan också material för självbedömning och stöd för den fortsatta kunskapsutvecklingen. När det gällde självbedömningen var det naturligt att

knyta materialet till den redan befintliga Europeisk Språkportfolio (Europarådet/Skolverket, 2004). För att stödja såväl elever som lärare i det fortsatta arbetet med elevens kunskapsutveckling togs bland annat en kunskapsprofil fram som kan användas över tid och som syftar till att ge en så tydlig bild som möjligt av elevernas starka och svaga sidor i engelska.

Utprovning och analys

Konstruktionsarbetet ledde till att ett stort antal uppgifter utvecklades. Det var uppgifter som prövade olika färdigheter och på olika svårighetsnivåer. Många av de här uppgifterna valdes sedan ut och har varit föremål för storskaliga utprovningar vid skolor i hela landet. Eftersom det är en flexibel start för engelskundervisningen i Sverige, har det också varit viktigt att få deltagande elever som startat med engelska i olika årskurser vid utprovningarna av uppgifter för de yngre eleverna. Eftersom en tanke med materialet var att de olika uppgifterna skulle rikta sig till elever beroende på kunskapsnivå och inte primärt i vilken årskurs eleverna gick, beslutades att varje uppgift skulle prövas ut i åtminstone två olika årskurser. Det var också av intresse att försöka få reda på om elever efter några års engelskundervisning hade olika kunskapsnivåer beroende på i vilken årskurs undervisningen startat.

Ett mycket stort antal slumpvis valda elever från åk 1 till åk 6 och deras lärare har deltagit i utprovningarna. Deras synpunkter i elev- och lärarenkäter efter att ha prövat ut olika uppgifter har varit ovärderliga för utvecklingen av materialet. Allt material som kom in från de utprovande klasserna databehandlades och resultaten analyserades. Utprovningsresultaten visade bland annat de olika uppgifternas svårighetsgrad, och en lämplig rangordning kunde därför göras med avseende på materialets progression. Skillnaderna i resultat för eleverna i åk 4 kunde inte härledas till i vilken årskurs engelskundervisningen påbörjats, förutom i den skriftliga produktionen, där utprovningen visade att skillnaden var relativt stor. De elever som hade fått pröva på fri skriftlig produktion redan i åk 3 visade en större förmåga att uttrycka sig i skrift i åk 4.

Analysen av utprovningsresultaten i kombination med synpunkterna från elever och lärare låg sedan till grund för valet av de uppgifter som nu finns med i materialet. I det följande kommer frekventa synpunkter från utprovande lärare och elever att återges som direkta citat, i samband med att uppgifterna presenteras.

Innehåll

Diagnosmaterialet innehåller uppgifter som prövar såväl reception som produktion och interaktion. Här finns två uppgifter för respektive färdighet tala/samtala och skriva samt fem uppgifter vardera för färdigheterna lyssna och läsa, i spannet åk 1–6. Svårighetsgraden går från nybörjarnivå till nivå A2.1 i GERS, även om vissa uppgifter ger elever möjlighet att också visa att de har kommit längre än så i sin språkutveckling. A2.1 är den nivå som motsvarar den lägsta godtagbara nivån för betygssteget E i åk 6. I materialet finns också delar som fokuserar reflektion, medvetenhet och ansvar, som t.ex. självbedömning och kamratbedömning. I lärarmaterialet finns förutom instruktioner, didaktiska kommentarer och bedömningsanvisningar med exempel och kommentarer även underlag för dokumentation, bedömningsunderlag för muntliga och skriftliga uppgifter samt en kunskapsprofil.

Fokus: Produktion och interaktion

Den uppgift som fokuserar *mundlig kommunikativ förmåga* för de allra yngsta eleverna, är en samtalsmodell som bygger på en bild. Samtalet förs mellan läraren och en elev i taget. Läraren följer ett speciellt samtalschema med en inbyggd progression. Samtalsmodellen kan användas flera gånger under de första årens undervisning i engelska och ger då en tydlig bild av elevens språkutveckling. Till uppgiften finns en instruktionsfilm (i tre delar) för lärare. Filmen visar hur ett samtal kan gå till och det finns också några exempelsamtal som ljudfiler. Samtalsmodellen är utprövad i åk 1–4 och har som syfte, förutom att vara underlag för bedömning av elevers muntliga kommunikativa förmåga, även att ge den mycket unga eleven en positiv känsla av att tala engelska och göra honom/henne medveten om sina kunskaper. Samtalet blir också ett bra tillfälle för eleven att lära sig ytterligare engelska.

Lärare:

”Uppskattat, roligt, lärorikt!”

”Det händer aldrig att man sitter med ett barn och pratar engelska i vanliga fall, men det gav mig en hel del tycker jag. Upptäckte hur jag måste förändra mitt sätt att undervisa till viss del.”

För de lite äldre eleverna finns en modell för parsamtal (elev/elev) som är utprövad i åk 4–6.

ATT BEDÖMA SPRÅKLIG KOMPETENS

”Eleverna var mycket positiva. Roligt att få prata engelska och se att det fungerar – att man kan.”

”Jag märkte att jag måste jobba mer med att eleverna får ställa frågor.”

Elever:

”Det var ganska svårt att komma på vad jag skulle säga när jag dessutom behövde kunna säga det på engelska. Men det är bra att träna på att prata engelska, man lär sig bra på sådana här saker.”

Också *den skriftliga kommunikativa förmågan* fokuseras i två uppgifter. Den första, som är utprövad i åk 3 och 4, kan användas så fort eleverna börjar kunna formulera sig i skrift. Underlaget är en tankekarta och eleverna ges här tillfälle att skriva om det allra mest närliggande och vardagliga som rör dem själva. För en del av eleverna blev den här uppgiften det första de prövade på att skriva på egen hand. Det gällde framför allt de elever i åk 3 som startat med engelska i åk 3. Trots att de uppgav att det var ganska svårt, svarar ändå två tredjedelar av eleverna att det var en rolig uppgift. Att, som nedan, uppmana eleverna att skriva på svenska när de inte kommer på vad ett ord heter på engelska, är inte att rekommendera. Det är lätt att fastna i en sådan strategi och eleven kan få svårigheter att producera texter som är begripliga för en person med engelska som förstaspråk och som inte kan svenska. Bättre är då att från början förklara att det de skriver ska vara begripligt och att man kan försöka att skriva det man vill uttrycka på ett annat sätt.

Lärare:

”Vi har pratat engelska, så de blev lite förvånade när de skulle skriva om sig själva. Jag skriver som jag tror att orden stavas, sa många.”

”Jag sa till barnen att skriva ordet på svenska om de inte kunde eng. ordet.”

Elever:

”Jag tyckte att det var ganska svårt för jag kom inte på ett ord som jag skulle skriva på engelska. Det ordet var skolämne. Men annars var det roligt och man fick träna när man skrev på engelska.”

Uppgiften för de äldre eleverna kräver ett något större språkligt omfång. Som stöd för sitt skrivande får de förslag i punktform på vad de kan skriva om och det finns också bilder som kan ge inspiration. Elever i åk 5 och 6 har prövat ut uppgiften.

BEDÖMNING I ENGELSKA FÖR YNGRE ÅLDRAR

Lärare:

”Bra uppgift! Roligt att skriva om egna erfarenheter!”

Elever:

“It was a bit hard to come on words, to explain. It was fun and a good test.”

“It was fun but a little hard sometimes to find the right word. But that is good I think. Then you learn to describe a word with other words.”

Fokus: Reception

Eftersom *förmågan att lyssna och förstå* är den förmåga som utvecklas först vid språkinlärning, kan elevernas förmåga att förstå mycket enkel och tydligt talad engelska prövas tidigt. I materialet finns en hörförståelseuppgift som är lämplig för de allra yngsta eleverna. De får lyssna till en mycket kort och enkel saga. Underlaget är tecknade färgbilder och eleverna väljer en bild av fyra genom att markera med ett kryss. Uppgiften är utprövad i åk 2–4 och den var mycket uppskattad, särskilt bland eleverna i åk 2 och 3. Uppgiften är som helhet lätt, för de äldre eleverna mycket lätt; särskilt gäller det den första frågan, som nästan alla elever klarade. Syftet med en så lätt start är att de unga eleverna ska känna att de förstår vad de ska göra, att de kan det här och att det är roligt.

Lärare:

”Lite svårt på slutet. Det är många ord vi inte gått igenom.”

”Kul, det var spännande att lyssna och fylla i rätt.”

Elever:

”Det var jätteroligt. Jag lärde mig mer engelska.”

”Det var roliga bilder och lätt var det också! Jag vill göra mer!”

Det finns ytterligare en uppgift för de yngre eleverna. Det handlar då om korta dialoger och eleverna kan, förutom bildalternativen, även se några förklarande ord till varje bild. För elever som kommit längre i sin språkutveckling finns tre uppgifter med varierande svårighetsgrad och format. De består av dialoger, ett radioprogram med korta reportage och intervjuer samt en längre berättelse. Den första uppgiften består av flervalsuppgifter, medan de två senare har ett blandat svarsformat, d.v.s. att eleverna ibland väljer ett svarsalternativ av fyra och ibland svarar med egna ord. Elever i åk 5–6 har prövat ut uppgifterna.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Lärare:

”För att även de som har svårt för engelska ska få en möjlighet att lyckas så tycker jag att nivån var OK.”

”Positiva elever, ivriga att börja.”

Elever:

“Well, it was kind of easy, but not that boring.”

“It was a very easy test. I can very much english from my games.”

De fem uppgifter som fokuserar *förmågan att läsa och förstå* prövar läsförmågan på olika sätt. För de yngsta eleverna som än så länge mest uppfattar detaljer i texten baserat på det ordförråd de har, finns en uppgift som är utprovad i åk 3 och 4. Den utgörs av en anslagstavla med korta uppmaningar på lappar. Texten på varje lapp ska kombineras med en bild som illustrerar uppmaningen på lappen.

Lärare:

”De tyckte om provet. Tyckte att det var lätt att göra. De ville ha fler.”

”De tyckte att det var roligt att kombinera text och bild.”

Elever:

”Jag tycker att testet är lagom lätt. Vissa ord förstod man inte men man kunde lista ut de svåra orden med hjälp av orden man kunde.”

”Jag förstod alla orden rätt så bra men det var ändå lite klurigt men roligt.”

För att genomföra de andra läsförståelseuppgifterna på ett bra sätt behöver eleverna förstå innehåll och sammanhang. Uppgifterna, som är utprovade från åk 4 till åk 6 beroende på svårighetsgrad, består av definitioner av ord, sökläsning och läsning av berättande text. För en av uppgifterna är svarsformatet blandat, på samma sätt som beskrevs för hörförståelsen för de äldre eleverna. En annan uppgift består av en lucktext med flerval, medan de två övriga är s.k. matchinguppgifter där det gäller att söka i texten för att kunna kombinera rätt text med ett antal givna beskrivningar.

Lärare:

”Bra att det var uppdelat på flera texter. Texterna hade bra innehåll som barnen kunde relatera till.”

BEDÖMNING I ENGELSKA FÖR YNGRE ÅLDRAR

Elever:

“Thought it was a fun and learning test. It wasnt close to boring.”

“It was a difficult test and I got tired at the first page ... but it was still good.”

Avslutningsvis

Det nationella diagnosmaterialet i engelska för åk 1–6 inleder den röda tråd av nationella bedömningsstöd och provmaterial i engelska som löper genom grundskolan och vidare genom gymnasieskolan. Förhoppningen är att *Diagnosmaterial i engelska för årskurs 1–6* ska komma till användning ute i skolorna och bidra till att ge eleverna en god start i engelska och en positiv insikt om hur bedömning är en del av lärandet. När den första versionen av materialet kom till skolorna 2011, verkade få lärare ha fått vetskap om detta och responsen lät vänta på sig. Att materialet i utökad version numera finns på Skolverkets bedömningsportal (www.skolverket.se) har underlättat tillgängligheten för lärare i engelska. Eftersom diagnosmaterialet inte är obligatoriskt och Skolverket således inte gör någon insamling av resultat, så som görs med de nationella proven, är det svårt att veta i vilken omfattning materialet används. Den respons som nått oss har emellertid varit odelat positiv. Det tycks som om materialet fyller den funktion som är dess syfte för de lärare som väljer att använda det som en del av sin undervisning.

Här kommer till sist två slutord från elever i åk 3, som prövat ut en hörförståelseuppgift. De får illustrera det självförtroende, när det gäller engelska, som lyser igenom så många av de yngre elevernas kommentarer.

Det var inte så svårt
för att vi kan engelska

Det var roligt för här
jag blir stor ska jag plugga
på engelska

Nationella prov och bedömningsstöd i engelska

Diagnosmaterial åk 1–6
Ämnesprov åk 6
Diagnosmaterial 7–9
Ämnesprov åk 9
Kursprov Engelska 5
Kursprov Engelska 6
Bedömningsstöd
Engelska 7

Referenser

- Andered, B. (2001). Europarådets Framework – en inspirationskälla för de nya kursplanerna. I Skolverket *Språkboken. En antologi om språkundervisning och språkinlärning* (s. 26–37). Örebro: db Grafiska.
- Europarådet/Skolverket (2009). *Gemensam europeisk referensram för språk: lärande, undervisning, och bedömning*. Stockholm: Skolverket. Hämtad 2018-05-10 från <https://www.skolverket.se/publikationer?id=2144>
- Europarådet/Skolverket (2004). *Europeisk språkportfolio 6–11/12–16 år*. Stockholm: Skolverket. Hämtad 2018-05-10 från https://www.skolverket.se/skolutveckling/larande/sprak/europeisk_sprakportfolio_1.83490
- Kursplanen i engelska, Läroplan för grundskolan 2011. Stockholm: Skolverket. Hämtad 2018-05-10 från https://www.skolverket.se/laroplaner_ammn_och_kurser

Är det farligt att köra MC?

Håkan Schönberg & AnnaKarin Lindqvist

Olika typer av flervalsfrågor, som t.ex. *multiple choice-items* (ofta förkortat MC), har en central plats i många prov. Det förefaller därför viktigt att resonera kring vilka risker och möjligheter som kan finnas, och hur bästa möjliga kvalitet kan uppnås vid konstruktion och användning av MC. Ett prov syftar till att ta reda på vad provtagaren kan inom ett visst kunskapsområde och avser ge provtagaren möjlighet att få visa det. I vissa kunskapsprov är MC av flera anledningar ett praktiskt och lämpligt format.

Detta kapitel inleds med en översiktlig genomgång av relevant litteratur som behandlar aspekter av konstruktion och användning av MC, med fokus främst på följande frågor:

- A Hur många svarsalternativ bör man använda?
- B Vilken roll spelar det att provtagare kan gissa?
- C Vilka typer av kunskaper och färdigheter kan MC mäta?
- D Vad är viktigt att tänka på vid konstruktion av MC?

Kapitlet redovisar även resultat från en storskalig utprovning av hörförståelseuppgifter våren 2016 för ett eventuellt framtida nationellt prov i Engelska 5, där samma uppgift prövades ut i två versioner: en med fyra svarsalternativ och en med tre.

Begreppslig bakgrund

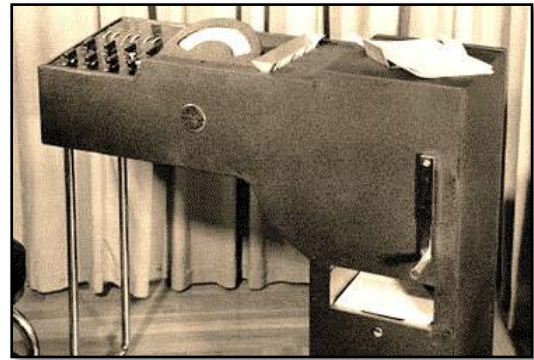
Multiple choice (MC) är förmodligen det vanligaste *selected response*-formatet (flerval) i storskaliga prov världen över och används även i de svenska nationella provmaterialen för språk. Det finns ett flertal varianter av flervalsformat, men i den här aktuella texten avses med MC den specifika uppgiftstyp som består av en fråga och ett antal svarsalternativ, av vilka provtagaren ska markera ett som det rätta genom att sätta ett kryss. Downing (2006) ger följande definition av *selected*

ATT BEDÖMA SPRÅKLIG KOMPETENS

response: "... examinees are required to choose an answer to a question or a statement from a listing of several possible answers" (s. 287).

MC som svarsformat introducerades i USA under första världskriget för användning i *Army Alpha*-testet i syfte att effektivt sortera och klassificera ett stort antal armérekryter som skulle matchas och utbildas för lämpliga uppgifter. Flervalsformatet visade sig lämpligare för ändamålet än öppna svarsformat där provtagaren själv ska formulera ett svar (Downing, 2006).

Ungefär 20 år senare introducerades maskinell rättning. Chappelle och Douglas (2006) nämner modellen *IBM 805 Test Scoring Machine* vilken lanserades redan 1935 och som hade kapacitet att rätta tio gånger snabbare än en människa. En erfaren operatör kunde rapportera upp till 1000 prov i timmen (16,7 prov/ minut) genom att släppa ned ett svarsblad i taget.



IBM Model 805 Test Scoring Machine
(IBM Corporation, 2002)

Lindblad (1990, s. 279ff) hänvisar till Spolsky (1976) och beskriver hur flervalsformatet efter andra världskriget introducerades i en svensk kontext:

”På vetenskapsteorins område var det *positivismen* som gällde och i fråga om mätningar givetvis *statistiken*. Det just sagda förklarar varför Spolsky kallar de nya tendenser som gradvis ökade i användning under 1950- och 1960-talen för *strukturellistisk-psykometriska*. I Sverige kan dessa tendenser sägas ha slagit igenom fullt ut i och med de nya centrala proven 1967 och, delvis, i de standardprov i engelska i årskurs 6 och 8 som introducerades i början av 70-talet.”

Lindblad noterar att ett typiskt inslag i de prov som konstruerades enligt dessa principer var att de byggdes upp av *discrete point items*, d.v.s. bedömdes som rätt eller fel (1 eller 0 poäng) och att detta tillät statistiska analyser. Dessa kunde i sin tur användas för att kontrollera och påverka provens mättillförlitlighet, reliabilitet. Lindblad (s. 281) konstaterar: ”Man kom därför att arbeta mycket med flervalsuppgifter, där i de flesta fall ett svar var rätt (eller i varje fall måste anses vara det bästa) och alla andra bedömdes som fel, även om graden av felaktighet kunde variera.”

A. Hur många alternativ bör man använda?

Rodriguez (2005) redovisar i en metaanalys hur forskare i över 90 år har försökt komma fram till lämpligt antal MC-alternativ. Den vanligaste frågan har varit huruvida de bör vara tre, fyra eller fem. Lee och Winke (2013) noterar att ett stort antal källor propagerar för tre svarsalternativ och menar att *“they are easier to write, have more effective distractors, and take less time to administer”* (s. 100).

Rodriguez (2005) för ett liknande resonemang: *“Using more options does little to improve item and test score statistics and typically results in implausible distractors”* (s. 11). Även Downing och Haladyna (1993) visar i en studie att tre svarsalternativ oftast är tillräckligt. Fler alternativ än så anses inte statistiskt funktionella, d.v.s. väljs inte av en tillräckligt stor procentandel provtagare. Downing (2006) konstaterar senare med viss förvåning att traditionen med antingen fyra eller fem svarsalternativ är stark, trots forskningsevidens.

Lee och Winke (2013) redovisar i en studie effekter av tre, fyra och fem svarsalternativ vid prövning av engelsk hörförståelse. Studien baseras på *College Scholastic Ability Tests* (CSAT) 2011, ett storskaligt *high stakes*-prov för högskoleintagning i Sydkorea (N=648 946). När studien genomfördes 2011 användes fem svarsalternativ i provet, som fram till 1993 hade haft fyra alternativ. En brett sammansatt grupp referenter och provtagare ombads ta bort det svarsalternativ de ansåg minst sannolikt som det rätta. Detta gjordes i tre olika prov i två omgångar för att skapa prov med respektive fem, fyra och tre svarsalternativ per fråga, vilket resulterade i nio olika prov (3x3). Dessa prov gjordes av närmare tre hundra studenter i tre försöksgrupper vilka gjorde tre av nio prov vardera med en veckas intervall. Grupperna testades diagnostiskt för jämförbarhet.

Lee och Winke menar att de observerade signifikant högre lösningsfrekvenser (medelvärden uttryckta i procent) för prov med tre svarsalternativ (d.v.s. de är lättare), vilket de anser ligga i linje med tidigare forskning. De fann däremot ingen signifikant skillnad när det gäller svårighetsgrad när de jämförde prov med fyra respektive fem alternativ, och kunde inte heller visa på något mönster avseende variationer i reliabilitetsdata mellan olika antal svarsalternativ och olika versioner. Lee och Winke spekulerar i huruvida de låga korrelationerna mellan svarsformat beror på att fler alternativ eventuellt kan involvera andra färdigheter än hörförståelse: *“... fewer options [...] may rely less on construct irrelevant skills such as testwiseness, L1 reading speed, and test anxiety.”* (s. 115ff). De för å andra sidan ett resonemang kring ett provs funktion och kontext –

provkonstruktörer behöver betänka statistiska, affektiva och kontextuella faktorer för att avgöra det optimala antalet svarsalternativ. Även små skillnader i provresultat kan behövas för att finmaskigt differentiera provtagares prestationer, och ett prov behöver finna acceptans bland alla inblandade. Ska provet resultera i ett flertal betygssteg eller bara ge godkänt/inte godkänt? Kan acceptans uppnås för prov med tre svarsalternativ om de – korrekt eller inte – uppfattas som lättare att gissa rätt på?

B. Vilken roll spelar det att provtagare kan gissa?

En kritik mot flervalssystemet grundar sig på att provtagaren förmodas lättare kunna gissa sig till rätt svar än vid öppna format. Detta har föranlett studier kring eventuella gissningseffekter på provresultat och reliabilitet. Wikström (2013) redovisar några argument och sammanfattar: ”Det tycks [...] råda enighet kring det faktum att gissning är ett problem och att risken för korrekt gissning ökar om frågorna är av bristande kvalitet” (s. 77f). Det framgår dock inte om det är storskaliga, utprovade prov eller klassrumsprov som avses. Enligt Lindblad (1990) bör lärare vara försiktiga med att använda flervalssystemet i egna prov och inte ha alltför stor tilltro till att dessa självklart skulle vara mer ’objektiva’ än prov med öppna svarsformat.

Downing (2006) påpekar att reliabilitetsdata skulle försvagas om slumpmässig gissning vore ett allvarligt problem. Han konstaterar att detta typiskt inte är fallet i väl utvecklade prov, framför allt inte i prov med ett stort antal items. Han menar att den statistiska sannolikheten i ett prov med 30 flervalssystemfrågor att få 70% rätt är mycket låg (0,0000356), vilket skulle tala emot att slumpmässig gissning i någon större utsträckning skulle kunna påverka ett provresultat. Lee och Winke (2013) refererar till Rodriguez (2005) som, liksom Haladyna (2004), avfärdar slumpmässig gissning som en statistiskt försumbar faktor och i stället diskuterar provtagares strategier: ”*Most test takers [...] do not blindly guess; rather, they make educated guesses*” (s. 102).

För att minska risken för att provtagare alltför enkelt ska kunna genomskåda felaktiga svarsalternativ menar Haladyna (2004) provkonstruktörer att använda rekommendationer för konstruktion av flervalssystem som tillkom under 1980-talet. Dessa är enligt Downing (2006) ”*based on a consensus of the educational measurement textbook authors and empirical research studies*” (s. 290). Rekommendationerna återfinns i förkortad form nedan.

C. Vilka typer av kunskaper och färdigheter kan MC mäta?

Den domän av kunskaper och färdigheter som avses mätas är av stor vikt vid val av format. En vanlig kritisk synpunkt kring flervalfrågor är att de mäter kunskap på lägre kognitiva nivåer, såsom enkla fakta och minneskunskaper, och att kunskaper på de högre kognitiva nivåerna hellre bör mätas på andra sätt (Wikström, 2013).

Downing (2006) hävdar å andra sidan att flervalformatet är det lämpligaste för att mäta kognitiv förmåga, framför allt på högre nivå, såsom problemlösning, syntes och utvärdering, och att det är mycket användbart och lämpligt för prov som avser inferenser till en vid domän av kunskaper, förmågor eller kognitiva färdigheter. En stor fördel som tas upp är att det flexibla och mångsidiga flervalformatet kan presenteras på alla nivåer av den kognitiva domänen. Detta gäller, enligt Downing, under förutsättning att provkonstruktionen genomförs på ett tillfredsställande sätt.

Haladyna (2004) tar upp två studier som jämför flervalformat med öppna svarsformat avseende de kognitiva nivåer som prövas och finner att det är en frågas utformning snarare än formattyp som styr dess kognitiva komplexitet. Downing (2006) refererar till Rodriguez (2003) som konstaterar att frågor med likvärdig komplexitet, oavsett flervalformat eller öppet svarsformat, uppvisar mycket hög korrelation. Downing hävdar att öppna svarsformat är mindre effektiva, ofta producerar mindre reliabla resultat och riskerar att täcka en för liten del av domänen, och konstaterar: "*Constructed response items should be used only for those content skill areas that cannot be measured by selected response items.*" (s. 288). Downing förespråkar dock användning av öppna svarsformat vid prövning av t.ex. skriftlig produktion och kognitiv förmåga med hög komplexitets- och inferensnivå. Haladyna å sin sida rekommenderar varierade svarsformat för att optimera mätning av olikartade innehåll och kognitiva processer, eftersom de öppna svarsformat som finns för att testa högre kognitiva nivåer är fler än flervalformaten.

D. Vad är viktigt att tänka på vid konstruktion av MC?

Såväl Downing (2006) som Haladyna (2004) återger de rekommendationer för konstruktion av MC som tidigare nämnts, och understryker att kritiken mot flervalssystemet oftast handlar om att det skulle generera bedömning av trivial kunskap på låg nivå; viss forskningsevidens tyder dock på att problemet inte ligger i själva flervalssystemet, utan i provkonstruktion som inte kvalitetssäkrats. Här följer en förkortad version av dessa rekommendationer.

- Undvik triviale innehåll.
- Se till att innehållet i varje item hålls oberoende av och inte överlappar med andra.
- Undvik items baserade på åsikter, eller börja med till exempel "According to ..."
- Ställ upp items vertikalt.
- Använd förkortningar sparsamt.
- Använd enklast möjliga språk så att läsförståelse inte interfererar med det som ska provas.
- Undvik onödig mångordighet för att minimera lästid, både i stam och alternativ, utan att detta inverkar på innehållet som provas och de kognitiva krav som ställs.
- Se till att anvisningarna i stammen är mycket klara.
- Undvik negationer i stammen. Om negation används, använd kursiv eller fet stil – negationer kräver större arbetsminne hos provtagaren.
- Se till att bara ett svar är rätt.
- Gör distraktörerna plausibla, gärna i sig korrekta påståenden, men felaktiga svar på frågan.
- Variera plats för rätt svar och balansera facit så att rätt svar återfinns lika ofta på de olika positionerna.
- Gör alternativen ungefär lika långa.
- Placera alternativen i logisk eller numerisk ordning, till exempel alfabetiskt eller med hänsyn till hur långa de är (kortast till längst eller vice versa).
- Konstruera innehållsligt och grammatiskt homogena alternativ.
- Undvik att ge ledtrådar till rätt svar genom till exempel grammatisk inkonsekvens, iögonfallande rätt svar samt par eller tripplar av alternativ.
- Använd vanliga felaktiga elevsvar på öppna frågor för att skapa distraktörer.
- Undvik humor i high-stakes-prov.

(Downing 2006, s. 294; Haladyna 2004, s. 99f)

Utprovning av en hörförståelseuppgift i två versioner

De nationella proven i engelska i Sverige är s.k. *proficiency tests*, d.v.s. de syftar till att pröva och ge en bild av elevernas allmänna språkfärdighet i engelska oberoende av var, när och hur denna färdighet uppnåtts, och genererar resultat i sex betygssteg. En del av provet avser pröva reception med fokus på hörförståelse, d.v.s. förståelse av talad engelska samt förmåga att tolka innehållet och därmed förmåga att lyssna på olika sätt och med olika syften. Elevernas förmåga inom olika områden provas oftast i integrerade former. Uppgifterna i hörförståelse kräver således, förutom hörförståelse, att eleverna också för att kunna besvara frågorna kan läsa och skriva på engelska. En grundläggande princip för proven är att eleverna ska ges möjlighet att visa så mycket som möjligt av vad de kan – inte primärt vad de inte kan. Detta innebär bland annat att provets olika delar är av skiftande slag, vad gäller till exempel svarsform.

Se http://nafs.gu.se/prov_engelska/engelska_gymn/bedomning

För att kvalitetssäkra nationella provmaterial och uppnå god tillförlitlighet är storskaliga utprovningar avgörande. Många elever ($N \approx 500$) och lärare vid slumpvis valda skolor genomför uppgifterna under så provliknande förhållanden som möjligt, granskar och kommenterar dem. Den information som då erhålls används vid bl.a. provsammansättning och fastställande av bedömningsanvisningar och betygsgränser.

Om uppgiften

För att kunna studera hur flervalsfrågor med tre eller fyra svarsalternativ fungerar, beslöts att en hörförståelseuppgift skulle provas ut i två versioner: en med tre svarsalternativ och en med fyra. Uppgiften innehöll sex frågor av flervalsformat och åtta med öppet svarsformat. Nedan redovisas några resultat av denna studie.

Uppgiften miniutprovades vid två tillfällen 2015 och utprovades därefter storskaligt 2016. För att kunna ta bort ett svarsalternativ användes resultatet av den sista miniutprovningen, och det minst frekvent valda, felaktiga svarsalternativet togs i en version bort inför storskalig utprovning. Läs- och svars pauser var lika långa i båda versionerna.

Bakgrundsdata om utprovningen

Båda utprovningsgrupperna uppvisade jämn könsfördelning. I gruppen som gavs tre svarsalternativ (här kallad **G3**) fanns något fler elever på yrkesförbere-

ATT BEDÖMA SPRÅKLIG KOMPETENS

dande program, och i gruppen som gavs fyra alternativ (**G4**) något fler på studieförberedande program. Samtliga utvärkningsgrupper genomför en s.k. ankaruppgift som syftar till att ge en indikation om gruppens färdighetsnivå. Enligt resultaten på denna ankaruppgift ligger G3 på en genomsnittlig färdighetsnivå (.63) och G4 över genomsnittet (.68).

Möjligheten att pröva ut den här aktuella uppgiften i två versioner visade sig relativt sent under planeringen av utvärkningsgruppen och kunde därför inte styra dess design. Detta medförde till exempel att gruppen som gavs tre svarsalternativ (G3) fick göra uppgiften som sista uppgift av tre (totalt 38 minuter), medan gruppen som gavs fyra alternativ (G4) fick göra uppgiften som första uppgift av två (totalt 31 minuter). Utöver gruppernas inbördes olikheter i fråga om programtillhörighet och färdighetsnivå (enligt ovan nämnda ankaruppgift), fanns alltså flera faktorer som försvårar jämförelser – grupperna fick i praktiken olika lång tid för att besvara frågorna eftersom läs- och svars pauserna var lika långa i båda versionerna, och uppgifterna gjordes i olika positioner och kontexter. I ljuset av detta bör resultaten som redovisas nedan tolkas med försiktighet.

De båda grupperna gav vid en s.k. *test-taker feedback* något olika respons på uppgiften. Samtliga elever som deltar i utvärkning ombeds fylla i en enkät för varje uppgift, dels med en Likert-skala (Yes 1–No 5), dels genom öppna kommentarer. Här redovisas endast medelvärden enligt Likert-skalan.

	G4	G3
1. I think the task was good	2.71	2.71
2. I think the task was easy	3.13	3.16
3. I think the task tested something important	2.76	2.58
4. I think I did well on this task	3.06	3.02
5. I think the task gave a fair picture of how well I understand English	3.05	2.78
6. There was enough time for the task	2.46	2.13

Tabellen till höger visar att G3 och G4 skattade uppgiften mycket lika avseende fråga 1, 2 och 4, *Good, Easy* och *I did well*.

Vissa skillnader finns däremot i svaren på fråga 6 (*Enough time*), 5 (*Fair picture*) och 3 (*Tested something important*). G4 anger en något högre grad av tidsbrist, i något lägre grad att uppgiften gav en rättvis bild av deras kunskaper och i något lägre grad att den prövade något viktigt. Denna skillnad i acceptans kan eventuellt härledas till det faktum att grupperna, som nämndes ovan, hade olika tidsramar för att läsa och besvara frågorna, eftersom G4 hade fyra svarsalternativ att ta ställning till. Även här bör dock resultaten tolkas med försiktighet.

ÄR DET FARLIGT ATT KÖRA MC?

Hur gick det?

Ett provs reliabilitet, eller tillförlitlighet, handlar om hur ”rättvis” och icke slumpmässig bedömningen är, avseende t.ex. jämförbarhet och likvärdighet (Erickson, 2009). Ett vanligt reliabilitetsmått är *Cronbach's Alpha* (Cronbach, 1951). Alpha bör ligga så nära 1 som möjligt för att uppgiften ska anses som fungerande. Tabellen till höger visar att både Alpha och lösningsfrekvens är något högre i G3. Skillnaden avseende lösningsfrekvens grupperna emellan är större vid MC-frågor. Detta indikerar, också med tanke på resultaten på ankaruppgiften som nämnts ovan, att tre svarsalternativ gjorde uppgiften lättare.

Resultat

	G4	G3
<i>Cronbach's Alpha</i>	.732	.765
Lf*, hela uppgiften	.61	.66
Lf, 6 MC-frågor	.54	.63

*Lf: Lösningsfrekvens, %

Här följer en redovisning av data rörande samtliga sex flervalsfrågor i den aktuella uppgiften samt några jämförelser mellan resultaten för tre respektive fyra svarsalternativ (G3 respektive G4).

I tabellerna nedan anges för varje fråga:

- poäng (0/1) för det givna svaret (**Pg**)
- medelpoäng på hela uppgiften (max 15p) för elever som valt ett visst svarsalternativ (**Medel**)
- svarsalternativ (**Alt.**)
- andelen elevsvar i procent per svarsalternativ (%), inklusive hur många som avstått från att svara (\emptyset)

Andelen elevsvar för det rätta alternativet benämns också frågans lösningsfrekvens. När det gäller att gå från fyra till tre svarsalternativ har, som tidigare nämnts, principen varit att ta bort den distraktor som dragit minst i miniutprovning inför utprovningen i G3.

Flervalsfråga 2 och 11

Gemensamt för fråga 2 och 11 är att de inte blev nämnvärt lättare när ett svarsalternativ togs bort. Elevsvarens spridning mellan distraktorerna uppvisar inte heller några större skillnader. Endast få elever har låtit bli att kryssa för något alternativ (\emptyset), och den distraktor som tagits bort till G3 attraherar endast i liten utsträckning i G4.

ATT BEDÖMA SPRÅKLIG KOMPETENS

G4

Pg	Medel	Alt.	%
----	-------	------	---

0	3.85	Ø	2.4
0	6.81	A	3.7
1	10.58	B	60.7
0	7.83	C	2.0
0	7.40	D	30.5
0	4.50	BC/CD	0.7

Fråga 2

G3

Pg	Medel	Alt.	%
----	-------	------	---

0	5.00	Ø	2.7
1	11.29	A	62.4
0	7.78	B	4.2
0	7.92	C	30.6

(Överstrukna distraktorer i G4 togs bort i G3.)

0	4.00	Ø	1.0
0	6.35	A	5.8
0	5.30	B	10.2
1	10.16	C	76.9
0	7.29	D	5.8
0	8.00	BC	0.3

Fråga 11

0	0.00	Ø	0.9
0	5.80	A	10.9
1	10.82	B	78.5
0	8.34	C	9.7

Flervalsfråga 7

I item 7 placerades distraktorn i alternativ C (G4) som A i G3, i enlighet med rekommendationer ovan om alternativens placering i relation till inbördes längd. Distraktorn attraherade något mer som C med fyra alternativ än som A med tre. Svarsmönstret är annars likartat. Primära och sekundära distraktorer tillsammans attraherar ungefär lika mycket i båda grupperna. Frågans lösningsfrekvens stiger däremot anmärkningsvärt när frågan har tre jämfört med fyra svarsalternativ. Detta kan bero på att distraktorn som togs bort attraherar en relativt stor andel elever som eventuellt har valt rätt alternativ i G3. Endast få elever avstod från att svara (Ø).

G4

Pg	Medel	Alt.	%
----	-------	------	---

0	6.66	Ø	1.0
0	7.86	C	29.2
0	8.78	B	15.6
0	7.90	A	10.8
1	10.66	D	43.4

Fråga 7

G3

Pg	Medel	Alt.	%
----	-------	------	---

0	5.44	Ø	2.7
0	8.98	A	26.4
0	8.31	B	13.3
1	10.96	C	57.6

(Överstruken distraktor i G4 togs bort i G3.)

ÄR DET FARLIGT ATT KÖRA MC?

Flervalsfråga 10

Item 10 uppvisar relativt stora skillnader. Lösningfrekvensen stiger anmärkningsvärt när uppgiften ges med tre svarsalternativ. Den primära distraktorn (A) attraherar färre elever i G3. Den sekundära distraktorn ligger däremot kvar på ungefär samma nivå, och den minst attraktiva distraktorn (bara i G4) väljs av mycket få elever. Andelen \emptyset är låg, men något högre i G4.

G4

Pg	Medel	Alt.	%
0	7.00	\emptyset	4.7
0	8.91	A	46.1
0	7.60	B	1.7
1	10.34	C	36.3
0	7.81	D	10.8
0	13.00	CD	0.3

Fråga 10

G3

Pg	Medel	Alt.	%
0	4.10	\emptyset	3.0
0	9.64	A	34.2
1	10.74	B	53.0
0	8.40	C	9.7

(Överstruken distraktor i G4 togs bort i G3.)

Flervalsfråga 6 och 14

I item 6 och 14 attraherar distraktorerna på likartat sätt. Lösningfrekvenserna är även för dessa items högre i G3 än i G4. Andelen \emptyset för item 6 i G4 är något högre än i G3, och förhållandevis många relativt högpresterande elever har valt att inte svara.

G4

Pg	Medel	Alt.	%
0	8.70	\emptyset	5.8
1	10.58	A	44.4
0	8.21	B	22.4
0	7.88	C	8.5
0	7.94	D	19.0

Fråga 6

G3

Pg	Medel	Alt.	%
0	5.22	\emptyset	2.7
1	11.24	A	56.1
0	8.47	B	26.1
0	8.46	C	15.2

G4

Pg	Medel	Alt.	%
0	3.25	\emptyset	1.4
1	10.00	A	65.1
0	8.10	B	19.3
0	7.50	C	8.8
0	7.93	D	5.4

Fråga 14

Pg	Medel	Alt.	%
0	2.66	\emptyset	1.8
1	10.77	A	70.6
0	8.55	B	18.5
0	7.73	C	9.1

(Överstrukna distraktorer i G4 togs bort i G3.)

Samtliga sex flervalsfrågor

Resultaten på ankaruppgiften tyder på att gruppen G4 befinner sig på en högre färdighetsnivå. Ändå är lösningsfrekvenserna högre i G3 än i G4 (Hela uppgiften: .66/.61, MC: .63/.54). Tre svarsalternativ tycks alltså ge ett lättare prov än fyra, vilket torde upplevas som logiskt. Inget i studien tyder på att tre alternativ skulle påverka reliabiliteten negativt, snarare tvärtom; för hela uppgiften var Alpha högre i G3. (I den här redovisade uppgiften hade frågor med öppet svarsformat för övrigt generellt högre Alpha än flervalsfrågorna.) Noteras kan också att distraktorerna som togs bort i G3 attraherade få elever i G4, utom i item 7 och i någon mån item 6.

Cronbach's Alpha If Item Deleted (här förkortat CAID) visar hela uppgiftens reliabilitetsvärde (*Cronbach's Alpha ...*) om det aktuella itemet tas bort (*... if item deleted*). Alltså: om Alpha går upp om ett visst item tas bort betyder det att detta item 'sänker' hela uppgiftens reliabilitetsvärde, d.v.s. detta item fungerar förmodligen inte optimalt och bidrar därmed inte till uppgiftens kvalitet.

När det gäller de sex flervalsfrågorna som ingick i uppgiften tycks item 2 och 11 vara de två mest reliabla av flervalsfrågorna i uppgiften enligt CAID-värdena i tabellerna nedan, och item 10 den minst reliabla.

G4			G3		
Hela uppgiften:			Hela uppgiften:		
Lf (%)	.61		Lf (%)	.66	
Alpha	.732		Alpha	.765	
MC nr	Lf (%)	CAID	MC nr	Lf (%)	CAID
2	.61	.711	2	.62	.750
6	.44	.730	6	.56	.759
7	.43	.728	7	.58	.767
10	.36	.741	10	.53	.777
11	.77	.710	11	.78	.750
14	.65	.733	14	.71	.762

Lf = Lösningfrekvens, %
CAID = Cronbach's Alpha if Item Deleted

Det är värt att understryka att de jämförelser som här görs mellan utprövningsgrupperna påverkas av flera faktorer som redan har berörts: utprövningsgrupperna befinner sig på olika färdighetsnivåer, uppgiften placerades på olika positioner i versionerna, utprövningsgrupperna hade i praktiken olika lång tid

ÄR DET FARLIGT ATT KÖRA MC?

på sig för att besvara flervalsfrågorna (eftersom G4 hade mer att läsa), och uppgiften blir olika svår i de två grupperna. Dessutom är det rimligt att tänka sig att de olika MC-items som ingick i uppgiften prövar olika sorters hörförståelse, även om de strikt formatmässigt är flervalsfrågor.

Det är därför lämpligt att vara försiktig med långtgående slutsatser avseende t.ex. huruvida flervalsfrågor är att föredra eller inte, hur många svarsalternativ provtagare ska erbjudas och hur svår eller lätt en uppgift blir med ett givet antal svarsalternativ.

Frågor och reflektioner

En studie som denna, avseende en jämförelse mellan tre eller fyra distraktorer i flervalsfrågor, väcker naturligtvis ett antal följdfrågor, som till exempel:

- Treval vid flervalsfrågor av MC-format verkar generellt inte påverka uppgiftens reliabilitet negativt. Är det då självklart att använda tre distraktorer överlag? Eller hur många svarsalternativ bör egentligen användas? Bör till exempel läs- och hörförståelse hanteras olika i detta avseende?
- Treval verkar göra uppgiften lättare. Riskerar MC-frågor med fler alternativ att störas mer eller mindre än MC med färre alternativ eller öppna frågor, av irrelevanta faktorer eller förmågor? Hur mycket spelar andra faktorer än de som avses prövas in när provtagare ställs inför en flervalsfråga, såsom t.ex. läshastighet, minneskapacitet, *testwiseness* och provnervositet?
- Som argument för prov med flervalsfrågor, framförs ibland att dessa är mer 'objektiva' än prov med öppna svarsalternativ. Är det så? Lindblad (1990) påpekar att det 'objektiva' i flervalsprov handlar om bedömningen – vad som ingår i provet och hur uppgifterna utformas är däremot subjektivt: "Detta förbises ibland som en följd av ordets makt över tanken." (s. 281) Ett liknande resonemang förs i Gemensam europeisk referensram för språk (Council of Europe/Skolverket, 2009, s. 186).
- Inför den omfattande digitaliseringen av de nationella proven i Sverige, som föreslås vara utbyggd 2022, är det inte orimligt att tänka sig att MC och andra flervalsformat blir populära eftersom de kan rättas maskinellt. Önskar då de som förespråkar effektiva prov med 'central' rättning att vi går tillbaka till en uppgraderad variant av 1930-talets armétester i flervalsformat och en centralt placerad *IBM Model 805 Test Scoring Machine*, som beskrevs i början av kapitlet? Kommer vi då att pröva det lätt mätbara

och 'objektiva' och väljer vi då bort domäner och färdigheter som inbegriper mer tidskrävande bedömningar?

- Det vore intressant att göra en liknande studie, med högre jämförbarhet avseende utprövningsgrupper, design och genomförande. Skulle en sådan studie ge liknande resultat som denna?

Överväganden kring vilka frågeformat som är lämpliga i en provuppgift bör utgå ifrån provets övergripande syfte, vilken typ av prov det är samt vad det avser pröva och indikera genom resultat. Denna studie utgick från en uppgift som avser pröva hörförståelse och som eventuellt kommer att ingå i ett nationellt, betygsstödande prov av *proficiency*-typ för gymnasieskolan. För att eleven ska ges goda möjligheter att visa sin förmåga varierar uppgiftstyper och skrivrespektive läsmängd hålls nere. I denna kontext förefaller MC med treval vara ett gott alternativ till öppna frågor.

Vid konstruktion av MC med tre distraktorer valdes att utgå från en version av uppgiften med fyra distraktorer och att, efter en mindre utprövning, stryka ett alternativ. Detta föll väl ut och kan vara ett led i att kvalitetssäkra de alternativ som behålls i en trevalsfråga. Vilka distraktorer som tas bort, och på vilka grunder detta görs, är emellertid utan tvekan en mångfasetterad process som behöver diskuteras, dokumenteras och motiveras. Det kan vara värt att notera att i den här aktuella uppgiften var frågor med öppet svarsformat överlag mer reliabla och hade överlag högre lösningsfrekvens än flervalsfrågorna, i båda utprövningsgrupperna.

Enligt Downing (2006) är utbildning av provutvecklare en betydande investering: "... *writing effective selected-response items is a non-trivial skill, requiring a great deal of costly professional time and effort from content experts*" (s. 292). Han noterar även: "*Per item costs are variable, but it is not unreasonable for a very high-stakes testing program to value its items at more than USD 1,000 per item.*" Thorndike (1967) beskriver itemkonstruktion som den förmodligen mest krävande tänkbara sortens kreativt skrivande: "*Not only must the item writer understand content measured by the item but must determine whether the cognitive demand will involve recall, understand, or application. Originality and clarity are key features of well written test items.*" (Haladyna, 2004, s. 65).

För att avslutningsvis återknyta till rubriken till detta kapitel, *Är det farligt att köra MC?*, skulle en slutsats kunna vara att konstruktionen behöver vara robust, säker och funktionell om vi planerar att köra MC. Detta gäller även underlaget: noggrann bearbetning och provkörning med miniutprövningar ger värdefulla

ÄR DET FARLIGT ATT KÖRA MC?

data om MC:ns egenskaper och kvaliteter, såsom dess stabilitet och tillförlitlighet. Provkonstruktörer bör också följa etablerade rekommendationer, eller trafikregler om så önskas, för att köra MC så säkert som möjligt.

Så här beskriver Downing (2006) konsten att köra MC:

"Item writing is both art and science. There are scientifically sound principles of item writing, but the creation of effective multiple-choice items requires the skillful application of these principles to the content to be tested. That is the art."

(s. 293)

Referenser

- Chapelle, C. A. & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press.
- Council of Europe/Skolverket. (2009). *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. Stockholm: Skolverket.
www.skolverket.se > Om Skolverket > Publikationer.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. I *Psychometrika*, September 1951, Volume 16, Issue 3, pp 297–334.
- Downing, S. M. (2006). Selected-Response Item Formats in Test Development. I S. M. Downing & T. M. Haladyna (Red.), *Handbook of Test Development* (s. 287–301). Mahwah: Lawrence Erlbaum Associates, Inc.
- Erickson, G. (2009). ”Att bäras åt” – Om den goda bedömningens flerfaldighet och ömsesidighet. I U. Tornberg, et al. *Språkdidaktiska perspektiv. Om lärande och undervisning i främmande språk* (s. 159–174). Stockholm: Liber.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. New York: Routledge Taylor & Francis Group.
- Lee, H. S. & Winke, P. (2013). The differences among three-, four-, and five-option item formats in the context of a high-stakes English language listening test. I G. Fulcher & A. Ginther (Red.), *Language Testing*, 30,(1) (s. 99–123). Singapore: SAGE Publications Ltd.
- Lindblad, T. (1990). Prov och bedömning i främmande språk. I K. Thorsén (Red.), *Undervisning i främmande språk. Kommentarmaterial Lgr 80*. (s. 274–298). Stockholm: Skolöverstyrelsen och Utbildningsförlaget.
- Projektet Nationella prov i främmande språk. Engelska 5 och 6. Bedömning.
http://nafs.gu.se/prov_engelska/engelska_gymn/bedomning
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. I *Educational Measurement: Issues and Practice*. 24(2), 3–13. Philadelphia: National Council on Measurement in Education.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Studentlitteratur. (1972). *Konsten att göra flervalsfrågor*. I allt väsentligt är innehållet hämtat från Bengt Ramund: Hur ska man fråga? – en artikel i *Undervisningsteknologi*, 1967(8). Lund: Studentlitteratur.

Wikström, C. (2013). *Konsten att göra bra prov – vad lärare behöver veta om kunskapsmätning*. Stockholm: Natur & Kultur.



Tre texter om test-taker feedback från utprövningar i moderna språk

Bedömningsstöden i moderna språk (franska, spanska och tyska) utvecklas av projektet Nationella prov i främmande språk (NAFS) vid Göteborgs Universitet på uppdrag av Skolverket och enligt samma principer som de nationella proven i engelska. Det tar mellan 1,5 och 2 år att konstruera ett prov och konstruktionen bygger dels på beprövad erfarenhet, dels på forskning och storskaliga utprövningar av provuppgifter. En nödvändig och viktig del i processen för att säkerställa hög kvalitet är att skolor runt om i landet medverkar i dessa utprövningar.

En nationell utprövning av provuppgifter sker varje vår på slumpvis valda skolor och resultaten från denna ligger sedan till grund för valet av uppgifter till framtida provmaterial i moderna språk. Såväl produktiva och interaktiva färdigheter (tala och skriva) som receptiva (hör- och läsförståelse) prövas ut och upp till 400 elever genomför varje provuppgift. Av praktiska skäl är dock volymen mindre för utprövningen av muntlig färdighet. De receptiva provuppgifterna har en variation av svarsformat, vanligen flervalsfrågor (3-val eller 4-val), öppna frågor (eleven svarar med ett eller flera egna ord), matchning (ett antal korta texter paras ihop med rätt rubrik eller svar) och lucktexter av olika slag (antingen där eleven själv ska skriva ett ord i luckan eller luckor med flervalsalternativ). Även andra format kan förekomma.

I samband med att elever och lärare genomför utprövningar ombeds de granska och kommentera provuppgifterna. Förutom upplysningar om validitet, reliabilitet och svårighetsgrad, det senare bl.a. genom så kallat P-värde (medelvärdet uttryckt i procent, d.v.s. i vilken utsträckning eleverna har klarat att lösa provuppgifterna), är svaren från elevenkäter (s.k. test-taker feedback) och lärarenkäter med kommentarer kring bl.a. upplevd svårighetsgrad, innehåll och allmänna synpunkter viktiga redskap i utvecklingen av kursprov. Elever och lärare ger sin respons i enkäterna dels genom att markera sina svar på femgradiga Likertskalor, dels genom att skriva kommentarer i ett öppet fält. Elevenkäten återfinns i samma häfte som uppgifterna som prövas ut och besvaras i direkt anslutning till genomförandet; bortfallet blir därför försumbart.

Efter att utprövningsmaterialet inkommit vidtar olika typer av analyser. Enkätsvaren sammanställs och analyseras tillsammans med övrigt statistiskt underlag och utgör en viktig del av validitetsarbetet i den fortsatta processen fram till provets färdigställande. Sett ur ett internationellt perspektiv är det ovanligt att elevers kommentarer i så stor utsträckning och på ett så systematiskt sätt ligger till grund för utvecklingen av nationella provmaterial i främmande språk på samma sätt som i Sverige.

Nedan följer tre texter där informationen från elevenkäter till ett antal storskaliga utprövningar i moderna språk sammanställs och analyseras utifrån olika perspektiv. *Was piensan les élèves? – Test-taker feedback i samband med utprövningar i tyska, spanska och franska* (Sebestyén & Albinsson) är en jämförande studie av elevenkäter och resultat från tre olika utprövningar av läsförståelse. Därefter följer *Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ* (Finndahl & Perrotte), som behandlar elevers synpunkter på och upplevelse av utprövningar i reception samt produktion och interaktion i franska. Avslutningsvis analyseras attityder, utprövningsresultat och självskattningar från elevenkäter till utprövningar av spanska receptiva provuppgifter i texten *Motiverad för uppgiften? – Test-taker feedback ur ett genusperspektiv* (Hedenbratt & Axelson). Avsnittet avslutas med en gemensam referenslista för alla tre texterna.

Was piensan les élèves?

– Test-taker feedback i samband med utprövningar i tyska, spanska och franska

Åsa Sebestyén & Hans Albinsson

Inledning

Sedan början av 2000-talet har olika elev- och lärarenkäter funnits med som en viktig del av de storskaliga utprövningarna i moderna språk. Enkäterna innehåller ett antal påståenden med tillhörande så kallade Likertskalor samt plats för öppna kommentarer. Hur många steg som ska finnas i en Likertskala är en omdiskuterad fråga. I Nafs-projektets enkäter har man valt att använda fem steg, för att genom ett mittalternativ ge respondenterna möjlighet att förhålla sig neutrala till ett påstående.

Eleverna besvarar enkäten i direkt anslutning till utprövningstillfället och nya uppgifter provas ut varje år. Detta innebär att provtagarnas upplevelser av och synpunkter på provuppgifterna systematiskt samlas in och tas i beaktande som en naturlig del av validitetsarbetet vid sammansättningen av kursprov. Värdet av att samla in synpunkter från provtagare har poängterats av bland andra Alderson, Clapham och Wall (1995), Bachman och Palmer (1996) samt Erickson (1996; 2010), inte minst från kvalitetssynpunkt och relaterat till den utökade syn på validitet, som uttrycks bl.a. av Messick (1989), och som innefattar både användning och konsekvenser av bedömningar. Ett annat av utprövningens huvudsyften är att testa provuppgifternas reliabilitet, spridning och svårighetsgrad (det senare redovisat i form av P-värden, eller lösningsfrekvenser i procent, där ett högt P-värde innebär att uppgiften var ”lätt”).

Underlaget för denna studie utgörs av enkätsvar och lösningsfrekvenser (P-värden) från de senaste utprövningarna av läsförståelseuppgifter för Moderna språk steg 4 (vanligen motsvarande åk 2 i gymnasieskolan), dvs. utprövningarna i franska (2014), spanska (2015) och tyska (2016). Inom begreppet ”läsförståelseuppgifter” ryms texter med tillhörande frågor av olika format, vanligen flervalsfrågor, frågor med fria (öppna) svar (ibland förekommer flerval och

ATT BEDÖMA SPRÅKLIG KOMPETENS

öppna svar i kombination), matchningsuppgifter (ett antal texter som ska paras ihop med rätt rubriker eller svar) och olika typer av lucktexter.

I denna text görs en jämförelse mellan elevsynpunkter från enkäterna till de olika provuppgifterna beträffande upplevd svårighetsgrad, skattning av egna resultat och den ”faktiska” svårighetsgraden, dvs. P-värdet. Dessutom undersöks vilka provuppgifter som enligt respondenterna har ett intressant innehåll.

Följande frågeställningar ligger till grund för undersökningen:

Kan man skönja ...

- några allmänna tendenser i enkätsvaren kring hur eleverna uppfattar olika svarsformat eller uppgiftstyper och deras svårighetsgrad?
- något samband mellan elevernas skattning av sina egna resultat och det faktiska utfallet, dvs. P-värdet?
- skillnader mellan de tre språken franska, spanska och tyska när det gäller elevernas uppfattning av provuppgifterna?

De tre påståenden från elevenkäterna som har valts ut för analys är de som var innehållsmässigt gemensamma för de tre språken och där svaren därför kan anses vara jämförbara. Som tidigare nämnts uppmanas eleverna att markera sina svar i enkäten på en Likertskala (1–5) där 1 betyder *Ja, absolut* och 5 *Nej, absolut inte*. Alternativ 3 kan ses som en neutral åsikt – varken ja eller nej.

	<i>Ja, absolut</i>			<i>Nej, absolut inte</i>	
1 Det var en intressant läsförståelse. (fra, spa) / Det var en intressant och rolig läsförståelse. (ty)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Uppgiften var lätt. (fra, spa) / Texten var svår. (ty)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Jag tror att jag har klarat den här uppgiften bra.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Andra synpunkter än de du redan lämnat i kryssvaren.

Formuleringen av påstående 3 var identisk i de tre enkäterna. Nr 1 och 2 såg däremot något olika ut (se ovan). Detta kan naturligtvis påverka utfallet i enkätsvaren, men eftersom syftet här främst är att undersöka om några allmänna tendenser kan urskiljas mellan elevernas upplevelser av olika uppgiftstyper, deras

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

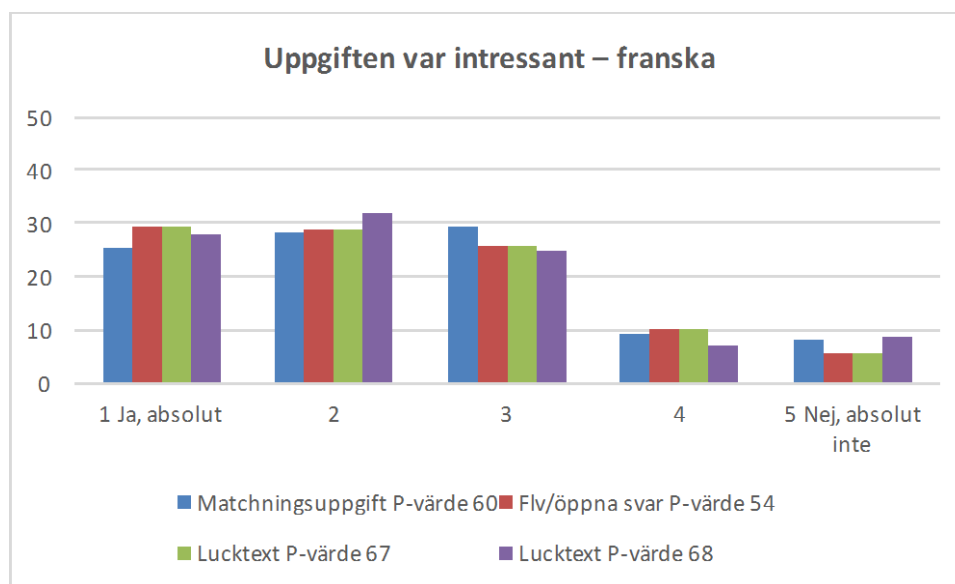
skattningar av svårighetsgrad och egen prestation i förhållande till P-värdet, är enkätsvaren ändå intressanta i sitt sammanhang.

Resultatgenomgång

Provdelarnas innehåll

Uppgiften var intressant. Eleverna som har deltagit i utprövningarna i Moderna språk har en genomgående positiv inställning till innehållet i läsförståelseuppgifterna. Allra mest positiva var eleverna som gjorde **den franska utprövningen**. Med endast ett undantag markerar hälften eller fler av eleverna i den franska utprövningen 1 eller 2 på Likertskalan, gällande påståendet *Det var en intressant uppgift*. Tre av uppgifterna får till och med 60 procent eller fler markeringar på 1 och 2. Antalet markeringar på extremvärdet 5 *Nej, absolut inte* är följaktligen lågt, vanligen mellan 2 och 8 procent.

Diagrammet nedan illustrerar utfallet i elevenkäten beträffande fyra provuppgifter med olika svarsformat. Skillnaden i svårighetsgrad är relativt stor uppgifterna emellan (se P-värdet), men värderas trots detta av eleverna som ungefär lika intressanta. Det genomsnittliga Likert-värdet för dessa fyra uppgifter ligger mellan 2,3 och 2,5.



Figur 1 Test-taker feedback, *Uppgiften var intressant* – Franska

Av eleverna som genomförde **den spanska respektive tyska utprövningen** anger 40–60 procent att de tyckte att större delen (tre fjärdedelar) av läsförståelseuppgifterna var *intressanta*. Även övriga uppgifter fick relativt positivt omdöme, strax under 40 procent av eleverna uppger att dessa uppgifter är *intressanta*. Antalet markeringar på alternativ 5, dvs. absolut inte intressanta, var i genomsnitt 8 procent i de spanska och tyska utprövningarna.

Vid en jämförelse mellan de tre olika språkens enkäter syns inget generell mönster gällande relationen mellan hur intressant uppgiften upplevs vara och olika svarsformat. Däremot kan ett visst samband utläsas beträffande svårighetsgrad eftersom eleverna ofta, men inte alltid, upplever de lättaste provuppgifterna som något mer intressanta. Oavsett såväl upplevd svårighetsgrad som P-värde finns endast en uppgift i hela underlaget där fler elever väljer extremvärdet 5 (*Nej, absolut inte*) än 1 (*Ja, absolut*), nämligen en lucktext.

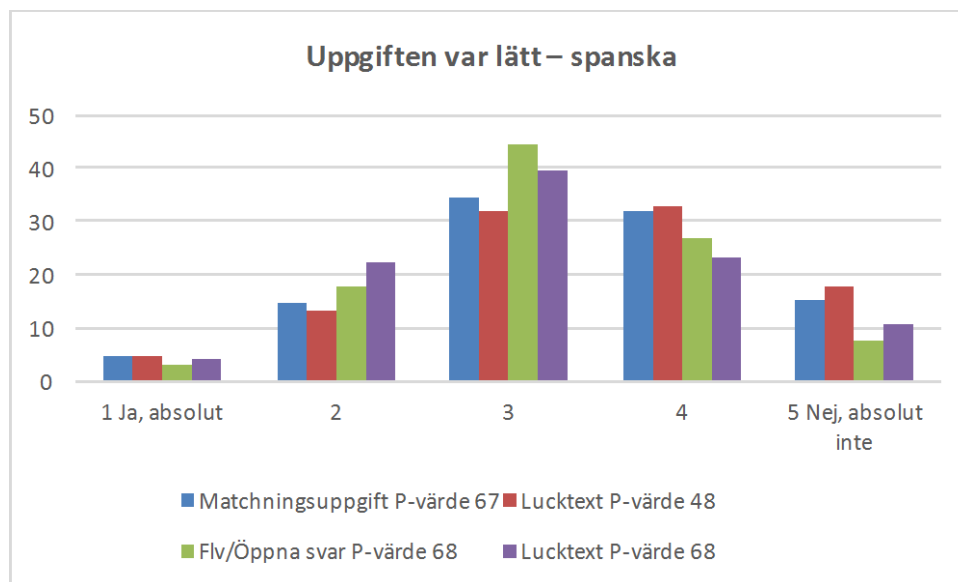
Upplevd svårighetsgrad

Uppgiften var lätt. Eleverna från **den franska utprövningen** väljer i större utsträckning alternativ 4 och 5 (dvs. *Nej* som svar på påståendet *Uppgiften var lätt*), än tvärtom. Två tredjedelar av uppgifterna upplevs som svåra. Fyra av dessa uppgifter (alla fyra med olika svarsformat) uppfattades dessutom av minst hälften av eleverna som mycket svåra (Likert 5, *Nej absolut inte [lätt]*). Tilläggas bör att P-värdena på dessa fyra uppgifter också var relativt låga (mellan 53–57%), vilket betyder att elevernas skattning här är tämligen realistisk. Men det finns också elever som väljer att markera 1 och 2 på skalan beträffande samma uppgifter. I genomsnitt 13 procent tyckte alltså att även dessa provuppgifter var lätta. I hälften av fallen är detta antal till och med över 16 procent. Detta belyser vikten av en variation av uppgifter med olika svårighetsgrad för att alla elever ska ges möjlighet att visa vad de kan och faktiskt också *känna* att de får göra det. I två fall överskattar eleverna svårighetsgraden något. Det gäller två provdelar, en lucktext och en matchningsuppgift som båda har ett P-värde på eller över 60 procent.

Även i enkäterna från **den spanska utprövningen** väljer eleverna i större utsträckning alternativ 4 och 5 än 1 och 2. I Likertskalorna till hälften av provdelarna väljer 40 procent eller fler alternativ 4 och 5. Den av dessa uppgifter som upplevdes som allra svårast (en lucktext) angavs som svår (Likert 4 eller 5) av 50 procent av eleverna. Denna uppgift hade också med god marginal lägst P-värde (48 %), vilket gör elevernas skattning rimlig i sammanhanget (se röd

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

stapel i diagrammet nedan). Den uppgift som upplevs som näst svårast (47 % av eleverna markerar här 4 och 5) är en matchningsuppgift med P-värdet 67 procent.



Figur 2 Test-taker feedback, *Uppgiften var lätt* – Spanska

Det anmärkningsvärda med matchningsuppgiften (blå stapel i diagrammet) är att den samtidigt är en av de tre uppgifter i utprövningen som helhet som har ett P-värde över 66 procent. Den tillhör alltså utprövningens lättaste, trots att den upplevs som svår eller till och med mycket svår. De övriga två uppgifterna med likartad svårighetsgrad (en lucktext med flervalsalternativ och en uppgift med blandat format, båda med P-värdet 68 %), har också skattats av eleverna som betydligt lättare (ca 33 % markerar där 4 och 5 i skalan). Denna jämförelse illustreras i ovanstående diagram.

Texten var svår (tyska). För att kunna jämföra elevernas uppfattning om svårighetsgrad har vi i analysen av svaren från **den tyska enkäten** vänt på skalan till påståendet *Uppgiften var svår* och kommer istället härnäst att referera till denna fråga som *Uppgiften var lätt*, där 1 motsvarar *Ja, absolut* och 5, *Nej, absolut inte*.

Minst 40 procent av eleverna som svarade på den tyska enkäten markerar att en majoritet av uppgifterna var svåra och valde 4 eller 5 i skalan. I tre fall väljer t.o.m. 50 procent av eleverna alternativ 4 och 5. En av dessa tre provuppgifter var en längre text med blandat svarsformat (flerval och öppna svar). Uppgiften hade också lågt P-värde jämfört med övriga i utprövningen (P-värde 56 %, jämfört med genomsnittet på 61 %), vilket gör elevernas skattning rimlig.

Däremot sticker de två andra uppgifterna ut – båda matchningsuppgifter – då de, precis som i spanskans fall, samtidigt har högt P-värde: 65 procent respektive 69 procent. Dessa två uppgifter hör till och med till de tre uppgifter som har högst P-värde i utprovningen som helhet.

Fenomenet att elever överskattar svårighetsgraden vid matchningsuppgifter, i synnerhet vad gäller läsförståelse, har även observerats i andra sammanhang. T.ex. vid studier av elevenkäter och nationella prov i engelska har matchningsuppgifter visat sig uppfattas som väsentligt svårare än de i realiteten är (Erickson, 1996 & 2006; Skolverket, 2004).

Skattning av resultat

Jag tror att jag har klarat den här uppgiften bra. Det tredje påståendet från enkäten uppmanar respondenterna att skatta sina egna resultat på provdelarna. Antalet elever som väljer extremvärdet 1, dvs. *Ja, absolut*, är i genomsnitt sammanlagt 6 procent i de tre utprovningarna. Eleverna markerar i större utsträckning 5, *Nej, absolut inte* (i genomsnitt 13 %). Inga större avvikelser förekommer mellan språken eller mellan uppgiftstyper.

De fem provuppgifter i **den franska utprovningen** som får flest markeringar på alternativ 5, *Nej, absolut inte* får i genomsnitt 15 procent. Fyra av dessa fem uppgifter är de som nämndes ovan, dvs. de *upplevt* svåraste uppgifterna i utprovningen och samtidigt de med lägst P-värde, vilket betyder att elevernas skattning även här är realistisk.

Den femte uppgiften, en matchningsuppgift, sticker däremot ut från de andra. P-värdet (60 %) ligger nära genomsnittet (62 %), men uppgiften har ändå den *näst högsta* andelen markeringar på alternativ 5 på skalan, *Nej, absolut inte*. (Detta kan jämföras med den uppgift som hade *allra flest* markeringar på alternativ 5 och som hade ett P-värde på 54 %.) Återigen tyder utfallet på att elever underskattar sina resultat på matchningsuppgifter, i synnerhet de med ett (relativt) högt P-värde.

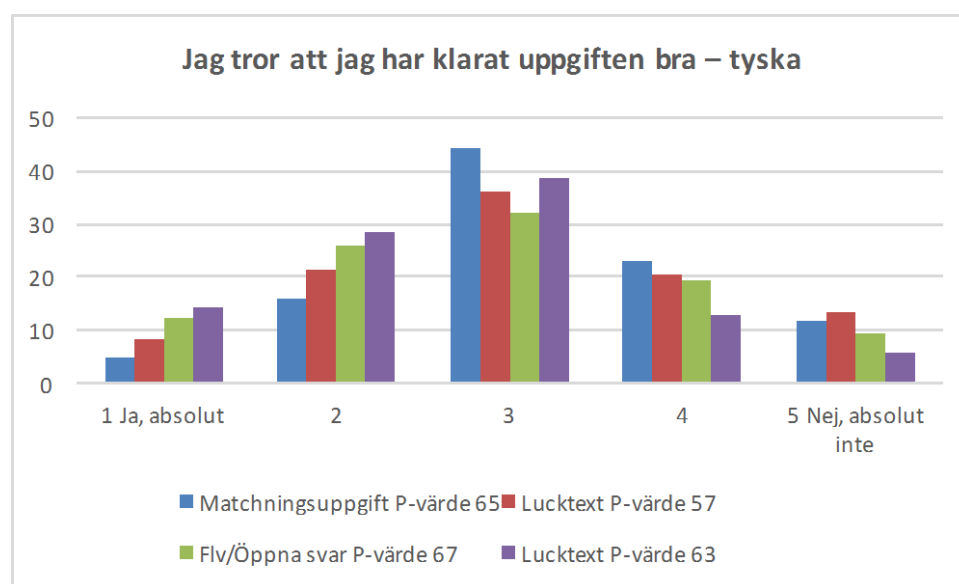
När det gäller **den spanska utprovningen** har två av provuppgifterna fått 15 procent eller fler markeringar på 5, *Nej, absolut inte*. Den ena är lucktexten, som tidigare nämnts, som har lägst lösningsfrekvens av alla uppgifter i utprovningen vilket gör elevernas skattning realistisk. Den andra är matchningsuppgiften som faktiskt hör till de lättaste enligt P-värdet.

De fyra provuppgifter i **den tyska utprovningen** som hade flest antal markeringar på alternativ 5, *Nej, absolut inte* (minst 12 %) var dels en uppgift med

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

blandat format (flerval/öppna svar), dels en lucktext, båda med relativt låga P-värden (56 respektive 57 %) samt en flervalsuppgift och en matchningsuppgift. Flervalsuppgiften var dock förhållandevis lätt enligt P-värdet (65 %), vilket indikerar att eleverna här har underskattat sin egen prestation. Detsamma gäller matchningsuppgiften, som hade lika högt P-värde som flervalsuppgiften (65 %), men där eleverna ändå upplever att resultatet blivit lågt.

Följande exempel på relationen mellan hur intressant provuppgiften upplevs vara, skattning av egen prestation och P-värde är hämtat från den tyska utprövningen. De två uppgifter som där fick högst andel markeringar på 1 och 2, var de två uppgifter som eleverna också upplevde att de hade klarat bäst (39–43 % markerar att de tror att de har klarat den här uppgiften bra). Dessa två uppgifter hade dessutom högt P-värde (63 %, respektive 67 %: lila respektive grön stapel i diagrammet nedan). Här tycks således finnas ett tydligt samband mellan högt P-värde och positiv skattning av eget resultat. Vid en jämförelse med matchningsuppgifterna, båda med ett lika högt P-värde (65 % respektive 69 %), framkommer dock samma mönster som i franska och spanska: Endast 21 procent av eleverna har här angett att de tror sig ha klarat matchningsuppgiften med ett P-värde på 65 procent bra (blå stapel nedan).



Figur 3 Test-taker feedback, Jag tror att jag har klarat uppgiften bra – Tyska

I diagrammet ovan görs en jämförelse mellan elevernas skattning av sina resultat beträffande en av den tyska utprövningens allra svåraste provdelar (en lucktext med ett P-värde på 57 procent – röd stapel) och ytterligare två provuppgifter

med varierande P-värde. Som framgår av diagrammet har eleverna i jämförelse värderat sin egen prestation på matchningsuppgiften som betydligt sämre än den i själva verket var.

Diskussion

Informationen som framkommer i elevenkäterna är av stor vikt vid sammansättningen av nya prov i moderna språk och ligger till grund för såväl val av provdelar, som ordningsföljd och tidsramar. På så sätt skapas goda förutsättningar för eleverna att visa sina färdigheter. Se vidare texten *Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ* (Finndahl & Perrotte) som i en genomgång av elevsynpunkter från receptiva och produktiva provuppgifter i franska diskuterar autenticitetsaspekten i förhållande till elevernas upplevelse av ett antal provuppgifter.

Erickson har i en internationell studie (2010) visat att elever ofta uttrycker att det som utmärker ”dåliga” prov är att de är för ensidiga och inte ger dem möjlighet att fullt ut visa sin förmåga. Det är följaktligen viktigt att eleverna får möta en hög grad av variation i det färdiga provmaterialet, t.ex. med olika innehåll, frågeformat och svårighetsgrad.

Sammantaget är majoriteten av eleverna som deltog i de franska, spanska och tyska utprovningarna positiva till innehållet i provuppgifterna. Ett flertal faktorer kan naturligtvis ha betydelse för elevernas uppfattning, förutom att det handlar om tre olika språk. Kringfaktorer som exempelvis gruppstorlek, grupp-sammansättning eller hur lätt respektive svårt skolan har haft att rekrytera behöriga lärare till ett visst språk kan också påverka hur väl förberedda eleverna i en viss språkgrupp känner sig inför att lösa provuppgifterna. Trots detta är skillnaderna mellan attityderna i enkäterna marginella språken emellan. Varken avseende hur intressant en uppgift upplevs vara, eller dess P-värde kan något genomgående samband med svarsformat skönjas. En intressant iakttagelse är att de franska eleverna är allra mest entusiastiska över uppgifternas innehåll, trots att flera av de franska provdelarna både upplevdes som – och är – svåra enligt P-värdet. Huruvida detta har med elevernas allmänna inställning till språket och studierna i sin helhet att göra eller snarast är relaterade till de konkreta uppgifterna går inte att avgöra med de data som föreligger här men vore givetvis en intressant aspekt för fortsatt forskning.

Vid en analys av elevers självskattningar ur ett genusperspektiv, finns studier som indikerar att flickor, jämfört med pojkar, oftare tenderar att retrospektivt

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

underskatta sina egna resultat (Erickson, 1996 & 2006). Detta har observerats i olika sammanhang och diskuteras också av Hedenbratt & Axelson i texten *Motiverad för uppgiften? – Test-taker feedback ur ett genusperspektiv* i denna rapport.

Vid en jämförelse kan en viss relation utläsas mellan provuppgifternas P-värde och den upplevda svårighetsgraden (*Uppgiften var lätt*) samt om eleverna tror att de klarat uppgiften bra. Generellt sett skattar eleverna svårighetsgrad och eget resultat på ett realistiskt sätt. Ett avvikande mönster som dock framträder är att matchningsuppgifter (i synnerhet de med ett högt P-värde) oftare än andra format upplevs som svårare än de i själva verket är.

Matchningsuppgifterna avviker dock inte nämnvärt från övriga uppgifter när det gäller hur intressanta de upplevs vara. Elevernas markeringar befinner sig där på genomsnittet, både beträffande antalet *Ja* (Likert 1 och 2) och *Nej* (Likert 4 och 5). Detta kan tyckas anmärkningsvärt eftersom eleverna samtidigt upplever just dessa provdelar som extra svåra och det naturliga kanske hade varit att anta att de därmed inte uppfattas som så intressanta. Enligt tidigare studier har det dock visat sig att uppgiftens svårighetsgrad tycks spela mindre roll vid elevernas uppfattning (positiv, såväl som negativ) om en uppgift (Erickson, 2006).

Ett antagande kan vara att eleverna, trots den upplevda svårighetsnivån vid matchningsuppgifter, ändå tycker att uppgiftstypen som sådan är relevant och/eller att texterna har ett intressant innehåll, vilket också poängteras av flera respondenter i enkätens öppna fält. Det anknyter också till resultaten från en internationell studie av elevuppfattningar (Erickson & Gustafsson, 2005), där utmaning var en av de egenskaper som enligt ett antal elever angavs som positiv. Följande exempel från elevenkätens öppna kommentarer till en av matchningsuppgifterna (från den tyska utprövningen) får här sammanfatta denna studie:

”Lite oklart vad vissa texter handlade om, vilket jag förstår är hela poängen.”

ATT BEDÖMA SPRÅKLIG KOMPETENS

”Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ”

Ingela Finndahl & Frank Perrotte

I denna text kartläggs elevers synpunkter från franska elevenkäter gällande upplevd meningsfullhet och hur intressant en provuppgift är. Såväl receptiva färdigheter (höra och läsa) som produktiva (skriva, tala och samtala) undersöks och detta görs utifrån påståenden i de Likertskalor som används, nämligen *Uppgiften var intressant* samt *Uppgiften testar något som är bra att kunna*. Först förklaras varför dessa två påståenden återfinns i elevenkäter, sedan redovisas och analyseras resultaten.

Varför är autenticitetsaspekten viktig i ett prov?

Som provutvecklare är det viktigt att alltid försöka skapa ett så bra prov som möjligt – men vad utmärker egentligen ett bra prov? God validitet och god reliabilitet (dvs. sträva efter att provet testar det som ska testas och genererar användbara resultat, samt att resultatet är tillförlitligt) säkerställs genom att i konstruktionsfasen noggrant titta på centralt innehåll och mål- och ämnesplaner, samt noggrant analysera olika typer av resultat efter det att provuppgifter utprovats i stor skala.

Men det finns även andra aspekter som gör att ett prov upplevs som ett välfungerande och bra prov – nämligen provets meningsfullhet. Om prov i moderna språk känns meningsfulla och intressanta för eleverna, bidrar de med stor sannolikhet till en ökad måluppfyllelse, samtidigt som de konkretiserar de nationella kurs- och ämnesplanerna. I syftestexten för ämnet moderna språk kan man läsa att: *”Undervisningen i ämnet moderna språk ska syfta till att eleverna utvecklar kunskaper i målspråket och omvärldskunskaper samt tilltro till sin egen förmåga att använda språket i olika situationer och för skilda syften. Eleverna ska ges möjlighet att, genom användning i funktionella och meningsfulla sammanhang, utveckla en allsidig kommunikation.”* Vidare ska eleverna utveckla *”[...] kunskaper om livsvillkor, samhällsfrågor och kulturella företeelser i olika sammanhang och områden där språket används”*

(Skolverket, 2011) – allt för att elevernas lust att lära sig ett nytt språk och dess kultur ska främjas.

Vidare bygger de svenska kursplanerna i moderna språk och engelska till stor del på GERS, den gemensamma europeiska referensramen för språk (Council of Europe/Skolverket, 2009), vars grundsyn är så kallad handlingsorienterad, dvs. funktionell och kommunikativ. Där beskrivs språkanvändningens områden som aktiviteter, domäner och kontexter, vilka är starkt kopplade till autenticitet och språklig funktion.

Elevernas kommunikativa språkliga kompetens bör stimuleras i olika kontexter och språkliga aktiviteter med hjälp av varierande uppgifter. *”Den nya kursplanen betonar vikten av att ta vara på språket i omvärlden, till exempel i medier. Detta är ett ställningstagande som grundar sig både på forskning och på Skolverkets undersökningar som visar att motivationen och lärandet ökar när eleverna får tillgång till ett intresseväckande och levande språk från olika sammanhang.”* (Kommentarmaterial till kursplanen i moderna språk, 2011, s. 6.) Stiggins (1987) hävdar att *“Performance assessments call upon the examinee to demonstrate specific skills and competencies, that is, to apply the skills and knowledge they have mastered.”*

Att även bedömning ska vara så autentisk som möjligt, s.k. “authentic assessment” är något som genomsyrar bedömningsforskningen. Wiggins (1993) definierar det som *“... engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively. The tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field.”* I språkdidaktisk forskning påpekas ideligen vikten av att, oavsett färdighet, väcka elevernas inre motivation, “the intrinsic motivation”, där just autenticitet är en del i att skapa meningsfullhet för eleverna; *“Therefore, you might strive to focus your students on interesting, relevant subject matter content that gets them more linguistically involved with meanings and purposes and less with verbs and prepositions.”* (Brown, 2007)

Hur konkretiseras autenticitetsaspekten i provutvecklingen?

I konstruktionsprocessen används, i så stor utsträckning som möjligt, material som är autentiskt och som kan passa aktuell målgrupp, till exempel radioreklam eller en ansökan. Sverige, såväl som många andra länder, har inte detaljstyrda styrdokument. Detta skapar en relativt stor frihet för läroboksförfattare, lärare och även provkonstruktörer att framställa material som ska passa aktuellt steg,

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

dvs. nivå i språket både innehållsligt och språkligt. För att säkerställa att uppgifterna känns stimulerande och intressanta, bidrar även verksamma lärare i konstruktionsfasen av ett prov.

I en av NAFS-projektets gemensamma principer för utveckling av prov i moderna språk och engelska fastslås att ”*ambitionen är att såväl material och uppgifter ska vara så autentiska som möjligt*” (Göteborgs universitet/NAFS). I provmaterialet innebär det att texter, samtal men även situationer, i så hög grad som möjligt, ska vara äkta. Det kan till exempel vara korta faktatexter eller utdrag ur en skönlitterär text, en intervju från ett radioprogram eller ett rollspel som efterliknar en autentisk situation. Givetvis måste man ibland förenkla, sakta ner tempot och anpassa uppgifterna efter steget – men ledstjärnan är att materialet ska ”kännas autentiskt”.

Med autenticitet menas också att provuppgifter ska vara på målspråket, dvs. både instruktioner och frågor i elevbladen. En ytterligare aspekt som höjer autenticiteten är att elevernas svar, texter och samtal som samlats in i samband med utprovning används i bedömningsanvisningarna.

Hur kan ett provs meningsfullhet undersökas?

En del av provprocessen är så kallad storskalig utprovning, vilket innebär att provuppgifter prövas ut i slumpvis valda grundskolor och gymnasieskolor i olika delar av landet. Syftet är att säkerställa provuppgifternas funktionalitet i vid bemärkelse eller, annorlunda uttryckt, deras validitet och reliabilitet. I varje utprovning finns även så kallade ankaruppgifter, d.v.s. gemensamma uppgifter för olika utprovningversioner och utprovningomgångar. Dessa uppgifter finns med för att ytterligare möjliggöra analys av varje utprovningsgrupps egenskaper och är en viktig del av provprocessens stabilitet. Efter att utprovningmaterialet skickats tillbaka vidtar olika typer av analyser och en viktig faktor som spelar in är givetvis elevernas upplevelse av provuppgifterna.

För att undersöka graden av meningsfullhet i provuppgifter, och därmed förhoppningsvis öka elevernas motivation att genomföra provet, används enkäter i dessa utprovningar, s.k. ”test-taker feedback”. För att mäta elevernas reaktioner används Likertskalan, som är en skala för att mäta olika attityder hos respondenter. Eleverna ombeds reagera på flera påståenden och i denna undersökning granskas två av dessa närmare, nämligen: *Uppgiften var intressant* och *Uppgiften testar något som är bra att kunna*, där eleverna ombeds gradera sina svar utifrån en sifferskala i fem steg där första steget motsvarar *Ja, absolut* och femte

ATT BEDÖMA SPRÅKLIG KOMPETENS

steget motsvarar *Nej, absolut inte*. Här finns också elevernas öppna svar att tillgå för att se vad de tycker mer specifikt om uppgiften. Även lärarna ges tillfälle att tycka till om uppgifternas relevans. I denna undersökning läggs fokus på det andra påståendet (*Uppgiften testar något som är bra att kunna*) och det är med hjälp av det som provuppgiftens meningsfullhet undersöks, d.v.s. om provuppgiften är användbar ur ett elevperspektiv. Även möjliga könsskillnader bland elevsvaren analyseras.

Enkät	Ja, absolut			Nej, absolut inte	
1. Uppgiften var intressant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Uppgiften testar något som är bra att kunna.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kommentar:				
				
				

Vi har valt att redovisa och diskutera resultat från storskaliga utprövningar i franska som fokuserar både receptiv förmåga (HF: hörförståelse, LF: läsförståelse) samt produktiv förmåga (FSP: fri skriftlig produktion och interaktion; FMP: fri muntlig produktion och interaktion):

- HF3 och LF3: våren 2015 utprövades receptiva provuppgifter på steg 3 (vanligen motsvarande åk 1 i gymnasieskolan). Varje provuppgift testades av cirka 400 elever.
- FSP3: våren 2016 storutprövades skriftliga provuppgifter på steg 3 (vanligen motsvarande åk 1 i gymnasieskolan). Varje provuppgift testades av cirka 270 elever.
- FMP2: våren 2015 utprövades muntliga provuppgifter på steg 2, (vanligen motsvarande åk 9 i grundskolan). Varje provuppgift testades av cirka 80 elever.

För elever som läser franska enligt kursplanen i moderna språk inom ramen för språkval i åk 9 var könsfördelningen läsåren 2012/13–2015/16 följande: 62 procent flickor och 38 procent pojkar (Skolverket, 2016). Könsfördelningen i de

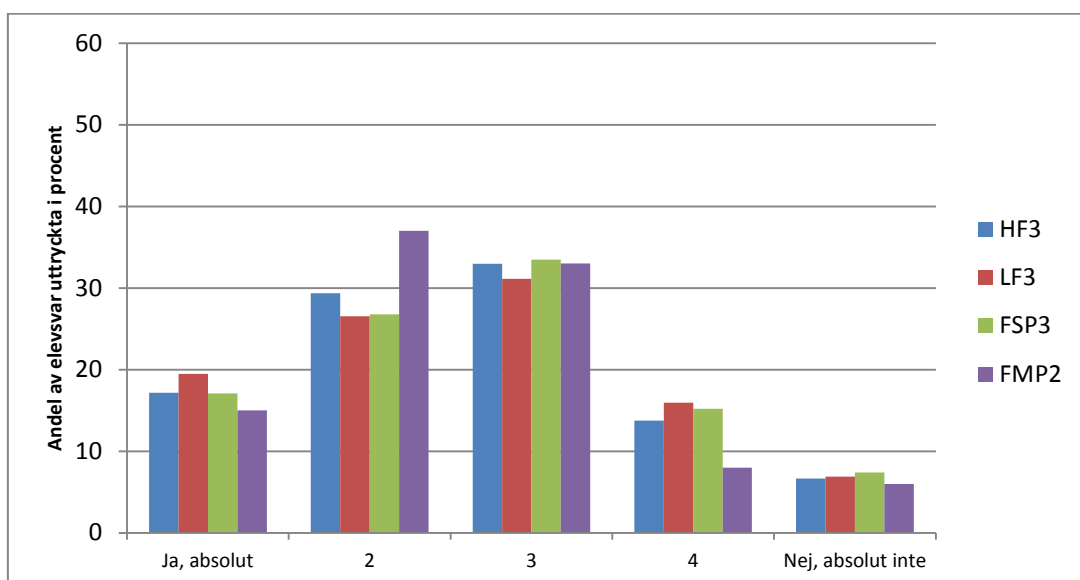
TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

storskaliga utprövningarna matchar den för hela landet och är därmed representativ i detta avseende.

Resultat och diskussion

Nedan följer en sammanställning av elevernas enkätsvar från de provuppgifter som prövades ut i stor skala och som sedan blev till prov. Beträffande HF3 och LF3 är resultaten baserade på flera provuppgifter som efter analys sattes samman till ett hörförståelseprov och ett läsförståelseprov.

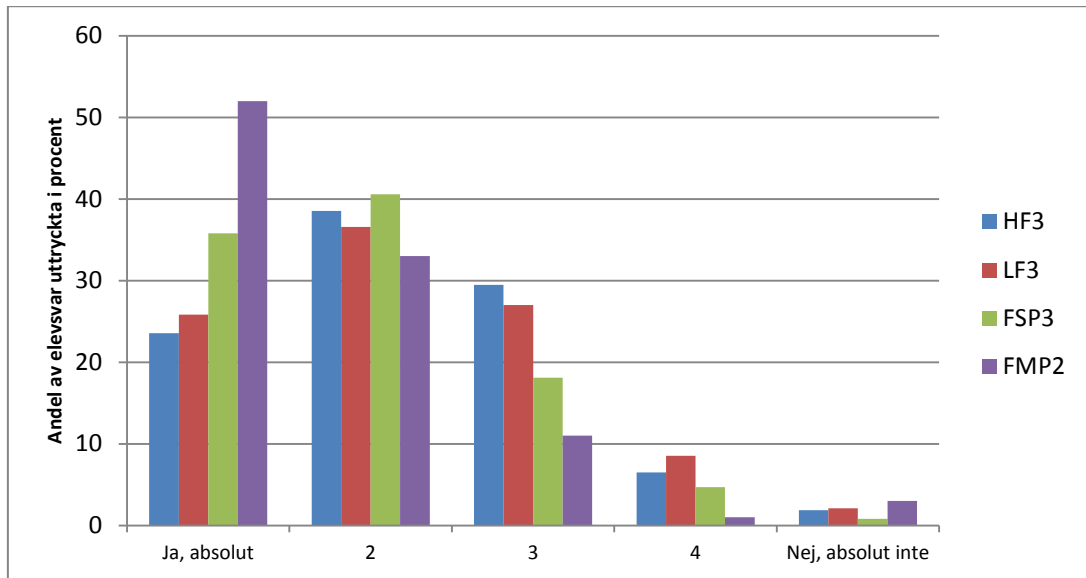
Nedan ses det sammantagna resultatet för HF3, LF3, FSP3 och FMP2 vad gäller påståendet *Uppgiften var intressant*.



Figur 1.1 Test-taker feedback, *Uppgiften var intressant*

ATT BEDÖMA SPRÅKLIG KOMPETENS

Nedan ses det sammantagna resultatet för HF3, LF3, FSP3 och FMP2 vad gäller påståendet *Uppgiften testar något som är bra att kunna*.



Figur 1.2 Test-taker feedback, *Uppgiften testar något som är bra att kunna*

Vad gäller intressefaktorn har alla fyra färdigheter likartade staplar (figur 1.1) och det finns en övervikt för stapel 1 och 2, jämfört med stapel 4 och 5. Detta bekräftas när man tittar på medelvärdet för varje färdighet där 1 motsvarar *Ja, absolut* och 5 motsvarar *Nej, absolut inte*: HF3 (2,77), LF3 (2,65), FSP3 (2,68), FMP2 (2,50). De redovisade medelvärden kan tolkas positivt eller negativt i förhållande till det neutrala värdet 3,00. Alltså svarar här eleverna att provuppgifterna är till övervägande del intressanta.

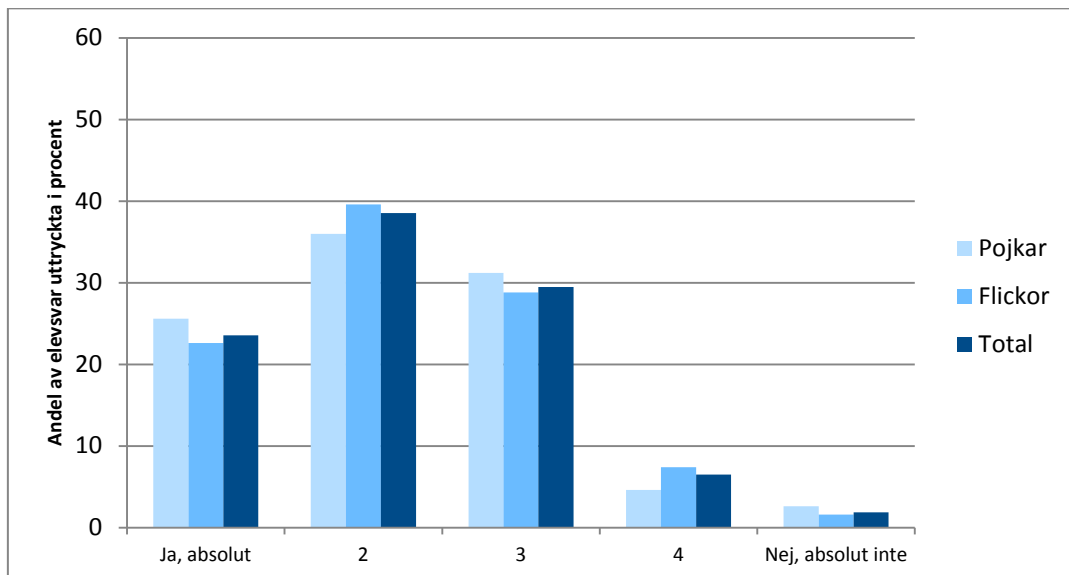
Figur 1.2 visar högre staplar i vänsterkanten än i högerkanten, oavsett färdighet, därmed visas här samma tendens som i figur 1.1 vad gäller övervikten på stapel 1 och 2 jämfört med stapel 4 och 5. Eleverna tycker att provuppgifterna testar något som är bra att kunna. Resultatet i figur 1.2 visar till och med att eleverna svarat ännu mer positivt än i figur 1.1, dvs. att övervikten är här ännu högre för stapel 1 och 2. Medelvärdet för påstående 2 för varje färdighet bekräftar detta: HF3 (2,27), LF3 (2,25), FSP3 (1,97), FMP2 (1,70). En förklaring är utprovningar i liten skala som går ut på att få fram provuppgifter som är i fas med styrdokumentet. Redan i dessa utprovningar ombeds eleverna att lämna synpunkter på de olika provuppgifterna. Denna förberedande konstruktionsfas görs för att stärka provuppgifternas funktionalitet och meningsfullhet.

Mer i detalj kan man se att resultatet i figur 1.2 för de produktiva färdigheterna skiljer sig något jämfört med de receptiva färdigheterna. I synnerhet för

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

FMP2, alltså den muntliga uppgiften för steg 2, där 52 procent av eleverna kryssade *Ja, absolut* medan 3 procent valde att kryssa *Nej, absolut inte*. Även i figur 1.1 ser man att FMP2 får något mer positiva siffror. Det kan säkerligen bero på den kommunikativa språksynen som genomsyrar den svenska ämnesplanen i moderna språk. För eleverna upplevs den muntliga förmågan som mycket viktig, medan andra förmågor kanske fått stå tillbaka något de senaste åren, till exempel förmågan att skriva. Utprövningarna i liten skala för FMP skiljer sig dock något från utprövningarna i liten skala för de andra färdigheterna: vid utprövningstillfället, möter ofta provutvecklare elever på skolorna och får på så vis direkt respons på hur uppgiften fungerar.

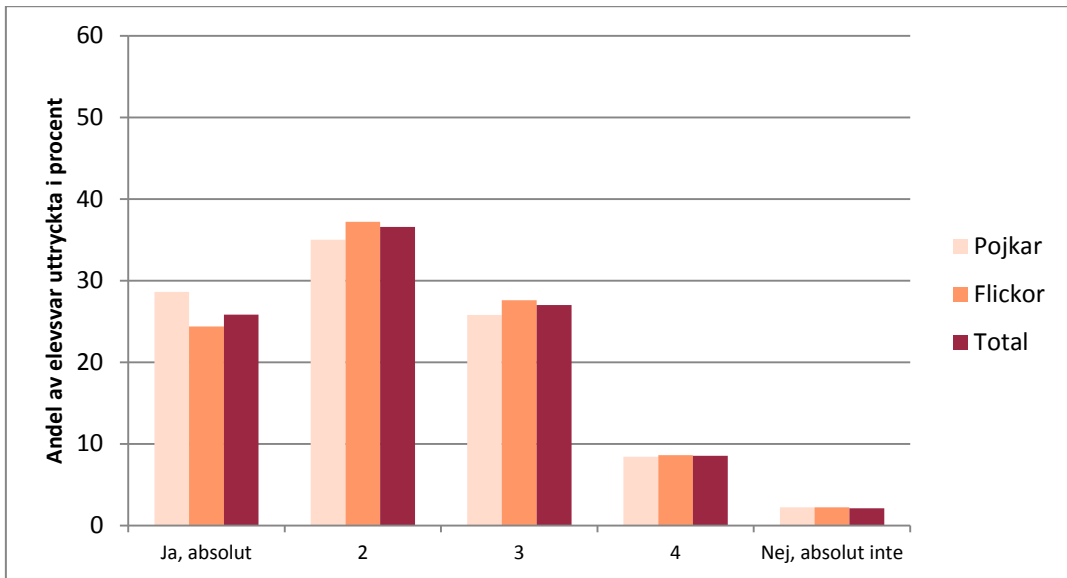
Nedan ses, utifrån kön, det sammantagna resultatet för Höra steg 3 vad gäller påståendet *Uppgiften testar något som är bra att kunna*.



Figur 2.1 Test-taker feedback, *Uppgiften testar något som är bra att kunna*. HF3; pojkar och flickor

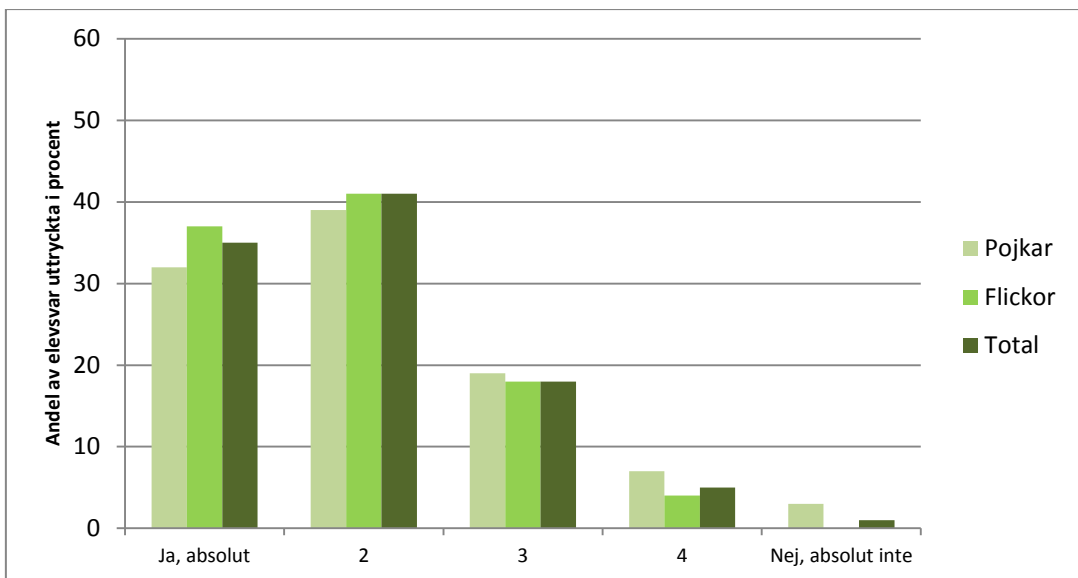
ATT BEDÖMA SPRÅKLIG KOMPETENS

Nedan ses, utifrån kön, det sammantagna resultatet för Läsa steg 3 vad gäller påståendet *Uppgiften testar något som är bra att kunna*.



Figur 2.2 Test-taker feedback, *Uppgiften testar något som är bra att kunna*. LF3; pojkar och flickor

Nedan ses, utifrån kön, det sammantagna resultatet för Skriva steg 3 vad gäller påståendet *Uppgiften testar något som är bra att kunna*.

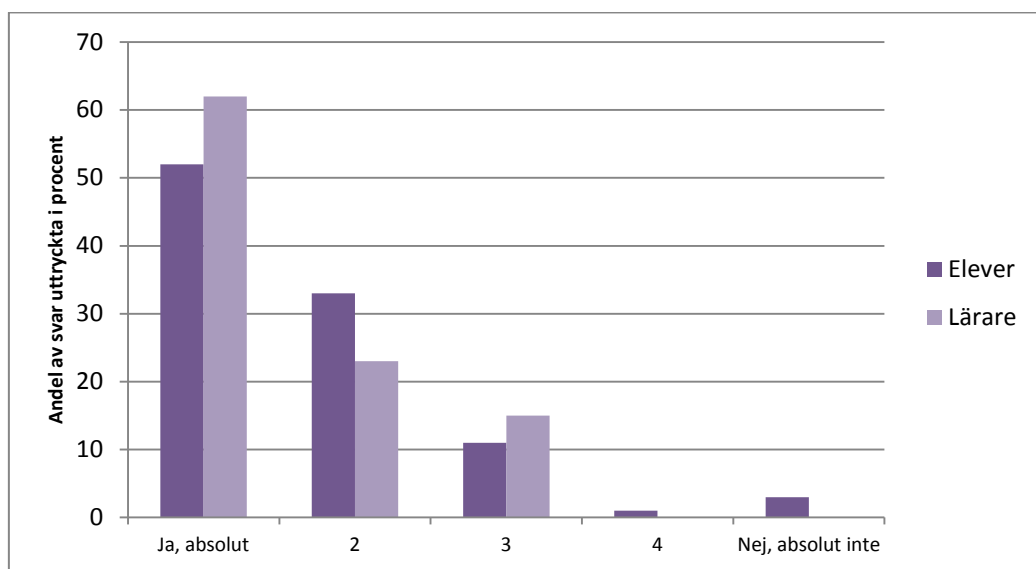


Figur 2.3 Test-taker feedback, *Uppgiften testar något som är bra att kunna*. FSP3; pojkar och flickor

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

För att undersöka provuppgifternas meningsfullhet är det också viktigt att titta närmare på eventuella könsskillnader för att undvika att missgynna den ena eller andra gruppen. I figur 2.1, 2.2 och 2.3 visar resultatet att både pojkar och flickor tycker väldigt lika om proven, och att man inte kan se några signifikanta skillnader mellan könen i upplevd meningsfullhet. Så är inte alltid fallet och man bör som provutvecklare vara medveten om sådana skillnader redan i utprovningssfasen, före storskalig utprovning. En liten skillnad går dock att utrona: nämligen resultatet på påståendet *Uppgiften testar något som är bra att kunna* som visar att medelvärdet hos pojkar i både HF3 (2,26) och LF3 (2,19) är något mer positivt än hos flickor (2,28 resp. 2,28) medan medelvärdet hos flickor i FSP3 (1,89) är något mer positivt än hos pojkar (2,10). Skillnaderna är dock små och bör tolkas med stor försiktighet. Föredrar pojkar provuppgifter med flervalformat och flickor provuppgifter med öppna, skriftliga svar? Ja, denna tendens tycks bekräftas vid närmare analys av elevsvaren i både enkäter och på provuppgifter i franska. Dessutom visar erfarenheter både från de nationella proven och andra studier, att såväl format som innehåll även kan ses ur ett genusrelaterat perspektiv (Bolger & Kellaghan, 1990).

Nedan ses det sammantagna resultatet för Tala och samtala steg 2 vad gäller påståendet *Uppgiften testar något som är bra att kunna*. Här finns både elevers och lärares svar.



Figur 3 Test-taker feedback, *Uppgiften testar något som är bra att kunna*. FMP2; elever och lärare

ATT BEDÖMA SPRÅKLIG KOMPETENS

Utprovningarna innehåller också enkäter till lärare eftersom även deras uppfattningar och reaktioner är mycket värdefulla i provkonstruktionsprocessen. Lärarna ombeds tycka till om provuppgifternas utformning, innehåll, svårighetsgrad etc. Även om antalet lärarsvar naturligtvis är mycket lägre än antalet elevsvar, kan man i viss mån göra kvantitativa analyser. Tycker lärare och elever lika? I figur 3 ser man att både lärare och elever anser att den muntliga uppgiften för steg 2 är mycket meningsfull. Resultaten på påstående 2 är väldigt lika i procenttal för lärare och elever d.v.s. att samstämmigheten mellan lärare och elever, vad gäller FMP2, är god. Medelvärdet hos lärare är 1,53 och 1,70 hos elever. Generellt sett kan man dock inte enbart utgå från att så alltid är fallet. I lärarenkäterna ligger också fokus på de öppna kommentarerna, dvs. att lärarsvaren huvudsakligen analyseras kvalitativt och kan ge oss värdefull information om hur provuppgifterna fungerar.

Nedan följer några autentiska lärarkommentarer om HF3 och LF3:

Bra tempo och tydligt tal. Inga större konstigheter – inte för akademiskt innehåll och inte heller några försvårande brytningar.

Det var svårt att hitta hörövningar liknande dessa. Det mesta är kopplat till kursboken. UR program är mycket bra, men där har eleverna förstärkning av kroppsspråk.

Mer elevnära ungdomliga texter t.ex. bloggar. Mail? Facebookinlägg? Ungdomsfokus? Normkritiskt tänk? Aktualiteter från gångna året?

Jag personligen tyckte att det var en bra uppgift men insåg att eleverna hade svårigheter att välja rätt svar. Blev mycket överraskad.

På det hela taget upplevde eleverna hörförståelsen som ganska svår. Få elever hade över 30 poäng, de flesta låg runt 20 poäng. Samtidigt tycker jag att nivån är rimlig.

När vi hade genomfört testen kom det många klagomål om hur svårt det var. Och i några enstaka fall kan jag förstå vad de menade, t ex i delar av hörförståelsen. MEN, jag kan egentligen inte se att de gjorde så annorlunda ifrån sig än vad grupper gjort tidigare år. De blev också själva bitvis positivt överraskade av resultatet.

I sin helhet upplevde jag provet som ett ”lätt” prov. Å andra sidan tror jag att många elever har en mycket bra chans att klara det, vilket är positivt. Ämnesinnehållet var relevant och väl överensstämmande med det centrala innehållet.

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

Innan provet skrev jag att det var bra med den progressiva svårighetsgraden. Någon/några av mina elever håller dock inte med mig. Våldigt bra med tidslängden, alla hann klart (tidigare har det varit ont om tid). Tydligt att mina elever behöver läsa mer på franska.

Helhetsintrycket är bra. Jag tycker det är bra att eleverna inte behöver formulera så många egna svar när det är hörförståelse, de har ofta svårt nog att hänga med i talet!

Uppgifter med bara kryssfrågor är läträttade men det är en stor risk att eleverna sneglar på grannen (särskilt när man har en stor grupp).

Nedan följer några autentiska lärarkommentarer om FSP3 och FMP2:

Kanske att ämnet känns ovant för elever men de flesta förstod ändå vad de skulle göra.

Vissa elever har svårt att förstå instruktionerna. Skulle man skriva [...] eller vad? Mycket bra uppgift! Verklighetsanknuten. [...] Ger bra utslag.

Ämnet stämmer väl överens med de teman som tas upp i steg 3. Även texttypen speglar kursen väl. Jag trodde att instruktionen skulle vara i för många steg med [...], men den verkar ha fungerat. Lagom många frågor som fortfarande lämnar utrymme för fantasin. Bra med bilder, särskilt som ett komplement till instruktionen. Provet passar för såväl svaga som starka elever. Eleverna kan anpassa texterna till sin egen nivå – bra! Överlag ser jag att de elever som verkligen utnyttjade tiden lyckades bäst. Även någon svag text finns här, men den skulle ha varit ännu sämre med kortare tid.

I elevenkäterna ligger snarare fokus på elevernas svar på flervalsfrågorna (Likertskalan), men elevsvar kan också analyseras kvalitativt då öppna svar finns att tillgå. Även om en majoritet av eleverna väljer att inte skriva någon kommentar alls och att kommentarfälten oftast lämnas tomma, ges ändå många värdefulla kommentarer av elever.

Nedan följer några autentiska elevkommentarer om FSP3 och FMP2:

Jag tyckte att det var lite roligt men supersvårt.

Det var bra med förslag på vad man kunde skriva om. Dock var det lite svårt att skriva i imperfekt, så jag blandade mellan presens och passé composé.

Ökar ni valfriheten så är uppgiften grym.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Var tvungen att stanna kvar två timmar i skolan för att göra denna uppgift. Att det fanns så många sidor att skriva på gjorde mig stressad.

Jag upplevde provet svårt men det beror nog på att mina kunskaper brister, så det är svårt att säga vilken nivå det är på. Personligen har jag lättare att svara på frågor istället för att skriva fritt. Kan man kombinera både och kanske?

Om det hade funnits bilder istället för textämnena hade det varit bättre.

Det var bra att få en chans att prata på franska.

I det stora hela var uppgiften ganska enkel, dock var instruktionen på uppgift 2 väldigt svår att förstå vilket gjorde hela uppgiften lite för krånglig. Men det var i alla fall roligt att genomföra provet.

Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ.

Avslutande reflektioner

Sammantaget är eleverna i högsta grad delaktiga i provutvecklingen genom att de genomför provuppgifterna och även som viktiga synpunktsgivare. Elevsvaren i de storskaliga utprövningarna visar att autenticitetsaspekten i provuppgifterna har hög relevans i såväl receptiva färdigheter (höra och läsa) som produktiva (skriva, tala och samtala). Överensstämmelsen är god vad gäller intresse och meningsfullhet, och resultat för de produktiva färdigheterna visar på ännu bättre siffror. Alla fyra färdigheter är viktiga i en språkförmåga, men enligt dessa elevenkäter tycker eleverna att just provuppgifter som fokuserar på skriftlig och muntlig produktion och interaktion är något bättre än de som prövar den receptiva förmågan.

Det bör framhållas att resultaten på enbart dessa två påståenden aldrig kan tillåtas styra valet av provuppgifter. Dock ses resultaten som en viktig indikator utifrån vilka vidare kvalitativa analyser görs. För att få svar på varför eleverna tycker som de gör i elevenkäterna skulle man behöva göra mer kvalitativa undersökningar, och här finns utrymme för fortsatta undersökningar och fördjupad forskning.

Motiverad för uppgiften? – Test-taker feedback ur ett genusperspektiv

Lena Hedenbratt & Patrik Axelson

Inledning

Underlag – utprovning av receptiva uppgifter för spanska 2/språkval åk 9

Under vårterminen 2014 gjordes en storskalig utprovning av uppgifter för hör- och läsförståelse med företrädesvis elever i åk 9 på grundskolan som läste spanska som språkval och ett mindre antal elever som läste spanska 2 på gymnasienivå, (totalt 42 st.). I underlaget för denna studie finns *per provuppgift* ca 300 elevresultat (hur eleverna lyckats med uppgifterna och deras kommentarer) samt elevdata (skola, kön) från olika skolor i landets alla delar. Urvalet av skolor gjordes slumpvis, utifrån Skolverkets skoldatabas, där landets alla skolor finns registrerade. Av det totala antalet elever som medverkat i utprovningen valdes dessa ca 300 ut för analys med hjälp av ett slumpurval. Detta gjordes dels på grund av alltför omfattande insänt material, dels med hänsyn till risken att en viss grupp (klass/skola) blir överrepresenterad. Könsfördelningen (ca 55 procent flickor och 45 pojkar) är dock representativ för utprovningen som helhet.

I direkt anknytning till hör- eller läsförståelseuppgifterna gavs eleverna möjlighet att svara på fem respektive sex frågor (beroende på vilken språklig färdighet som prövades) enligt en 5-gradig Likertska. (Att använda sig av ett ojämnt antal alternativ, som fem, kan möjliggöra att det mittersta alternativet ses som ett mellanläge för dem som väljer att förhålla sig neutrala till påståendet.)

Eleverna ombads ta ställning till ett antal påståenden om uppgifterna var intressanta, deras svårighetsgrad, mängd svåra ord, instruktionernas tydlighet, talhastighet (hörförståelse), hur väl de bedömer att de lyckats och om de fått tillräckligt med tid att göra delprovet (höra/läsa). De markerade sina uppfattningar från *Ja, absolut* (1) till *Nej, absolut inte* (5) och fick också möjlighet att uttrycka egna synpunkter i ett öppet fält. Föga förvånande var svarsfrekvensen i de öppnafälten väsentligt lägre än i delarna där eleverna skulle välja bland fasta

ATT BEDÖMA SPRÅKLIG KOMPETENS

svarsalternativ, i genomsnitt ca 8 procent (läsförståelse) respektive 10 (hörförståelse), jämfört med 97–98 procent för frågor med fasta svarsalternativ. Möjligheten att med egna ord uttrycka synpunkter om utprovningssuppgifterna utnyttjades alltså i liten utsträckning av eleverna i spanska. Som jämförelse kan nämnas senare års utprovningar till de receptiva delarna i det nationella provet i engelska i årskurs 9 (2014, 2015) där det i genomsnitt var 58 procent av eleverna som valt att lämna en kommentar i det öppna svarsfälten.

	<i>Ja, absolut</i>			<i>Nej, absolut inte</i>	
Det var en intressant uppgift.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgiften var svår.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jag tror att jag har klarat uppgiften bra.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Övriga kommentarer:					

Frågeställningar

De frågor vi hoppas finna svar på och belysa i denna studie har främst ett *genusperspektiv* och fokuserar pojkars respektive flickors resultat samt deras attityder till provuppgifterna i spanska:

- Vilka skillnader mellan pojkars och flickors attityder till provuppgifterna kan utläsas utifrån svaren i elevenkäterna?
- Vilka skillnader kan man se mellan pojkars och flickors resultat på varje provuppgift?
- Vilka skillnader framträder mellan pojkars och flickors uppfattning om hur väl de klarat uppgiften jämfört med deras resultat på varje provuppgift?

Vi har valt att redovisa resultat från utprovningen av provuppgifter till bedömningsstöd för elever som läser spanska som språkval i *grundskolan*, eftersom det utan tvekan är den största elevgruppen som läser moderna språk i svensk skola idag: 56 procent av de 71 procent av eleverna i grundskolan som läser moderna språk (Skolverket, 2015) och därmed är en viktig grupp att fokusera. Dessutom är detta den enda elevgrupp i moderna språk som var representerad i den

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

Internationella Språkstudien 2011 (Skolverket, 2012) vars resultat utgör en intressant jämförelse.

I denna studie redovisas endast elevernas enkätsvar på de utprovade provuppgifter som senare valdes ut för att ingå i receptivt prov för spanska 2/språkval åk 9 år 2015. Uppgifterna prövades ut i olika versioner, med olika uppgifter. Följaktligen är det inte alltid samma elever som svarat på varje enkät. Enkäterna innehöll fler påståenden, som tidigare nämnts, men givet studiens fokus kommer endast svaren på följande tre att redovisas här:

- Det var en intressant uppgift
- Uppgiften var svår
- Jag tror att jag har klarat uppgiften bra

De elevsvar som redovisas i tabellerna nedan visar vad eleverna (uppdelade per kön) tyckt om uppgifterna. Både svarets frekvens i varje kategori (1–5), uttryckt i procent och medelvärdet per kategori redovisas. Orsaken till detta är att frekvenser ger en tydligare bild av spridningen i svaren, medan medelvärden gör utfallet på de olika uppgifterna lättare att jämföra. I samband med frågan om eleverna tror sig ha klarat uppgiften bra har vi även valt att redovisa P-värden³ för att kunna jämföra elevens uppfattning om sin förmåga med deras uppvisade resultat (P-värdet).

I kommentarerna till tabellerna (med Likertskalan 1–5) ser vi 3 som ett neutralt värde och beskriver 1 och 2 som *intressant*, *svår*, *klarat bra* och 4 och 5 som *mindre intressant* etc. Det kan vara värt att påpeka att alternativ 3 är det mest valda alternativet i de allra flesta uppgifter. Det *medelvärde* som redovisas längst ner i varje tabell kan alltså tolkas som antingen ett positivt eller negativt värde i förhållande till 3,00. På frågorna *intressant*, *klarat bra* eftersträvas ett lågt värde < 3, medan frågan om *svår* idealt skulle ha ett neutralt värde (3) eller > 3 för att tolkas som positivt. Vi kommenterar endast signifikanta könsskillnader. Övriga resultat kan utläsas av tabellerna.

³ Det proportionella värdet, P-värdet eller svårighetsnivån för en uppgift, även kallat *lösningfrekvensen*, står i detta sammanhang för det uträknade medelvärdet och räknas fram genom att dividera det antal som besvarat uppgiften rätt med det totala antalet som besvarat uppgiften. P-värdet kan anta ett värde mellan 0 och 1 där 0 anger att ingen svarat rätt och 1 anger att alla svarat rätt.

Resultatgenomgång

Nedan presenteras endast **en tabell per färdighet** och tillhörande kommentar. Den *sammanfattande* kommentaren per färdighet baseras dock på analyser av samtliga uppgifter som valdes ut till provet 2015 (5 läsförståelse- och 5 hörförståelseuppgifter).

Läsförståelse

Tabell 1. Resultat av elevenkäten för Uppgift 1 fördelat på **pojkar** respektive **flickor**.

Flerval och öppna svar Uppgift 1	Det var en intressant uppgift			Uppgiften var svår			Jag tror jag har klarat den bra				
	Pojkar	Flickor	Differens	Pojkar	Flickor	Differens	Pojkar		Flickor		Differens andel
							Andel	P-värde	Andel	P-värde	
1: Ja, absolut	7%	10%	-3%	25%	21%	4%	3%	0,52	2%	0,78	1%
2	23%	21%	2%	33%	32%	1%	26%	0,57	18%	0,66	8%
3	36%	39%	-3%	32%	35%	-3%	40%	0,45	34%	0,55	6%
4	19%	22%	-3%	9%	10%	-1%	19%	0,37	25%	0,46	-6%
5: Nej, absolut inte	15%	8%	7%	1%	1%	0%	11%	0,40	21%	0,37	-10%
Medelvärde	3,11	2,98	0,13	2,28	2,38	-0,10	3,09	0,46	3,46	0,53	0,37

Uppg. 1 är en lite längre lästext med både flervalsfrågor och frågor där eleverna själva ska skriva ett eller flera ord. 30 procent av båda könen tyckte att uppgiften var *intressant* och ungefär lika många var av motsatt åsikt. Dock utmärker sig pojkarna genom att i högre grad välja alt. 5, *Nej absolut inte*. Drygt hälften av eleverna uppfattade uppgiften som *svår*, och en fjärdedel trodde sig ha *klarat uppgiften bra*, men här är skillnaden stor mellan könen; 29 procent av pojkarna men endast 20 procent av flickorna trodde så. Fler elever (40 procent) trodde *inte att de klarat den bra*, 30 procent av pojkarna och 46 procent av flickorna. 21 procent av flickorna valde alt. 5, *Nej, jag tror absolut inte att jag klarat den bra*, vilket även stämmer med deras resultat (dessa elevers P-värde var 0,37). De uppvisade alltså en tämligen realistisk uppfattning om den egna förmågan. Däremot anger pojkarna i mycket större utsträckning att de klarat uppgiften bra (alt.1–2) trots att deras P-värde är betydligt lägre än flickornas, vilket tyder på att de i viss mån överskattade den egna förmågan.

Om man sammanfattar resultaten från samtliga fem läsförståelseuppgifter kan konstateras att den andel elever som totalt sett tycker att uppgifterna är intressanta är ungefär lika stor som de som är av motsatt åsikt vad gäller de tre första uppgifterna, 30–35 procent. De två sista uppgifterna tycker ca 40 procent

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

av eleverna är mindre intressanta. Den populäraste uppgiften var nr 3, en kort lästext med blandade svarsformat. Det är även den uppgift som uppfattats som något mindre svår. I övrigt anser de flesta elever att uppgifterna är svåra och en stor andel elever (35–45 procent) tror inte att de klarat uppgifterna bra. Flickorna har lyckats bättre med alla provuppgifterna, även i de fall de inte uppfattat det så. Särskilt tydligt är detta i uppgift 2 där flickornas P-värde var mycket högre än pojkarnas. Ett återkommande mönster är att pojkarna i högre grad än flickorna tycker att uppgifterna är mycket svåra och absolut inte intressanta, men tror sig ha klarat uppgifterna bättre. Möjliga orsaker till detta diskuteras vidare under *Diskussion*.

Hörförståelse

Tabell 2. Resultat av elevenkäten för Uppgift 1 fördelat på pojkar respektive flickor.

Matchning Uppg. 1	Det var en intressant uppgift			Uppgiften var svår			Jag tror jag har klarat den bra				
	Pojkar	Flickor	Differens	Pojkar	Flickor	Differens	Pojkar		Flickor		Differens andel
							Andel	P-värde	Andel	P-värde	
1: Ja, absolut	4%	4%	0%	39%	35%	4%	4%	0,82	3%	0,82	1%
2	12%	9%	3%	32%	36%	-4%	23%	0,64	10%	0,83	13%
3	34%	42%	-8%	23%	17%	6%	35%	0,57	37%	0,62	2%
4	24%	31%	-7%	4%	10%	-6%	24%	0,49	31%	0,61	-7%
5: Nej, absolut inte	27%	15%	12%	2%	2%	0%	14%	0,37	19%	0,48	-5%
Medelvärde	3,58	3,43	0,15	1,99	2,07	-0,08	3,21	0,55	3,53	0,62	-0,32

Uppg. 1 är en matchningsuppgift som eleverna får lyssna på en gång. Ungefär lika andelar pojkar och flickor, ca 15 procent, angav att den var *intressant*. Betydligt fler (48 procent) tyckte inte så och uttryckte sig övervägande negativt och det finns stora skillnader mellan könen: 27 procent av pojkarna valde alt 5, *Nej absolut inte* (12 procent fler än flickorna). En stor andel elever av båda könen uppfattade uppgiften som *svår*, 71 procent. Totalt 44 procent trodde *inte att de klarat den bra*, 38 procent av pojkarna och 50 procent av flickorna. 14 procent av pojkarna valde alt. 5, *Nej, jag tror absolut inte att jag klarat den bra*, vilket överensstämmer med deras resultat (dessa elevers P-värde var 0,37). De uppvisade alltså en realistisk uppfattning om sin förmåga. Av flickorna var det 19 procent som valde alt. 5 och dessa elevers P-värde var 0,48, vilket tyder på mindre tilltro till den egna förmågan. Pojkarna anger däremot i mycket större utsträckning att de

ATT BEDÖMA SPRÅKLIG KOMPETENS

klarat uppgiften bra – se alt. 2 – trots att deras P-värde är betydligt lägre (0,64) än flickornas (0,83). Dessa pojkar överskattade alltså i viss mån den egna förmågan.

Om man sammanfattar resultaten från samtliga fem hörförståelseuppgifter kan konstateras att den andel elever som totalt sett tycker att hörförståelserna är intressanta bara är hälften så stor som den som är av motsatt åsikt. Detta gäller fyra av delarna (ca 20 respektive 40 procent). Jämfört med läsförståelseuppgifterna, där andelen elever som totalt sett tycker att uppgifterna är intressanta är ungefär lika stor som de som är av motsatt åsikt, är skillnaden mycket stor. Den populäraste hörförståelseuppgiften, nr 3, är en intervju som eleverna fick lyssna på två gånger, med blandade svarsformat, även om den uppfattats ungefär lika svår som de övriga, vilket torde peka på att just innehållet uppfattats som intressant.

Jämfört med läsförståelseuppgifterna tycker en större andel elever att hörförståelseuppgifterna är svåra (50 respektive 68 procent om man lägger samman alla de analyserade uppgifterna i resp. färdighet). En stor andel elever (36–44 procent) tror inte att de klarat uppgifterna bra, vilket är en liknande siffra som för läsförståelseuppgifterna. I fyra av de fem uppgifterna är det flickorna som visar på betydligt mindre tilltro till den egna förmågan, trots att deras P-värden är högre än pojkarnas. Denna tendens förekommer även i Gudrun Ericksons studie *Test-taker feedback – Elevers synpunkter som del av arbetet med nationella prov i engelska* (Erickson, 1998):

”Svaren på den retrospektiva självbedömningsfrågan visade att flickor betydligt oftare än pojkar underskattade sina prestationer. I synnerhet gällde detta fri skriftlig produktion, vilket får betraktas som anmärkningsvärt, eftersom flickor inom detta delområde oftast presterar bättre och även av sina lärare bedöms som bättre än pojkar” (s. 89)

Flickorna har för övrigt lyckats bättre med alla provuppgifterna. Ett återkommande mönster är att pojkarna i högre grad än flickorna tycker att uppgifterna är mycket svåra och absolut inte intressanta, men tror sig ha klarat uppgifterna bättre. Möjliga orsaker till ovanstående diskuteras vidare under *Diskussion*.

Det bör påpekas att en viss osäkerhetsfaktor kring utfallet kan finnas, exempelvis vad gäller elevernas motivation och attityder samt i vilken utsträckning de gjorde sitt absolut bästa när uppgifterna prövades ut (se vidare under *Diskussion*). Alla involverade lärare har kanske inte tolkat informationen på samma vis och följaktligen inte heller förmedlat exakt samma information till eleverna vad beträffar t.ex. besvarandet av elevenkäterna.

Diskussion

Vilka skillnader mellan pojkars och flickors attityder till provuppgifterna kan utläsas utifrån svaren i elevenkäterna?

Det var en intressant uppgift

Som kommenteras ovan är det totalt sett en större andel elever som uppvisar en neutral eller negativ attityd till provuppgifterna (främst vad gäller hörförståelse) än tvärtom. Läsförståelseuppgifterna mottogs dock med något större entusiasm. För påståendet *Det var en intressant uppgift* kan utläsas att flickorna är mer centrerade till mitten av skalan än pojkarna vad gäller både hörförståelse och läsförståelse. Pojkarna har, i något större utsträckning än flickorna, valt de yttersta stegen i skalan och är därmed mer utspridda över de fem nivåerna. Detta gäller framför allt hörförståelse, men det yttersta steget, alt. 5, *Nej, absolut inte* valdes av pojkar i större utsträckning på *alla* provuppgifterna. Vad beror då elevernas måttligt entusiastiska reaktioner på? Dels finns det en generell tendens att till största delen kryssa i det neutrala alternativet 3, (där man återfinner den största andelen enskilda svar på alla enkätens frågor, med några få undantag). Det ligger kanske i sakens natur att en viss andel elever generellt inte upplever en provsituation som något positivt. Följaktligen förhåller sig en stor del av eleverna, enligt enkäterna, neutrala till uppgifterna.

Att pojkarna uppvisar en mindre positiv attityd till uppgifterna än flickorna kan ha många individuella förklaringar, men en sådan kan vara kopplad till *motivation*. I doktorsavhandlingen *Motivation och inläring ur ett genusperspektiv* (Jakobsson, 2000) framförs idén att i den aspekt av motivation som handlar om *intresse* för innehållet i det som skall läras, speglas den vanliga ordningen genom att pojkar/unga män är mer intresserade av manligt kodade ämnesområden, som exempelvis naturvetenskap och teknik, medan flickor/unga kvinnor intresserar sig mer för kvinnligt kodat innehåll, bland annat språk, litteratur och historia. En annan aspekt av motivation är målorientering med två varianter, lärandemål och prestationsmål.

”Med lärandemål strävar individen efter att öka sin kompetens, t.ex. genom att förstå något på ett bättre sätt eller behärska något man inte kunde tidigare. Med prestationsmål lär man för att få sin kompetens bekräftad, att få gynnsamma omdömen om sin förmåga och/eller att undvika ogynnsamma sådana ...” (Jakobsson, 2000, s. 45).

Tämligen genomgående i den forskning som finns har flickor oftare lärandemål eller kombinationen lärande- och prestationsmål, vilket är det mest gynnsamma för inläring, vilket är en möjlig anledning till att de uttrycker sig mindre negativt om uppgifterna och även lyckas bättre med dem. (Jakobsson, 2000)

I doktorsavhandlingen *L3 Motivation* (Henry, 2012) med syfte att undersöka motivationen för att lära sig ett tredje språk samtidigt som man lär sig engelska från klass 4–9 och att dessutom se till könsskillnader, analyseras bl.a. påverkan av självbilden för L2 engelska på L3 motivation. L2 och L3 (*second* och *third language*) refererar till det andra respektive tredje språket man lär sig efter modersmålet. Några av slutsatserna visar på att en hög grad av jämförelser mellan självbilder för L3–L2 kan tolkas negativt för det tredje främmande språket i relation till självbilden för L2 engelska, och därmed kan vara en av anledningarna till låg motivation att lära sig ett tredje främmande språk (franska, spanska, tyska). Speciellt tydligt var detta för pojkarna som deltog i undersökningen (s. 104). Henry hävdar t.o.m. att:

“As the results of the research reported here suggest, motivation to learn L3s declines during secondary education, with declines being particularly steep for boys. One of the factors responsible appears to be the negative effect of L2 English on motivation to learn other less socially dominant L3s.” (s. 123).

Det skulle kunna vara så att medan elever har mycket positiva upplevelser av olika uppgifter i engelska, som många uppfattar som lätta (och därmed intressanta), uppfattar de i hög grad uppgifterna i spanska som svåra (och därmed mindre intressanta).

I den *Internationella språkstudien 2011* (Skolverket, 2012) som prövade och jämförde 15-åringars kunskaper i L2 och L3 i 14 europeiska länder, visar analyser av elevenkäterna att svenska elevers attityder avviker från eleverna i andra länder, samt mellan engelska och spanska. När eleverna anger hur användbart det studerade språket är, ligger svenska elevers attityder till engelskan i topp tillsammans med flera andra länder. Ett mer förvånande resultat, eftersom en så hög andel svenska elever väljer spanska som modernt språk, är att spanskan hamnar i den absoluta botten jämfört med andra länders L3, då svenska elever uppger att spanskan knappast är användbar för dem.

Uppgiften var svår

Skillnaden mellan könen för påståendet *Uppgiften var svår* är generellt så liten att den inte är signifikant. (Dock har pojkarna här i större utsträckning än flickorna

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

valt det yttersta steget, alt. 1, Ja, *absolut* på alla uppgifterna i utprövningen.) Även här kan utläsas att flickorna, med några få undantag, är mer centrerade till mitten av skalan än pojkarna. En stor andel elever av båda könen ansåg alltså att uppgifterna var svåra. Tilläggas bör dock att de *utprövande lärarna* till allra största del (i genomsnitt 80 %) bedömde den språkliga svårighetsgraden i uppgifterna som *lagom* eller *lätt* och att de speglar kunskapssynen i kurs- resp. ämnesplanen för aktuell nivå.

Om Henrys slutsatser ovan stämmer, dvs. att elever på högstadienivå som studerar L3 tappar sin motivation delvis p.g.a. jämförelser med resultaten i engelska, och att detta visar sig än mer tydligt bland pojkar, kan deras något mindre positiva attityder till de spanska uppgifternas svårighetsgrad vara åtminstone ett av skälen till dessa attityder.

En annan förklaring kan vara s.k. ”antipluggkultur” som visar sig något tydligare bland pojkar. Inga Wernersson skriver i diskussionen till sin rapport *Könskillnader i skolprestation – idéer om orsaker* (SOU2010:51) bland annat:

”... att ”antipluggkultur”, som tidigare identifierats som arbetarklassrelaterat avståndstagande från en medelklassdominerad skola, brett ut sig och nu inkluderar också medelklassens pojkar. Antipluggkultur handlar om att vara ”cool”, populär och ha roligt. Att vara duktig i skolan och arbeta hårt för att lyckas utgör antitesen och innebär risk för socialt misslyckande.” (s. 58)

Men hon skriver också:

”Antipluggkultur återfinns dock också bland flickor (åtminstone i England). Ett förslag till förklaring av antipluggkulturens utbredning är att den växt fram som försvar både för social självbild (bl.a. maskulinitet) och för akademisk självbild (rädsla för misslyckande).” (s. 58)

Vilka skillnader kan man se mellan pojkars och flickors resultat på varje provuppgift?

Som kommenterats ovan har flickorna lyckats bättre än pojkarna med alla provuppgifterna i utprövningen, även om inte heller deras resultat kan anses vara helt tillfredsställande med tanke på att den avsedda svårighetsgraden är anpassad till centralt innehåll och kunskapskrav i gällande kursplan. Men vad beror då dessa, och framför allt pojkarnas, svaga resultat på? De enskilda förklaringarna är förstås lika många som eleverna, men en kan vara *motivation*. Det mest vanligt förekommande är att lösningsfrekvensen (P-värdet) stiger vid skarpt läge, dvs. då uppgifternas ges som ”riktigt prov” jämfört med vid utprövningen,

p.g.a. starkare motivation hos eleverna, vilket i sin tur leder till bättre prestationer. Är då detta mönster tydligare hos pojkarna? Möjligen, om pojkar i högre utsträckning än flickor har s.k. prestationsmål (Jakobsson, 2000) så kan detta påverka att de inte anstränger sig tillräckligt och i högre grad uttrycker en negativ attityd till utprovningssuppgifterna eftersom dessa uppgifter inte är så kallade *high stakes*-prov. En sådan skillnad mellan könen har dock inte kunnat påvisas i det begränsade antal elevsvar som inrapporterats (se nedan). Resultaten förbättras i ungefär samma utsträckning mellan utprovning och *high stakes*-prov vad gäller båda könen.

Vad gäller det bedömningsstöd för receptiva färdigheter (spanska 2/åk 9) som publicerades våren 2015, inrapporterades direkt till NAFS-projektet 130 lärarenkäter och 1 412 elevresultat, fördelat på 93 procent grundskolor och 7 procent gymnasieskolor. Även här var fördelningen ca 55 procent flickor och 45 pojkar. Resultaten påverkas dock av vilka lärare som skickat in materialet, den s.k. selektionseffekten, och följaktligen kan inte reliabla jämförelser göras mellan utprovningar och skarpt läge beträffande test-tagarnas uppfattningar och resultat. Detta måste beaktas vid läsningen av de resultat som presenterats i denna studie. Så länge användandet av provmaterialet för moderna språk inte är obligatoriskt på skolorna och enbart erbjuds som bedömningsstöd, inte som nationella prov, är det svårt att på ett tydligare sätt kartlägga såväl elevers som lärares uppfattningar samt elevers resultat vid *high stakes*-prov, detta eftersom Skolverket hittills inte gjort insamlingar av provresultat för moderna språk.

I den *Internationella språkstudien* (Skolverket, 2012) framträder också en motsägelsefull bild av svenska elevers resultat i engelska och spanska. Den andel som inte nådde upp till A1 på GERS-skalan var knappt synlig i engelska (knappt 1,5 procent) medan det i spanska var mellan 24–45 procent som inte nådde A1 i de tre färdigheter som testades (läsa, höra och skriva). Vid en närmare titt på könsskillnaderna framträder samma mönster som i denna studie, d.v.s. att flickorna presterar signifikant bättre än pojkarna vad gäller både läsa och höra och påtagligt bättre vad gäller skriva. Det finns många skäl till att eleverna presterar sämre i spanska, men förenklat kan man hävda att i Sverige är engelska och spanska, såväl inom den svenska skolan som utanför, något av varandras motsatser. Spanska är ett valbart ämne som börjar studeras som tidigast i klass 6 eller senare, har låg exponering utanför ämnets ram, få föräldrar är kunniga i språket, det upplevs som mindre meningsfullt att lära sig och har låg status. Engelska är däremot ett obligatoriskt ämne som börjar studeras i relativt tidig

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

ålder, många föräldrar är kunniga i språket, det upplevs som mycket meningsfullt att lära sig och får därmed hög status. Engelska har en mycket hög exponering utanför skolans värld, inte minst i olika media, s.k. *extra-mural English*, vilket är ett fenomen som har definierats och utförligt beskrivits av Pia Sundqvist och Liss Kerstin Sylvén i *Extramural English in teaching and learning. From theory and research to practice* (Sundqvist & Sylvén, 2016). I Lena Börjessons och Sofia Nilssons text *Könsskillnader i engelsk språkfärdighet* i denna rapport beskrivs hur exponering för *extra-mural English* tycks ha en mer positiv effekt på pojkars utveckling av ordförråd än på flickors. I ljuset av detta är det speciellt intressant att jämföra resultaten i spanska med pojkars goda resultat på de receptiva delarna i de engelska nationella proven.

Vilka skillnader framträder mellan pojkars och flickors uppfattning om hur väl de klarat uppgiften jämfört med deras resultat på varje provuppgift?

Jag tror att jag har klarat uppgiften bra

Jämfört med föregående påstående är könsskillnaderna betydligt större för det som rör retrospektiv, uppgiftsrelaterad självskattning, som också torde vara kopplat till uppgifternas upplevda eller faktiska svårighetsgrad, men också till självförtroende och/eller mer eller mindre realistisk uppfattning om sin egen förmåga. Vad gäller hörförståelse tycks pojkarna generellt ha mycket större tilltro till den egna förmågan än vad flickorna uppvisar. Om man jämför siffrorna på elevernas retrospektiva skattning av sina prestationer med dessa gruppers P-värden kan man se att flickorna i högre grad klarat uppgifterna bättre än pojkarna, (detta gäller utprövningens alla delar) trots att de i lägre grad tror sig ha klarat uppgifterna bra. Ibland tenderar flickorna att underskatta sin egen förmåga, ibland visar de en mera realistisk syn på sin egen förmåga än pojkarna. Könsskillnaderna syns framför allt och ännu tydligare bland dem som svarat positivt, *Ja, absolut* eller *Ja* (alt. 1–2). Pojkarnas P-värden per uppgift vad gäller hörförståelse är alltid lägre än flickornas, och med enstaka undantag gäller detta även läsförståelseuppgifterna.

Att pojkar uppvisar något större självförtroende än flickor visar sig vad gäller skattning av den egna förmågan efter genomförande av prov, vilket också bekräftas av lärare. Det är intressant att ställa sig frågan varför flickor tenderar att retrospektivt sett underskatta sin prestation, medan det för pojkar är tvärtom. Enligt Erickson och Åberg-Bengtsson (2012) kan man konstatera att detta är

ATT BEDÖMA SPRÅKLIG KOMPETENS

relativt konsekvent, oberoende av färdighetsnivåer, med undantag av specifika uppgifter. Att pojkar i detta avseende tycks ha större tilltro till sin förmåga är ett fenomen som noterats tidigare och det visar sig också i de fall då flickors resultat är påtagligt bättre. Det är dock viktigt att ställa sig frågan vad eleverna egentligen svarar på, t.ex. om de ger uttryck för sin egen uppfattning eller om de försöker gissa vad andra, kanske inte minst läraren, anser (Erickson, 1998). Vad betyder det att ha klarat uppgiften bra? En del elever anser det kanske vara tillräckligt att ha klarat hälften (= godkänt?) medan andra strävar högre än så och jämför sina prestationer på s.k. *proficiency tests* med de resultat de brukar få på prov och läxförhör, där de fått möjlighet att förbereda sig, s.k. *achievement tests*.

Wernersson (SOU 2010:51) refererar även till Jakobsson (2000) beträffande motivation relaterad till akademisk självvärdering, de förväntningar man har på sina egna prestationer. Den samlade bilden visar att flickors och pojkars akademiska självvärdering är kopplad till genusmärkningen av ämnen, men, skriver Jakobsson:

”I de fall flickorna har mindre positiv värdering av sin kompetens verkar det dock inte påverka deras prestationer negativt.” (Jakobsson, 2000, s. 54).

Jakobsson avslutar sin avhandling med ett konstaterande som skulle kunna vara *ett* av svaren på frågan varför flickor lyckas bättre men inte litar till den egna förmågan i samma utsträckning som pojkarna:

”Paradoxalt nog skulle vi kunna säga att det är just på grund av sin relativa underordning, som flickor får högre betyg i skolan. Man kan uttrycka det som att flickor kompenserar en ”brist” de i realiteten inte alltid har (föreställningen om att de inte är logiska och smarta) och därför blir duktigare genom att anstränga sig mer. Pojkar klarar sig inte lika bra för att de litar på en förmåga (föreställningen att de är logiska och smarta) de inte alltid har. Då tar man det mer lättsamt och presterar följaktligen sämre.” (Jakobsson 2000, s. 208)

Som nämnts tidigare analyseras i ***L3 Motivation*** (Henry, 2012) bl.a. påverkan av självbilden för L2 engelska på L3 motivation. Kanske kan det vara så att en hög grad av jämförelser mellan självbilder för L3–L2 gör att eleverna som besvarat enkäterna både upplever de spanska uppgifterna som *mycket svåra* jämfört med engelska på samma nivå och därmed också värderar sina egna prestationer lågt. Det är i så fall troligt att många elever känner en viss uppgivenhet vad gäller spanskstudierna och resultaten, vilket också speglas av de många avhoppet från spanska mellan åk 7 och 9 i grundskolan. Av de 46 612 som läste spanska i åk 7 i oktober 2014 återstod 40 794 i oktober 2016 (Skolverket, 2017). 12 procent

TRE TEXTER OM TEST-TAKER FEEDBACK FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

av eleverna har följaktligen hoppat av sina spanskstudier och då inkluderas inte de eventuella ytterligare avhopp som kan komma att ske under resten av läsåret 2016/2017.

Avslutande reflektion

De nationella bedömningsstöden i moderna språk hjälper lärarna i hela landet att på ett likvärdigt sätt pröva och bedöma de fyra färdigheterna: höra, läsa, tala, skriva. Vi hoppas att materialen har en *washback*-effekt på undervisningen – att man för att eleverna ska lyckas bra kontinuerligt arbetar med både receptiva och produktiva kompetenser, t.ex. genom att eleverna både får höra och tala mycket spanska i klassrummen.

Avslutningsvis är vår förhoppning att även spanska (som är det mest populära valet bland moderna språk i svensk skola) liksom engelska framöver kommer att upplevas som mycket meningsfullt att lära sig och därmed få högre status som skolämne, både bland flickor och pojkar. För att åstadkomma en utveckling som går i en positiv riktning bör vikt läggas vid att tydligare kommunicera värdet av att behärska ett tredje språk. En sådan utveckling, anser vi, förutsätter en samlad kraftansträngning från både lärare, skolledare och myndigheter. Och inte minst viktigt är det att den politiska nivån genom sina beslut visar på vikten av ökade språkkunskaper, detta genom att göra det obligatoriskt att studera ett modernt språk i grundskolan.

Referenser

- Alderson, C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. & Palmer, A. (1996). *Language Testing in Practice – Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bolger, N. & Kellaghan, T. (1990). Method of Measurement and Gender Differences in Scholastic Achievement. In *Journal of Educational Measurement*, volym 27, No 2.
- Brown, D. H. (2007). *Teaching by Principles, An Interactive Approach to Language Pedagogy (Third Edition)*. Pearson Longman.
- Erickson, G. (1996). *Lärare och elever tycker – om standardprovet i engelska. Redovisning av två enkätundersökningar*. Göteborg: Göteborgs Universitet, Institutionen för pedagogik.
- Erickson, G. (1998). Test-taker feedback – Elevers synpunkter som del av arbetet med nationella prov i engelska. I B. Ljung & A. Pettersson, (red.),

ATT BEDÖMA SPRÅKLIG KOMPETENS

- Perspektiv på bedömning av kunskap* (s. 79–100). Lärarhögskolan i Stockholm, Institutionen för pedagogik.
- Erickson, G. (2006). Bedömning av och för lärande: En kollaborativ ansats i arbetet med nationella prov i språk. I U. Tornberg (red.) *Mångkulturella aspekter på språkundervisningens kommunikativa praktiker* (s. 187–207). Örebro: Örebro Universitet, Pedagogiska institutionen.
- Erickson, G. (2010). Good Practice in Language Testing and Assessment – A Matter of Responsibility and Respect. In T. Kao & Y. Lin (Ed.) *A new Look at Language Teaching and Testing: English as a Subject and Vehicle* (s. 237–258). Selected papers from the 2009 LTTC International Conference on English Language Teaching and Testing. March 6–7. Taipei, Taiwan: Bookman Books Ltd.
- Erickson, G. & Åberg-Bengtsson, L. (2012). A Collaborative Approach to National Test Development. In D. Tsagari & I. Csépes, (Ed.), *Language Testing and Assessment*. Frankfurt am Main: Peter Lang.
- Göteborgs universitet: Projektet nationella prov i främmande språk (NAFS). Hämtat 2018-05-10, från <http://nafs.gu.se/>
- Henry, A. (2012). *L3 Motivation*. Göteborgs universitet, Institutionen för pedagogik och specialpedagogik.
- Jakobsson, A.K. (2000). *Motivation och inläring ur ett genusperspektiv*. Göteborgs universitet, Institutionen för pedagogik och didaktik.
- Messick, S.A. (1989). Validity. I R. L. Linn (red.), *Educational Measurement* (Third edition, s. 13–103). New York: American Council on Education/Macmillan.
- Skolverket (2004). *Engelska i åtta europeiska länder – en undersökning av ungdomars kunskaper och uppfattningar*. Stockholm: Skolverket, Rapport.
- Skolverket (2007). *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. Stockholm: Fritzes.
- Skolverket (2011). *Kursplan i moderna språk*. Hämtad 2017-04-24, från <https://www.skolverket.se/undervisning/grundskolan/laroplan-och-kursplaner-for-grundskolan/>
- Skolverket (2012). *Internationella språkstudien 2011*. Stockholm: Fritzes.
- Skolverket (2016). *Moderna språk inom ramen för språkval i årskurs 6, 7, 8 och 9 läsåren 2012/13–2015/16*. Hämtad 2018-05-10, från https://www.skolverket.se/statistik_och_utvardering/statistik_i_tabeller/grundskola/skolor_och_elever/skolor_och_elever_i_grundskolan_lasaret_2015_16_1.241617
- Skolverket (2017). SIRIS. Kvalitet och resultat i skolan. Hämtad 2018-05-10, från http://sir.is.skolverket.se/siris/ris.elever_gr.mod_sprak
- Stiggins, R. J. (1987). The design and development of performance assessments. *Educational Measurement: Issues and Practice*, volym 6, 33–42.

TRE TEXTER OM TEST-TAKER FEEDBACK
FRÅN UTPRÖVNINGAR I MODERNA SPRÅK

- Sundqvist, P. & Sylvén, L. K. (2016). *Extramural English in teaching and learning. From theory and research to practice*. London: Palgrave Macmillan.
- Wernersson, I. (2010). *Könnskillnader i skolprestation – idéer om orsaker (SOU 2010:51)*. Stockholm: Utbildningsdepartementet.
- Wiggins, G. P. (1993). *Assessing student performance*. San Francisco: Jossey Bass Publishers.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Könsskillnader i engelsk språkfärdighet

Lena Börjesson & Sofia Nilsson

Projektet Nationella prov i Främmande Språk (NAFS) vid Göteborgs universitet utvecklar nationella prov och bedömningsstöd i engelska, franska, spanska och tyska på uppdrag av Skolverket. I samband med prov samlas data in dels av Statistiska centralbyrån (SCB), dels av projektet NATIONELLA prov i Främmande Språk (NAFS) som också samlar in data i samband med storskaliga utprovningar. Dessa data möjliggör analyser och fördjupningar av olika slag och ger möjlighet till analys av utfall och tendenser. Bl.a. undersöks hur väl olika elevgrupper lyckas på de olika delproven och uppgifterna. En av de tendenserna är att pojkar och flickor presterar ungefär lika bra till och med nivån Engelska 5 på gymnasiet. I kurserna Engelska 6 och 7 kan man däremot observera vissa skillnader mellan manliga och kvinnliga elevers resultat. Den här rapporten avser att beskriva en bild som framträder, sätta in den i ett sammanhang och diskutera möjliga orsaker.

Allmänt om ämnesproven och kursproven i engelska

De nationella proven i engelska utvecklas på uppdrag av Skolverket vid Göteborgs universitet inom projektet NATIONELLA prov i Främmande Språk (NAFS). Proven är s.k. *proficiency tests* som syftar till att pröva och ge en bild av elevernas allmänna språkfärdighet oberoende av var, när och hur denna färdighet förvärvats. Planeringen och konstruktionen av proven baseras på analyser av de nationella styrdokumenterna och på nationell och internationell forskning kring språkinlärning och bedömning av språkfärdighet sammanfattade i ett antal gemensamma principer för de nationella provmaterialen i språk.

De olika delarna i proven, *Focus: Reading*, *Focus: Listening*, *Focus: Speaking* och *Focus: Writing* anknyter också till Europarådets ramverk för språkinlärning och språkbedömning, *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning* (Council of Europe/Skolverket, 2009) med dess rubriceringar reception, interaktion och produktion. Detta dokument genomsyras av en handlingsorienterad språksyn, och de svenska kurs- och ämnesplanerna i språk har en tydlig koppling till detta internationellt betydelsefulla dokument. Nämnas kan

att svenska styrdokument i språk alltsedan tidigt 80-tal präglats av denna kommunikativa och funktionella språksyn.

Arbetet med proven sker i en kollaborativ process där ett stort antal elever och lärare deltar och lämnar värdefulla synpunkter, vilket bidrar till provens kvalitet och validitet (se Gudrun Ericksons text *Bedömning av språklig kompetens* samt Erickson & Åberg-Bengtsson, 2012). De olika delarna i respektive prov prövas ut i stor skala på olika program vid slumpvis valda skolor i hela landet. Under arbetet med att ta fram ett kurs- eller ämnesprov sammanträder referensgrupper vid flera olika tillfällen.

I det följande kommer fokus att ligga på resultaten på de receptiva delproven, d.v.s. de delprov som prövar förståelse av talat och skrivet språk. En ambition vid konstruktionen av de receptiva delproven i språk är att proven ska vara så allsidiga som möjligt beträffande både innehåll och svarsformat för att ge elever med olika språkfärdighetsprofiler möjlighet att komma till sin rätt. Det innebär t.ex. att det finns uppgifter med flervalsalternativ eller av matchningskaraktär och uppgifter där eleverna ska ge kortfattade svar med egna ord. Noggranna överväganden görs för att undvika *bias*, d.v.s. att vissa elever systematiskt gynnas eller missgynnas. Som underlag för uppgifter används texter och inspelningar från olika delar av världen där engelska används, och vid utveckling och sammansättning av proven tas hänsyn till ett antal parametrar som t.ex. genus, kultur, livsstils-, ålders- och identitetsrelaterade aspekter. Innan uppgifterna kommer in i det slutliga provet prövas de ut i olika versioner med en s.k. ankaruppgift i varje version. Denna används för att ge en indikation om sammansättningen av utprövningsgruppen, så att man kan göra en bedömning av svårighetsgraden på uppgifterna.

Insamlingar av resultat

Totalinsamlingar av resultat från de nationella proven görs årligen. SCB:s insamlingar ger information om provbetyg, kursbetyg samt resultat på delprovsnivå, däremot inte på uppgifts- och *item*-nivån.

Som ett komplement för att kvalitetssäkra och utveckla proven samlar NAFS-projektet vid Göteborgs universitet efter vårterminsprovets genomförande in dels ca 700–800 resultatprofiler, dels hela provhäften för elever födda vissa datum. De cirka 300 hela provhäftena medger kvalitativa och kvantitativa analyser av olika uppgifter, enskilda elevsvar och elevtexter.

Manliga och kvinnliga elevers resultat

Resultaten från de nationella proven i engelska i årskurs 6 och årskurs 9 visar generellt inga skillnader mellan flickors och pojkars provbetyg. Inte heller i Engelska 5, den första kursen i engelska på gymnasiet, finns några egentliga skillnader mellan kvinnliga och manliga elevers provbetyg (se vidare Skolverkets Statistik och utvärdering och resultatredovisningar för respektive prov).

Engelska 5 är en gymnasiegemensam kurs, d.v.s. kursen är obligatorisk för alla elever. Engelska 6 är obligatorisk för de högskoleförberedande programmen och valbar för elever på yrkesprogrammen. En genomgående observation, om man ser till utfallet på kursprovet i Engelska 6, är att de manliga eleverna lyckas bättre framför allt på de receptiva uppgifterna, det vill säga på läs- respektive hörförståelse. Redan i en rapport om utfallet på kursprovet för Engelska B (= Engelska 6 efter läroplansrevideringen 2011) för höstterminen 2001, alltså några år efter införandet av de nationella proven, står följande: *Som framgår av figur 4.2.1 lyckas männen bättre på de receptiva färdigheterna i kursprovet medan kvinnorna har bättre resultat på de produktiva. I provbetyget är betygsfördelningen mellan könen ganska jämn, men vad gäller kursbetyget får kvinnorna totalt sett något bättre betyg än männen.* (Skolverket, 2002). I en rapport från 2013 (Skolverket, 2013) ges en tillbakablick på resultaten från kursproven i Engelska A och B från åren 2001–2012). Analysen av kursproven i Engelska B handlar om könsskillnader, och denna rapport ligger delvis till grund för diskussionen i detta kapitel. Den bygger dels på deskriptiva analyser utifrån lösningsfrekvenser (medelvärde uttryckt i procent) och spridningsmått på olika delprov och uppgifter, dels på SCB-data angående andelen elever som fått de olika betygen på de olika delproven samt provbetyg och kursbetyg. Det är naturligtvis möjligt att analysera resultaten även med hjälp av andra metoder, men deskriptiva analyser enligt ovan ger en bra bild av hur väl elever klarar av olika uppgifter.

Tendensen beträffande de receptiva delproven kvarstår 2016, men könsskillnaderna angående de produktiva delproven d.v.s. *Focus: Speaking* och *Focus: Writing* är numera små, och de manliga och kvinnliga eleverna presterar ofta på ungefär samma nivå. Det innebär att männen genomgående har ett bättre sammanvägt provbetyg.

För att stödja en likvärdig och rättvis bedömning och betygssättning har även ett bedömningsstöd för Engelska 7 utvecklats på Skolverkets uppdrag. Kursen i Engelska 7 är valbar men ger meritpoäng vid antagning till högskolestudier. Det första kursprovet blev tillgängligt i mars 2016. Provet omfattas av

sekretess, men eftersom kursen är valbar är provet inte obligatoriskt. Provet innehåller liksom de nationella proven i engelska på lägre nivåer olika delprov som avser att pröva reception av talat och skrivet språk samt muntlig och skriftlig interaktion och produktion. Eftersom kursen är valbar görs inga slumpvisa insamlingar av elevresultat av SCB. Däremot har NAFS-projektet gjort egna insamlingar av resultat. Då alla elever som gått kursen inte genomfört provet och resultatinsamlingarna bygger på frivillighet måste resultaten tolkas med försiktighet. Dock är tendensen tydlig: de manliga eleverna lyckas bättre på alla uppgifter i provet och får således ett bättre provresultat, särskilt beträffande läs och hörförståelse men även vad gäller muntlig och skriftlig produktion och interaktion, och enligt de inkomna resultaten är könsskillnaderna till och med större än i proven i Engelska 6. Se rapport för kursprovet för Engelska 7, vårterminen 2016 ([Börjesson, 2016](#)).

Resultat på enskilda uppgifter

Beträffande kursproven i Engelska 6 (tidigare B) kan konstateras att de manliga eleverna presterar bättre på så gott som samtliga uppgifter i de receptiva delproven. Könsskillnaderna är minst beträffande de långa läsförståelsetexterna och de långa hörförståelseuppgifterna. Dessa har ett blandat svarsformat med både flervalsfrågor och öppna svar i vilka eleven får svara med ett eller flera ord. Störst är skillnaden mellan kvinnliga och manliga elevers resultat på korta, informationstäta uppgifter med flervalsalternativ eller matchning. Exempel på sådana uppgifter är t.ex. läsförståelseuppgifterna *Mini-texts* och *Find the Context*, och beträffande hörförståelse kan nämnas uppgifter av typen *News Items* eller *What's the Topic?*. Dessa uppgifter är ofta hämtade från nyhetsartiklar eller reportage och har i regel en ganska avancerad vokabulär. I dessa uppgifter är ett stort ordförråd troligen mycket utslagsgivande. Sambandet mellan god läsförståelse och ett stort ord- och frasförråd är väl dokumenterat i forskning (se forskningsöversikt bl.a. i Grabe, 2009).

En ytterligare svårighet tillkommer i hörförståelse och då främst beträffande de ovan nämnda korta, informationstäta inslagen av typen nyheter, där vokabulären är avancerad. Då krävs dessutom att ordförrådet är automatiserat (Buck, 2001). Kraven på en automatiserad förmåga ställs på sin spets i ett hörförståelseprov, där den som lyssnar inte kan kontrollera bearbetningen av inflödet av språk på samma sätt som i ett läsförståelseprov, där läsaren kan gå tillbaka i texten och läsa om valda delar. Automatiseringen gäller alla delar av

språket, som t.ex. en persons ordförråd; högfrekventa ord har kanske hunnit bli automatiserade medan lågfrekventa ord ännu inte har blivit det. Detsamma gäller förtrogenheten med komplicerade syntaktiska konstruktioner eller en viss accent eller dialekt. När elever tycker att någon talar för snabbt beror det i regel på bristande automatiserad förmåga att ta till sig det man hör, vilket i sin tur hänger ihop med färdighetsnivån. Buck hänvisar till studier som har visat att en komplex struktur tillsammans med en avancerad vokabulär, s.k. *cognitive load*, bidrar till att stora krav ställs på den kognitiva förmågan och att det för många kan vara en betydande svårighet i framför allt hörförståelse.

I proven ingår ibland också lucktexter som prövar förmågan att förstå sammanhang. Eleverna läser en text där enstaka ord saknas och fyller i de ord som passar i sammanhanget. De ord som ska fyllas i är ofta frekventa ord och själva texten är i regel inte avancerad. På något undantag när har de manliga eleverna under åren haft bättre resultat även på dessa uppgifter. Ännu något större är skillnaden för lucktexter med flervalsalternativ.

Vad som ovan sagts om resultaten i Engelska 6 gäller i minst lika hög grad för det första kursprovet för Engelska 7, vårterminen 2016. Särskilt tydlig är skillnaden på det receptiva delprovet. Nästan dubbelt så stor andel män som kvinnor har här betygen A och B; 27 procent av männen når de två högsta betygsstegen, medan bara 14 procent av kvinnorna når A eller B. I andra änden av skalan har 14 procent av kvinnorna betyget F men bara 5 procent av männen har detta betyg. Återigen måste här dock påpekas att provet inte är obligatoriskt och att SCB inte samlar in resultat, utan data bygger på de drygt 3 300 resultat som lärare skickat in på uppmaning av NAFS-projektet.

Möjliga förklaringar till skillnaden i resultat

Frågan är då varför könsskillnaderna ökar på de högre nivåerna i engelska, d.v.s. Engelska 6 och 7, och varför de manliga eleverna presterar bättre.

Kursernas innehåll – andra domäner?

I Engelska 6 är det innehåll som behandlas mera abstrakt och komplext än på lägre nivåer, eftersom det enligt ämnesplanen i större utsträckning än på lägre nivåer relateras till ämnen som rör samhällsfrågor, yrkesliv och studier och mindre till den personliga domänen. För Engelska 7 tillkommer dessutom att innehållet kan vara mer teoretiskt. Enligt ämnesplanen ska kommunikationens

innehåll i Engelska 7 bland annat relatera till ”teoretiska och komplexa ämnesområden”, vilket kan jämföras med ”konkreta och abstrakta ämnesområden” i Engelska 6 och ”aktuella områden” i Engelska 5. Ett rimligt antagande är att innehållet i de högre kurserna sammanfaller med en mer komplicerad meningsstruktur och en mer avancerad vokabulär.

I resultatredovisningarna från provet i Engelska 6 har under årens lopp orsaker till de manliga elevernas genomgående bättre resultat på de receptiva delproven diskuterats. En orsak som har lyfts fram är att det kan bero på att de ofta har ett bredare och mer avancerat ordförråd, och frågan har ställts om detta i sin tur kan härledas till aktiviteter utanför skolan. Att det bland gymnasieelever i Sverige finns könsskillnader beträffande ordförråd till männens fördel har visats bl.a. i en studie angående utvecklingen av ordförrådet hos CLIL⁴-elever på gymnasienivå i jämförelse med elever som enbart läste engelska som ett separat ämne (Sylvén, 2004). En av slutsatserna var att de elever som läser engelska texter utanför skolan, oavsett om de går i CLIL-klasser eller inte, var de som utvecklade sitt ordförråd mest och att de manliga eleverna hade ett större ordförråd än de kvinnliga. I ett senare projekt genomförde elever på gymnasienivå flera ordförrådstester och även där lyckades manliga elever bättre, vilket diskuteras i en artikel från 2014 (Sylvén & Ohlander).

I en doktorsavhandling från Karlstad universitet (Sundqvist, 2009) visas ett positivt samband mellan den totala tid ungdomar i grundskolans årskurs 9 ägnar sig åt ”fritidsengelska” och deras muntliga färdighet i engelska samt storleken på deras engelska ordförråd. Enligt Sundqvist påverkas pojkarnas resultat i engelska mer positivt av fritidsengelska än flickornas. Pojkarna ägnar mer tid åt engelska på fritiden och då framför allt åt läsning, dataspelande och surfande på Internet. Sådana sysselsättningar verkar ha en större positiv inverkan på ordförrådet än aktiviteter som att lyssna på musik och titta på film och TV, vilket flickorna i studien i större utsträckning föredrar.

I ytterligare studier, redovisade bl.a. av Sundquist och Sylvén (2012), har visats att interaktiva dataspel, som i större utsträckning spelas av pojkar, har en positiv inverkan på språkfärdigheten och i synnerhet ordförrådet.

⁴ Content and Language Integrated Learning

Selektionseffekt?

Frågan har också ställts om en av förklaringarna till att resultatskillnaderna mellan könen är större i Engelska 6 än i Engelska 5 och grundskolan skulle kunna vara att de elever som läser Engelska 6 utgör en selekterad grupp.

I **grundskolan** är engelska ett obligatoriskt ämne och de nationella ämnesproven i engelska årskurs 6 och 9 är obligatoriska att genomföra.

I **gymnasieskolan** är, som nämnts ovan, kursen Engelska 5 obligatorisk på samtliga program och läses alltså av alla gymnasieelever. Engelska 6 är behörighetsgivande för studier på högskolenivå och kursen är därför obligatorisk på högskoleförberedande program d.v.s. de estetiska, naturvetenskapliga, samhällsvetenskapliga och tekniska programmen. För yrkesprogrammen är Engelska 6 en valbar kurs. Engelska 7 är en valbar kurs för samtliga program. Liksom i grundskolan är de nationella kursproven för Engelska 5 och 6 obligatoriska att genomföra. Då kursen Engelska 7 är valbar är kursprovet inte obligatoriskt.

Tabell 1: VT16 – Elever med betyg i engelska och provbetyg i engelska i gymnasieskolan Engelska 5, 6 och 7 (data Skolverket/SCB)

Kurs	Antal elever med kursbetyg i resp. kurs	Antal elever med provbetyg	Varav kvinnor	Varav män
Engelska 5	80 179	75 617	36 969	38 648
Engelska 6	68 119	57 015	29 402	27 613
Engelska 7	35 400	i.u.*	i.u.*	i.u.*

NB: För att få ett provbetyg på kursprovet behöver eleven ha deltagit i samtliga delprov.

* Ingen uppgift. SCB samlar inte in data då kursprovet i Engelska 7 inte är obligatoriskt.

Som framgår av tabellen ovan sker en selektion som blir större för varje kurs i engelska inom gymnasieskolan.

Bortsett från att färre elever läser Engelska 6 än 5, och ytterligare färre läser engelska 7, hur ser selektionen ut? Väljer vissa elevgrupper bort engelska i högre grad än andra?

I ett försök att bilda sig en uppfattning kan man titta på de elever som fick betyg i Engelska 7 vårterminen 2016 och gå bakåt i tiden. Hur har fördelningen på kurser och program varit för den årskursen under deras gymnasietid 2014–2016? Det finns dock flera felkällor, t.ex. att elever byter mellan program, avbryter och återupptar sina gymnasiestudier och att en del läser kursen på en termin, men det kan ge en viss överblick över vilka elever som läst vilka kurser under den aktuella perioden.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Tabell 2: Antal elever med kursbetyg i respektive kurs fördelat på typ av program (data Skolverket/SCB)

Kurs – år	Antal elever med kursbetyg totalt	Varav högskolef. prog.		Varav yrkesprog.	
Engelska 5 – 2014	84 013	53 120	63%	30 893	37%
Engelska 6 – 2015	70 405	54 505	77%	15 900	23%
Engelska 7 – 2016	35 400	34 504	97%	896	3%

Tabell 3: Antal elever med kursbetyg i respektive kurs fördelat på kön (data Skolverket/SCB)

Kurs – år	Antal elever med kursbetyg totalt	Varav kvinnor		Varav män	
Engelska 5 – 2014	84 013	41 702	50%	42 311	50%
Engelska 6 – 2015	70 405	37 505	53%	32 900	47%
Engelska 7 – 2016	35 400	18 775	53%	16 625	47%

Tabell 4: Antal elever med kursbetyg fördelat på kön och yrkes- respektive högskoleförberedande program. Uppgifterna i procent är andel av samtliga elever med kursbetyg det året. (data Skolverket/SCB)

Kurs – år	Antal kvinnor med kursbetyg yrkesprog.		Antal män med kursbetyg yrkesprog.		Antal kvinnor med kursbetyg högskolef. prog.		Antal män med kursbetyg högskolef. prog.	
Engelska 5 – 2014	12 994	15%	17 899	21%	28 708	34%	24 412	29%
Engelska 6 – 2015	7 777	11%	8 123	12%	29 728	42%	24 777	35%
Engelska 7 – 2016	508	1%	388	1%	18 267	52%	16 237	46%

Efter Engelska 5 tycks män i något mindre utsträckning än kvinnor fortsätta att läsa engelska. Ett möjligt antagande är att manliga elever på yrkesprogram i lägre grad väljer att fortsätta med Engelska 6 än kvinnliga elever i samma situation. En faktor skulle kunna vara att det finns elever på t.ex. det kvinnligt dominerade Vård- och omsorgsprogrammet som för att uppnå högskolebehörighet till sjuksköterska-, arbetsterapeut- eller logopedutbildningar och liknande måste läsa Engelska 6. Å andra sidan skulle situationen kunna vara liknande för elever på t.ex. det manligt dominerade El- och energiprogrammet om högskolebehörighet krävs för vidare studier.

Som framgår av tabell 2 är andelen elever på yrkesprogram som väljer att läsa Engelska 7 mycket liten jämfört med andelen elever från högskoleförberedande program.

Sammanfattningsvis är det utifrån tillgängliga data från Skolverket/SCB svårt att dra några egentliga slutsatser om betydelsen av en eventuell selektionseffekt.

KÖNSSKILLNADER I ENGELSK SPRÅKFÄRDIGHET

Ytterligare ett sätt att undersöka en eventuell selektionseffekt skulle kunna vara att titta på den ankaruppgift som alltid används vid utprövningar av provuppgifter. I Engelska 6 används en ankaruppgift som består av 12 meningar med luckor. Den har använts i utprövningar sedan 2000 och genomförts av drygt 25 000 elever. Sex av dessa luckor användes också i utprövningarna 2015 och 2016 i Engelska 7. I ankaruppgiften för Engelska 7 finns alltså 6 uppgifter från Engelska 6 och 6 nykonstruerade. I de senare är kontexten mer formell.

Tabell 5: Lösningfrekvens gemensamma ankaruppgifter Engelska 6 och 7

Kurs	Lösningfrekvens kvinnor (p.value)	N	Lösningfrekvens män (p.value)	N
Eng 6	,59	13 218	,67	11 945
Eng 7	,74	1 554	,81	1 368

Som framgår av tabell 5 stiger lösningfrekvensen i Engelska 7 markant jämfört med Engelska 6 för de sex gemensamma *items* för både män och kvinnor. Skillnaden mellan mäns och kvinnors resultat kvarstår dock, vilket inte stödjer tanken på att en selektionseffekt skulle kunna vara orsaken till de skillnader i resultat mellan män och kvinnor som kan observeras i Engelska 6 och 7.

Frågor som behöver diskuteras mer

Det är, utifrån forskning som redovisats ovan, rimligt att anta att könsskillnaden på de högre nivåerna i engelska till en betydande del kan relateras till skillnader i ordförråd och vana att läsa och tillgodogöra sig mer teoretiska och avancerade texter och mer komplext talat språk. Frågan är om alla elever är fullt medvetna om att de högre nivåerna, enligt centralt innehåll och kunskapskrav, ställer större krav på förmågan att förstå och uttrycka sig i mer komplexa och formella sammanhang, och att det för detta krävs en bredare och djupare kunskap om ord och språkliga strukturer samt en högre grad av precision?

Svenska elever har goda kunskaper och färdigheter i engelska, vilket har bekräftats i internationella undersökningar senast i *The European Survey on Language Competences (ESLC)*, på svenska *Internationella språkstudien*, som genomfördes 2011 ([Skolverket, 2012](#)). Studien visade att svenska niondeklassares kunskaper i engelska är mycket goda, och Sverige placerade sig i topp bland de 13 deltagande länderna. I studien ingick också en enkät som visade att svenska elever har en positiv inställning till engelska, och de tycker också i stor utsträckning att språket är lätt att lära sig.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Men hur länge kan man leva på gamla meriter? Finns det en risk att många elever är nöjda med sin nivå i engelska och tycker att kunskaperna räcker till utan att de behöver anstränga sig så mycket? I ett mail i samband med det nya kursprovet för Engelska 7 skrev en lärare: ”*Det är en kamp att synliggöra progressionen mellan kurserna 5–7. För kurs 5 räcker det att ha kollat på några serier typ ”Vänner” eller ”How I met your mother”. Kurs 6 klarar många med ett nödrop och sedan ska de gå kurs 7.*” En annan lärare gjorde i lärarenkäten följande reflektion: ”*Resultaten tyder på att eleverna rent allmänt bör läsa mer, inte minst tyngre texter, där det krävs viss analys för att uppfatta det som egentligen sägs.*” Kanske måste dessa frågor diskuteras mera i lärarutbildningarna och i klassrummen?

En del elever tillägnar sig troligen i viss utsträckning de kunskaper som krävs för en högre nivå i engelska utanför skolan, och som framgått ovan tycks detta framför allt gälla manliga elever. Dock behöver alla elever nå den nivå som krävs för att tillgodogöra sig texter och talat språk där innehållet kan vara komplext och abstrakt, ordförrådet relativt avancerat och meningsbyggnaden mer komplicerad. Det är i första hand en didaktisk fråga såtillvida att det är angeläget att lärare och elever diskuterar hur man kan gå tillväga för att utveckla dessa förmågor. Det innebär inte att de kvinnliga eleverna måste börja spela interaktiva dataspel i större utsträckning, även om det verkar vara ett av sätten att utveckla sin språkliga förmåga. Dock torde motivation vara en grundläggande faktor. Alla lärare är naturligtvis medvetna om att motivation är avgörande när det gäller att lära sig, och det finns många studier och teorier att luta sig mot. I en översikt i Sundquist & Sylvén (2016) redovisas t.ex. flera studier som visar att det gagnar både lärare och elever och ökar elevernas motivation om läraren tar reda på mer om elevers olika engagemang utanför skolan och visar att dessa kan vara relevanta utgångspunkter för ett resonemang om hur man kan förbättra sin språkliga förmåga.

Bland material som finns att tillgå via Skolverkets webbplats för att hjälpa elever att bli mer medvetna om sitt lärande och reflektera över sin språkinläring och för att ge uppslag till diskussion kan nämnas bl.a. [Europeisk språkportfolio](#) för grundskolan och [Europeisk språkportfolio 16+](#) för gymnasiet. För gymnasiet finns också ett [självbedömningsmaterial](#) (Skolverket, 2014) med olika checklistor. Ett annat sätt kan vara att eleverna under ett par veckors tid för dagbok över vilka fritidsaktiviteter de är engagerade i som har med engelska att göra och funderar över vilken typ av *input* (och eventuell *output*) dessa involverar.

Det är angeläget att följa kunskapsutvecklingen i engelska, inte minst manifesterad genom betyg och resultat på de nationella proven, eftersom det naturligtvis är viktigt att så många elever som möjligt når en nivå i engelska som är i linje med kunskapskraven i de högre kurserna på gymnasiet och som är tillräckligt hög för att eleverna ska få bra möjligheter att klara framtida studier.

Referenser

- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Börjesson, L. (2016). *Rapport om kursprovet i Engelska 7 vårterminen 2016*. Hämtad 2018-05-10 från <http://nafs.gu.se/publikationer>
- Council of Europe/Skolverket. (2009). *Gemensam europeisk referensram för språk: lärande, undervisning, bedömning*. Hämtad 2018-05-10 från www.skolverket.se/publikationer?id=2144
- Erickson, G. & Åberg-Bengtsson, L. (2012). A Collaborative Approach to National Test Development. I Tsagari, D. & Csépes, I. (red.), *Collaboration in Language Testing and Assessment* (s. 93–108). Frankfurt am Main: Peter Lang.
- European Commission/SurveyLang (2012). *First European Survey on Language Competences: Final Report*. Hämtad 2018-05-10 från http://ec.europa.eu/dgs/education_culture/repository/languages/policy/strategic_framework/documents/language_survey_final_report_en.pdf
- Europarådet/Skolverket (2004). Europeisk Språkportfolio 12–16 år. Hämtad 2018-05-10 från http://www.skolverket.se/skolutveckling/larande/sprak/europeisk_sprakportfolio_1.83490
- Europarådet/Fortbildningsavdelningen vid Uppsala universitet. (2014). *European Language Portfolio 16+*. Hämtad 2018-05-10 från <http://www.fba.uu.se/verksamhet/sprakportfolio/>
- Grabe, W. (2009). Teaching and Testing Reading. I Long, M. H. & Doughty, C. J. (eds). *The Handbook of Language Teaching* (s. 441–462). Chichester: Wiley & Blackwell
- Skolverket (2002). Gymnasieskolans kursprov vt 2002. En resultatredovisning.
- Skolverket (2014). *Självbedömning i engelska 5 på gymnasial nivå*. Hämtad 2018-09-28 från https://bp.skolverket.se/web/bs_gy_engeng05/enskilda-uppgifter
- Skolverket (2012). *Internationella språkstudien 2011*. Hämtad 2018-05-10 från https://www.skolverket.se/statistik_och_utvardering/internationella_studier/sprakstudien_eslc
- Skolverket (2013). *En tillbakablick på de nationella proven i gymnasieskolan: Kursproven i engelska 2001–2012*. Hämtad 2018-09-28 från

- <https://www.skolverket.se/publikationer?id=3255>
Skolverket (2016). Betyg och studieresultat. Hämtad 2018-09-28 från <https://www.skolverket.se/skolutveckling/statistik/sok-statistik-om-forskola-skola-och-vuxenutbildning>
- Sundquist, P. & Sylvén, L. K. (2012). World of VocCraft: Computer Games and Swedish Learners' L2 Vocabulary. I Reinders, Hayo (Ed.), *Digital Games in Language Learning and Teaching* (s. 189–209). Basingstoke: Palgrave MacMillan.
- Sylvén, L. K. (2004). *Teaching in English or English Teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. Doctoral Dissertation at Göteborg university. Acta Universitatis Gothoburgensis.
- Sundqvist, P. (2009). *Extramural English Matters – Out-of-School English and Its Impact on Swedish Ninth Graders – Oral Proficiency and Vocabulary*. Dissertation. Karlstad. Hämtad 2018-05-10 från <http://www.diva.portal.org/smash/record.jsf?pid=diva2%3A275141&dswid=7290>
- Sylvén, L. K. och Ohlander, S. (2014). The CLISS Project: Receptive Vocabulary in CLIL versus non CLIL Groups. *Moderna Språk* 108(2) s. 80–114.
- Sundquist, P. & Sylvén, L. K. (2016). *Extramural English in Teaching and Learning From Theory and Research to Practice*. London: Palgrave Macmillan.

Usage i kursprovet för Engelska 7

John Lövenadler & Iani Hoff

1. Introduktion

Nationella provmaterial har som syfte att stödja en likvärdig och rättvis bedömning och betygssättning, och materialen för engelska och moderna språk utvecklas på uppdrag av Skolverket inom projektet Nationella prov i Främmande Språk (NAFS) vid Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet. Vårterminen 2016 gavs för första gången ett (icke obligatoriskt) kursprov i gymnasiekursen Engelska 7. Provet omfattar fem delar: *Speaking*, *Reading*, *Listening*, *Writing* och *Usage*, där *Usage* är den del som i viss mån skiljer sig från övriga ämnes- och kursprov i engelska. De olika delarna i kursproven anknyter till Europarådets referensram för språkinlärning och språkbedömning, *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning* (Council of Europe/Skolverket, 2009) och dess rubriceringar 'reception', 'interaktion' och 'produktion'. De ämnesplaner i språk som trädde i kraft 2011 har en tydlig koppling till detta internationellt betydelsefulla dokument (Skolverket, 2011).

Att pröva *Usage* är något som många av de större internationella provinstituten gör, även om formaten kan skilja sig från det i Engelska 7 och provtypen ibland benämns *Use of English*. *Cambridge Certificate in Advanced English* har numera ett Paper som kallas *Reading and Use of English*, efter att tidigare ha haft ett renodlat *Use of English*-Paper. I *Reading and Use of English* för Cambridge CAE 2015 kommer nära hälften av poängen från *usage*-uppgifter (36 av 78p) (University of Cambridge ESOL, 2015). TOEFL iBT har *usage*-relaterade uppgifter integrerade i sin *Reading Section* i form av ordförklaringar/synonymer, och International Baccalaureate (IB) använder sig av tolv enordsluckor av totalt 53 poäng i provet för *English B, Higher Level, Paper 1* (IBO, 2013).

De format som använts för att pröva *Usage* i kursprovet för Engelska 7 är *Gap-texts* (lucktexter, antingen öppna eller med flervalsoalternativ), *Mini-gap* (i kursprovet kallade *Find the Right Word*) samt *Rephrase* (en uppgift som i provet

ATT BEDÖMA SPRÅKLIG KOMPETENS

benämns *What's Missing? – Rephrase*).⁵ Syftet är att ge eleverna möjlighet att visa prov på bredd, variation och precision i språket. Nedan visas exempel på dessa tre format.

Exempel på items i uppgiften *Gap-text*:

Edwin Battistella, an English professor at Southern Oregon University and author of a book on public apology, says that too often apologies say very little, stating the obvious and avoiding _____ (1). The letter to Ms. Rothenberg “is not so much an apology as an excuse. They’re not really sorry for anything they did. They’re sorry for the injury”, he said. “It’s an attempt to manage the situation rather than _____ (2) it.

- 1 A misbehaviour
B responsibility (X)
C enemies
D obscurity
- 2 A finalize
B understand
C resolve (X)
D avoid

Exempel på items i uppgiften *Mini-gap*:

1. Above all, researchers must come to _____ with the fact that they can no longer take the political setting as given. (terms/grips)
2. The results of the campaign were overwhelmingly positive: more than 85 per cent of the employees _____ of the decision to have smoke free zones. (approved)

Uppgiften *Rephrase* bygger på att provtagarna ska parafrasera en angiven mening med hjälp av ett nyckelord som måste användas i omformuleringen. Detta nyckelord är angivet i fetstil och instruktionen säger att ordet måste ingå i parafraseringen och att dess form inte får ändras. I luckan ska tre till fem ord inklusive nyckelordet rymmas.

Exempel på items i uppgiften *Rephrase*:

1. He realised that a window had been left open.

someone

He realised that _____ a window open.

(He realised that **someone** had left a window open)

2. The famous trade union activist and songwriter Joe Hill was born and raised in Sweden.

origin

Joe Hill, the famous trade union activist and songwriter, was _____.

(Joe Hill, the famous trade union activist and songwriter, was *of Swedish origin*)

⁵ Även på lägre nivåer av de nationella proven i engelska finns *Gap-texts* (både öppna och med flerval) med i den receptiva delen av provet. Där finns också ofta en uppgift benämnd *One Word Gaps*, vilket motsvarar det som i denna artikel kallas för *Mini-gap*.

2. Koppling till ämnesplan och kunskapskrav

Usage-delen i kursprovet för Engelska 7 syftar till att pröva såväl receptiva som produktiva förmågor. Med utgångspunkt i formuleringarna i ämnesplanen och kunskapskraven för kursen finns kanske den mest direkta kopplingen till produktiv språkförmåga. Den term som i de svenska kurs- och ämnesplanerna för främmande språk genomgående används för att beskriva förmågan att formulera sig enligt målspråkets normer är *tydlighet*. I kunskapskraven för Engelska 7 uttrycks detta på följande sätt, med utgångspunkt i ett centralt innehåll som anger att kraven ska relateras till *komplexa och formella* texter:

Betyget E: 'Eleven kan förstå **huvudsakligt innehåll och uppfattar väsentliga detaljer**, samt **med viss säkerhet** även underförstådd betydelse, i talad engelska i relativt snabbt tempo, och i skriven engelska, i olika genrer och av avancerad karaktär /.../ kan eleven formulera sig varierat, tydligt och **strukturerat**. Eleven kan även formulera sig med flyt och **viss** anpassning till syfte, mottagare och situation.'

Betyget C: 'Eleven kan förstå **huvudsakligt innehåll och uppfattar väsentliga detaljer**, samt **med viss säkerhet** även underförstådd betydelse, i talad engelska i relativt snabbt tempo, och i skriven engelska, i olika genrer och av avancerad karaktär /.../ kan eleven formulera sig varierat, **nyanserat**, tydligt och **strukturerat**. Eleven kan även formulera sig med flyt och anpassning till syfte, mottagare och situation.'

Betyget A: 'Eleven kan förstå **såväl helhet som detaljer**, samt även underförstådd betydelse, i talad engelska i snabbt tempo, och i skriven engelska, i olika genrer och av avancerad karaktär. ... I muntliga och skriftliga framställningar i olika genrer kan eleven formulera sig **med god precision**, varierat, **nyanserat**, tydligt och **välstrukturerat**. /.../ I muntlig och skriftlig interaktion i olika, även formella och komplexa, sammanhang kan eleven uttrycka sig tydligt, ledigt och med flyt samt **med flexibel och effektiv** anpassning till syfte, mottagare och situation.'

På nivå B2 i GERS, som i någon mån kan sägas motsvara baskraven i Engelska 7, finns följande formulering i relation till *kommunikativa språkliga kompetenser* (Council of Europe/Skolverket, 2009, s. 109–111):

Har ett stort ordförråd inom sitt område och när det gäller de flesta allmänna ämnen. Kan variera formuleringarna för att undvika alltför många upprepningar [...]

Generellt stor lexikal korrekthet [...]

Har god grammatisk behärskning [...]

I Skolverkets kommentarmaterial *Om ämnet engelska* (Skolverket, 2015) definieras begreppet *tydlighet* på följande sätt i samband med produktiva förmågor: 'Begreppet *tydligt* handlar om precision i språket /.../ Uttrycket **tydligt** innebär att eleven kan använda vanliga grammatiska strukturer och mönster och uttrycka sig förhållandevis exakt och korrekt i muntliga och skriftliga framställningar.'

Det är också viktigt att poängtera att vad som är språkligt 'tydligt' är beroende av kontext och grad av komplexitet i en text. Man kan således argumentera för att längre och mer avancerade texter behöver ett högre mått av språklig precision för att bli läs- och tolkningsbara, åtminstone om graden av ansträngning som krävs av läsaren inte ska bli orimligt hög. En provdel som specifikt prövar elevernas kunskaper om komplexa eller specifika ord och fraser i mer formella och akademiska texter kan därför betraktas som en potentiellt relevant del i ett kursprov som syftar till att mäta språklig förmåga på en mer avancerad nivå. Föreliggande studie syftar till att undersöka och diskutera hur och i vilken utsträckning *usage*-delen i kursprovet för Engelska 7 uppfyller detta mål.

3. Teoretiska utgångspunkter

Som tidigare nämnts ger ämnesplanens formuleringar i relation till produktiv språklig förmåga ett konkret stöd för att inkludera en *usage*-del i kursprovet. Delen avser alltså att ge eleverna möjlighet att visa sin förmåga när det gäller att uttrycka sig med tydlighet, precision och variation i mer formella och komplexa sammanhang. Att konkret pröva elevernas förmåga att producera mer avancerade strukturer och hitta specifika ord och fraser är exempel på något som ibland benämns 'the weak sense of performance assessment' (Weigle, 2002, s. 46ff). I det som benämns 'the strong sense of performance assessment' testas däremot enbart 'successful task completion', som inte bara omfattar rent språkliga förmågor utan också sådant som har att göra med exempelvis omvärldskunskap. Med utgångspunkt i Weigles diskussion kan man dra slutsatsen att den starka formen inte nödvändigtvis är mer lämplig än den svaga, utan detta beror på vilka syften ett prov har och hur man bäst uppnår reliabilitet och validitet i relation till förmågan att producera texter på avancerad nivå. En maximalt autentisk uppgiftstyp är därmed inte nödvändigtvis det mest lämpliga om man specifikt vill pröva provtagarnas kunskaper om ord, fraser och grammatiska konstruktioner i de sammanhang där dessa beståndsdelar i den kommunikativa förmågan bedöms spela en viktig roll. Givetvis är en mer formbaserad provdel

inte lämplig som enda mått på språklig produktiv förmåga, men som ett komplement till skrivuppgifter av mer autentisk typ går det däremot att argumentera för dess relevans.

I kursprovet för Engelska 7 har *usage*-delen inte integrerats i den produktiva utan i den receptiva delen, vilket bland annat hänger samman med att kunskaper om ord, fraser, kollokationer och grammatiska strukturer är lika betydelsefulla för språklig reception (Koda, 2005; Grabe, 2009). *Gap-texts*, antingen öppna eller av flervalstyp, syftar exempelvis till att pröva förmågan att förstå sammanhang och att hitta rätt ord eller fras i en större språklig kontext. *Usage* är med andra ord ett exempel på integrering av läs- och skrivförmåga, där det krävs att man både förstår och kan tolka betydelse och funktion hos ord och meningar, och att man utifrån denna förståelse kan producera ord eller fraser som kompletterar meningar på ett strukturellt och stilistiskt fungerande sätt. I relation till vedertagna metoder för att pröva receptiv förmåga kan provformatet relateras till det som kallas för *gap-filling tests* (Alderson, 2000, s. 209), och syftet med att inkludera sådana uppgifter i kursprovet är bland annat att förbättra provets validitet:

It is now generally accepted that it is inadequate to measure the understanding of text by only one method /.../ Good reading tests are likely to employ a number of different techniques, possibly even on the same text, but certainly across the range of texts tested. (Alderson, 2000, s. 206)

Vad gäller fokus på receptiv förmåga syftar uppgifterna *Mini-gap* och *Rephrase* till att pröva förståelse på fras- och meningsnivå, medan uppgiften *Gap-text* även prövar förståelse på styckes- och textnivå (Alderson 2000, s. 209). Dessutom har *usage*-uppgifterna alltså en mer uttalad produktiv aspekt än vad som finns i de övriga mer renodlade receptiva delarna av kursprovet (*Reading* och *Listening*).

Alderson (2000, s. 34ff) menar att det i samband med läsförståelse på ett främmande språk är av avgörande betydelse, ‘... what sort of linguistic knowledge is needed, and how much of it? Inevitably, the answer will in part depend upon the nature of the text and the outcomes of reading that are expected ...’. En naturlig konsekvens av detta är att betydelsen av fördjupade språkliga kunskaper ökar allteftersom texterna blir mer komplexa och avancerade. Frågan gällande exakt vilka språkliga kunskaper som har störst betydelse för läsförståelse är dock komplicerad. Alderson (2000, s. 35), exempelvis, menar att ‘... measures of readers’ vocabulary knowledge routinely correlate highly with measures of reading comprehension, and are often, indeed, the single best predictor

of text comprehension'. Shiotsu & Weir (2007), å andra sidan, menar att sambandet i själva verket är ännu starkare mellan läsförståelse och grammatisk förmåga, medan Jeon & Yamashita (2014) i sin metastudie inte finner någon signifikant skillnad mellan vokabulärkunskap och grammatisk förmåga när det gäller samband med läsförståelse. Sammantaget är det dock tydligt att såväl vokabulärkunskaper som kunskaper om fraser och grammatiska strukturer har stor betydelse när det gäller att förstå och tolka avancerade texter.

En viktig faktor som betonas av exempelvis Berman (1984) är betydelsen av att snabbt och effektivt kunna processa komplicerade grammatiska strukturer. Med hänvisning till bland annat Bermans studie menar Alderson (2000, s. 37) att '... the ability to parse sentences into their correct syntactic structure appears to be an important element in understanding text' och att '... the ability to process complex syntax may be more important for the understanding of detailed information in sentences than for the understanding of the gist of a text'. Buck (2001, s. 6ff) resonerar på liknande sätt när det handlar om hörförståelse, och betonar bland annat betydelsen av *automatiserade* språkförmågor. Buck skiljer mellan *controlled processes* och *automatic processes*, och menar att medan kontrollerade processer utgörs av kognitiva aktiviteter på ett medvetet plan, innebär automatiserade processer aktiviteter som i huvudsak sker undermedvetet. Han menar att denna distinktion är mycket betydelsefull när det gäller att förklara hörförståelseförmåga i ett främmande språk, där inlärarna ofta befinner sig på varierande nivåer i utvecklingen av automatiserade förmågor:

Given the speed and complexity of normal speech, the more automatic the listener's processing, the more efficient it will be, and the faster it can be done; and conversely, the less automatic the processing, the more time will be required. For language learners with less automatic processing, comprehension will suffer. /.../ At a certain speed, their processing will tend to break down completely, and they will fail to understand much at all. (Buck, 2001, s. 7)

Betydelsen av automatisering och förmågan att processa och tolka mer komplext språk är något vi kommer att återkomma till i diskussionen nedan.

Ett prov kan dessutom analyseras i relation till begreppet *test usefulness* (Bachman & Palmer, 1996). I Bachman & Palmers modell (1996, s. 17ff) definieras ett provs 'användbarhet' utifrån de sex faktorerna *reliability*, *construct validity*, *authenticity* (i vilken utsträckning en uppgift liknar språkanvändning i det verkliga livet), *interactiveness* (i vilken utsträckning provtagarnas intresse och bakgrundkunskaper engageras i uppgiften), *impact* (eventuella positiva *washback-*

effekter för lärare och elever) och *practicality* (genomförbarhet). Även dessa faktorer kommer vi att återkomma till nedan.

4. Analys av provdelar, uppgifter och items

Som utgångspunkt för diskussionen kring hur *usage*-delen i Engelska 7-provet fungerar följer här en analys av de kvantitativa data som erhållits från provresultat i utprövningar och från själva provtillfället. Fokus kommer att ligga på att undersöka vilka statistiskt belagda samband (korrelationer) som finns mellan elevernas resultat på *usage*-uppgifterna och på de övriga receptiva uppgifterna i provet. En sådan analys kan bland annat ge indikationer på möjliga samband mellan uppgifter och olika items med avseende på vilka förmågor som prövas. Analysen syftar på detta sätt till att skapa en grund för att kunna diskutera vad *usage*-delen egentligen mäter, hur väl den fungerar, och hur detta kan relateras till lärares och elevers uppfattning om provdelens relevans.

4.1. Analys på provdelsnivå och uppgiftsnivå

Först och främst visar analysen att det finns betydande korrelationer mellan provresultaten i *Usage*, *Reading* och *Listening*, vilket tyder på starka direkta eller indirekta samband mellan det som prövas i de olika receptiva delarna. Om man betraktar de sammantagna korrelationerna mellan provdelarna *Usage* och *Listening* respektive *Usage* och *Reading* finns där dessutom en signifikant skillnad ($p = .01$) i det avseende att det finns ett starkare generellt samband mellan *Usage* och *Listening* ($r = .79$) än mellan *Usage* och *Reading* ($r = .76$).⁶ Dock är det så att denna skillnad till viss del är en konsekvens av det något större antalet uppgifter som finns i *listening*-delen i jämförelse med *reading*-delen, och alltså inte med automatik speglar ett starkare reellt samband mellan *Usage* och *Listening* än mellan *Usage* och *Reading*. Som kompletterande jämförelse kan man därför ta bort den uppgift i respektive del med lägst intern reliabilitet (Cronbach's Alpha), vilket leder till att *usage*-korrelationen baserad på *listening*-uppgifterna *Lecture* och *Matching* (totalt 24p) jämförs med korrelationen baserad på de tre längsta *reading*-uppgifterna (25p).⁷ Denna korrigerade jämförelse visar visserligen att skillnaden i korrelationen mellan *Listening* och *Usage* ($r = .77$) och korrelationen mellan

⁶ $r_{LisUsage} = 0.791$, 95% CI [.778-.803], $r_{ReadUsage} = 0.755$, 95% CI [.740-.769]

⁷ I kursprovet består delprovet *Listening* av tre delar som här benämns *Matching*, *Lecture* och *Interview* med hänvisning till deras format, medan *reading*-uppgifterna utgörs av fyra relativt formella artiklar med blandade MC och öppna frågor.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Reading och *Usage* ($r = .75$) har minskat något (se Tabell 1a), men att den fortfarande är signifikant ($p < .01$).⁸ Det är dessutom intressant att notera att bägge dessa korrelationer är starkare än den inbördes korrelationen mellan *Reading* och *Listening* ($r = .72$). Givetvis är det svårt att dra generella slutsatser om sambanden mellan *Usage* och *Reading* respektive *Usage* och *Listening* utifrån detta begränsade material, men det kan ändå vara av värde att diskutera vad det är i uppgifterna *Lecture* och *Matching* som ligger bakom ovan beskrivna mönster.

Tabell 1a: Korrelationer mellan kursprovets receptiva delar

	Listening_{Total}	Reading_{Total}	Listening_{24p}	Reading_{25p}
Usage	$r = .79$	$r = .76$	$r = .77$	$r = .75$

Tabell 1b: Korrelationer mellan hörförståelse- och *usage*-uppgifter

	Mini-gap	Rephrase	Gap-text	Usage (total)
<i>Lecture</i>	$r = .65$	$r = .61$	$r = .57$	$r = .70$
<i>Matching</i>	$r = .62$	$r = .58$	$r = .56$	$r = .67$

Tabell 1c: Korrelationer mellan receptiva delar och enskilda *usage*-uppgifter

	Mini-gap	Rephrase	Gap-text
Reading _{25p}	$r = .68$	$r = .64$	$r = .63$
Listening _{24p}	$r = .72$	$r = .67$	$r = .64$

De uppgifter inom *reading*- och *listening*-delarna som har starkast samband med *usage*-delen är de relativt avancerade *listening*-uppgifterna *Lecture* och *Matching*, där korrelationerna mellan *Usage* och *Lecture* uppgår till $r = .70$ och mellan *Usage* och *Matching* till $r = .67$ (Tabell 1b). Detta kan jämföras med *usage*-korrelationen för den återstående hörförståelseuppgiften *Interview* ($r = .62$) samt för de olika uppgifterna i *reading*-delen som samtliga visar korrelationer med *usage*-delen på mellan $r = .60$ och $r = .63$. Det tycks med andra ord finnas ett direkt eller indirekt samband mellan kunskaper om språklig struktur och förmågan att uppfatta innehållet och budskapet framförallt i relativt komplexa hörförståelseuppgifter som *Lecture* och *Matching*. Om man jämför sambandet mellan elevresultaten i exempelvis *Lecture* och de olika uppgifterna inom *usage*-delen (Tabell 1b) kan man notera att korrelationen är särskilt stark med uppgiftsformaten *Mini-gap* ($r = .65$) och *Rephrase* ($r = .61$), men aningen svagare med lucktexten med flerval, *Gap-text* ($r = .57$).

⁸ $r_{LidUsage} = .772$, 95% CI [.758-.785], $r_{ReadUsage} = 0.745$, 95% CI [.730-.760]

Man kan här notera att korrelationen mellan *Mini-gap* och *Lecture* till och med motsvarar medelkorrelationen mellan de olika uppgifterna inom *usage*-delen (trots att de olika *usage*-uppgifterna kan tyckas vara inbördes mer likartade). Detta tyder på att förmågan att komplettera en språklig struktur med rätt ord eller fras har ett starkt samband med de förmågor som krävs för att klara av en hörförståelseuppgift som *Lecture*. Om man studerar sambandet mellan *reading*-delen, *listening*-delen och de enskilda uppgifterna inom *usage*-delen (Tabell 1c) ser man att de två uppgifter som kanske kan uppfattas som mest grammatiskt fokuserade (*Rephrase* och *Mini-gap*) i själva verket har ännu starkare samband med *Listening* och *Reading* än den uppgift som man kan tycka mer specifikt borde pröva förståelse på textnivå (*Gap-text*). Sammantaget visar detta att även de uppgifter inom *usage*-delen som relaterar mer uppenbart till språklig produktion än till reception (*Rephrase* och *Mini-gap*) i själva verket har ett stabilt och starkt samband med de mer renodlat receptiva delarna *Reading* och *Listening*.

4.2. Analys på item-nivå

Genom att studera empiriskt material från utprovningarna går det även att analysera dessa samband på item-nivå, och därmed studera hur enskilda items korrelerar med *Usage*. Som utgångspunkt kan man ta de uppgifter som visade sig ha starkast samband med de förmågor som provas i *usage*-delen, nämligen *listening*-uppgifterna *Lecture* och *Matching* (Tabell 2). De items av totalt tolv i *Lecture* som då visar sig ha starkast korrelation med *Usage* är items 10 ($r = .49$), 2 ($r = .47$), 8 (2p-uppgift, $r = .47$) och 1 (2p-uppgift, $r = .46$). I *Matching* (10 items) finns de starkaste korrelationerna med *Usage* i items 4 ($r = .40$), 10 ($r = .34$), 8 ($r = .32$) och 9 ($r = .31$).

Tabell 2: Korrelationer (r) mellan *Usage* och specifika items i två hörförståelseuppgifter

	1	2	3	4	5	6	7	8	9	10	11	12
Matching	.26	.17	.17	.40	.20	.27	.27	.32	.31	.34		
Lecture	.46	.47	.38	.40	.16	.37	.42	.47	.32	.49	.24	.36

Den uppenbara frågan är då vad det är i delproven *Matching* och *Lecture* som ligger bakom dessa starka samband. *Matching* är en uppgift som består av 10 korta uppspelningar, där provtagarna ska koppla ihop var och en av dessa med korrekt ämne (av totalt 20 angivna ämnen). De icke korrekta svarsalternativen (distraktorerna) kan ofta kopplas till något som sägs i uppspelningarna, men de beskriver inte det viktigaste eller det egentliga innehållet. Möjligheten att lösa

ATT BEDÖMA SPRÅKLIG KOMPETENS

dessa uppgifter skulle därmed kunna hänga samman med förmågan att uppfatta avgörande detaljer i komplexa sammanhang, vilket i sin tur kan ha ett samband med väl utvecklad språklig förmåga (jfr Alderson 2000 och diskussionen i avsnitt 3). Har provtagare inte tillgång till ord och fraser i automatiserad form är det betydligt svårare att hinna uppfatta vad huvudpoängen egentligen är.

Som tidigare nämnts är den uppgift som uppvisar allra starkast samband med *usage*-delen hörförståelseuppgiften *Lecture*. Uppgiften bygger på en föreläsning med relativt formellt men inte alltför avancerat språk, och många av föreläsarens påståenden och argument upprepas dessutom flera gånger. I ett relativt perspektiv bör det därmed inte vara anmärkningsvärt krävande att förstå innehållet på en övergripande och generell nivå. Svårigheten ligger troligtvis snarare i att frågorna som ställs kräver en djupare och mer exakt förståelse av innehållet. Vad gäller lösningsfrekvenser är uppgiften förhållandevis svår och genererar dessutom relativt stora könsskillnader (Tabell 3).

Tabell 3: Uppgiften *Lecture* (lösningsfrekvens (*P*-värde) per item i utprovningen)

	1(1)	1(2)	2	3	4	5	6	7	8(1)	8(2)	9	10	11	12
Totalt	.64	.64	.65	.52	.21	.49	.32	.60	.39	.39	.50	.39	.24	.58
Män	.68	.68	.69	.58	.24	.52	.40	.67	.40	.40	.56	.49	.26	.61
Kvinnor	.60	.60	.62	.47	.18	.47	.27	.55	.38	.38	.45	.31	.23	.55

Uppgiften består av fyra öppna frågor (varav två kan ge dubbla poäng) och åtta flervalsuppgifter (MC). Som framgår av Tabell 2 är det framförallt fyra items som genererar mycket höga korrelationer med *Usage* och som därför direkt eller indirekt är relaterade till mer språkliga svårigheter. Av dessa fyra items är tre av typen med öppna svarsalternativ (1, 8, 10). Två av dessa är relativt svåra (item 8 och 10) medan ett är relativt lätt (item 1). Den övriga frågan med hög *usage*-korrelation är item 2 (flervalsfråga), som statistiskt sett är den allra lättaste frågan i uppgiften ($p = .65$). I detta item har man stor hjälp om man förstår ett par specifika lågfrekventa ord och fraser, vilket åtminstone delvis skulle kunna förklara den starka korrelationen med *Usage*. Visserligen går det att svara på frågan även genom mer implicita ledtrådar i föreläsningen, men eftersom detta inte är direkt nödvändigt om man förstår de avgörande fraserna blir möjligheten att svara på frågan starkt sammankopplad med den sortens språklig förmåga som prövas i *usage*-delen.

Varför korrelationen med *usage*-delen är så stark just i samband med hörförståelsens öppna items är dock intressant att diskutera. Som resultatet visar

gäller detta alltså oavsett om dessa items är lätta eller svåra, och oavsett om det gäller items som är värda 1p eller 2p. En tänkbar möjlighet är att öppna items av den här typen kräver en djupare förståelse, särskilt då provtagaren inte får något stöd för tolkningen av de olika svarsalternativens formuleringar (som i MC-items). Detta innebär att en provtagare som inte tillräckligt snabbt lyckas processa och tolka det relevanta avsnittet nästan blir helt utan möjlighet att kunna besvara den öppna frågan. Man kan anta att detta blir särskilt avgörande i mer avancerade och komplexa uppgifter, där öppna items inte går att lösa enbart genom att uppfatta tydliga och konkreta fakta eller påståenden.

Den allra starkaste korrelationen med *usage*-delen finns i item 10, som är relativt svår ($p = .39$) och som dessutom uppvisar stor könsskillnad ($p = .49$ för män och $p = .31$ för kvinnor). En närmare granskning av detta item antyder att det som till stor del kan vara avgörande här är förmågan att snabbt och effektivt processa innehållet i en strukturellt komplicerad grammatisk struktur av typen [*Denver's Mayor Phil Dawson signed into law an ordinance prohibiting the use of firearms*], som dessutom kommer ganska snart efter det innehåll som behöver tolkas för att lösa föregående item. Det är inte orimligt att tänka sig att möjligheten att snabbt uppfatta vad som sägs här kräver att man har en automatiserad språkförmåga och vana att processa mer avancerade strukturer (se diskussionen nedan samt Buck, 2001).

5. Lärares och elevers reaktioner på *usage*-delen i kursprovet

Som påpekas av Bachman (1990, s. 279), 'tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large'. Mot bakgrund av Bachmans påstående är det intressant att undersöka hur lärare och elever ser på *Usage* i kursprovet för Engelska 7 och därmed i vilken utsträckning provdelen kan sägas ha *face validity* och potential för *washback*.

5.1. Lärares reaktioner

Många lärare ger uttryck för positiva reaktioner i lärarenkäterna från både utprovning och provet vt 2016, exempelvis vad gäller kvalitet och relevans:

Säger mycket om elevens förmåga beträffande vokabulär, idiom, fraseologi och grammatik.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Usage-delen var väldigt bra. Nyttänk!!

Mycket bra! Optimal svårighetsgrad + väldigt bra utformad övning där flera färdigheter sätts på prov.

Vi behöver tänka om vad gäller just precision och språklig komplexitet.

Det är detta eleverna har svårt med ... Bra uppgift som borde tas med även i Engelska 6.

Det var ganska svårt, tyckte många elever, men jag tänker att det stämmer med vad man borde prestera på den här nivån. Dock är det ju många svaga elever som läser kursen för att få meritpoängen och de hamnar lågt.

Andra positiva kommentarer handlar om de *washback*-effekter som provet anses kunna ge:

Mycket värdefullt och bra underlag. Det ger goda riktlinjer om varthän eleverna ska sträva i kursen och har påverkat/hjälpt mig (positivt) inför nästa läsår och utformning av kursen.

Uppgifterna klargjorde intentionerna bakom formuleringarna i betygskrav och centralt innehåll och visade att kraven är högre än innehållet i våra läroböcker.

Uppgiften är mkt bra. Den är till stöd för undervisningen då den 'andas' högskoleförberedande.

Ovanstående lärare kan med andra ord sägas vara nöjda med att kursprovets *usage*-del prövar en färdighet som de anser är viktig, och den generella uppfattningen är att provdelen har hög validitet. I vissa fall framkom även kritiska kommentarer, vilka huvudsakligen handlade om i vilken utsträckning *usage*-delen är relevant som mått på kommunikativ förmåga, samt i vilken utsträckning eleverna är medvetna om svårighetsgraden och vad provdelen kräver av dem:

Det mesta handlade om fasta uttryck. Eleverna har fostrats till att hitta en utväg, när ordförrådet sviker. Detta var en chock för dem.

Kan nog vara bra att i instruktionerna tydliggöra att endast idiomatiskt 'perfekta' ord ger poäng, och att 'göra sig förstådd' inte duger. Jag upplevde uppgiften som i nivå med vissa provfrågor på antagningsprovet till mastersprogrammet för översättare...!

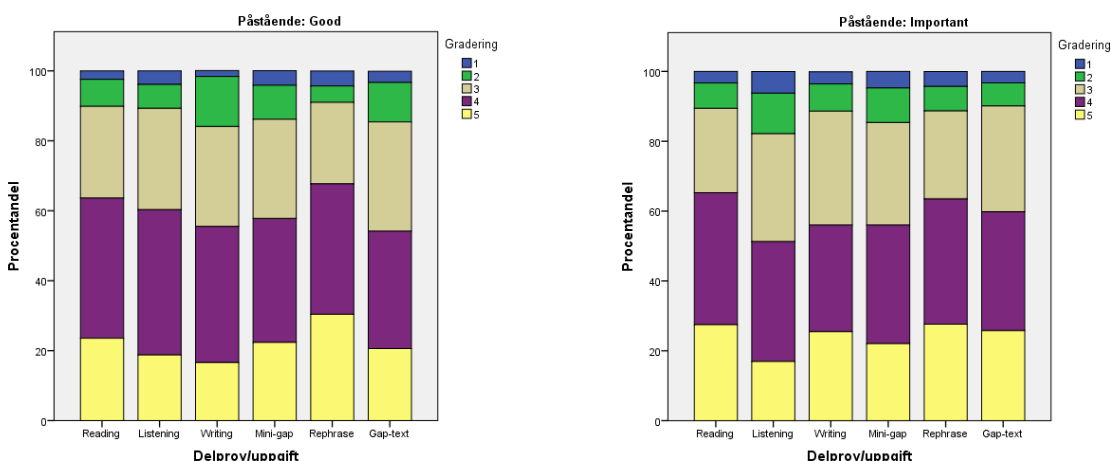
Värt att notera är också att majoriteten av lärarnas kommentarer handlar om *usage*-delens relation till produktiv snarare än receptiv förmåga, något vi kommer att återkomma till nedan.

5.2 Elevers reaktioner

I samband med utprövningarna samlades data in från provtagarnas egna upplevelser (s.k. *test-taker feedback*) av de olika provdelarna i form av femgradiga Likertskalor. Påståendena som gavs var bland annat ‘I think the task was good’, ‘I think the task tested something important’ och ‘I think the task gave a fair picture of my level of English’, där ‘5’ motsvarar ‘håller helt och hållet med’ och ‘1’ ‘håller inte alls med’. En sammanställning av dessa svar visar bland annat följande:⁹

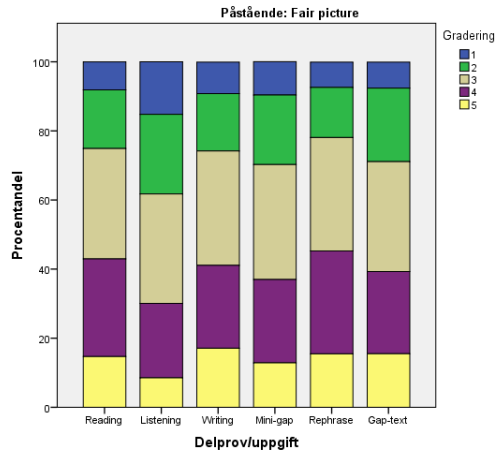
Tabell 4: Medelvärden i elevernas uppfattning om provdelar och provuppgifter

	Reading	Listening	Writing	Rephrase	Mini-gap	Gap-text
Good	3.75	3.64	3.54	3.85	3.63	3.57
Important	3.79	3.44	3.67	3,76	3.59	3.72
Fair picture	3.24	2.85	3.23	3.32	3.11	3.18



⁹ Observera att dessa svar är baserade på den totala utprövningsgruppen (ca 4 000 elever) indelade i fem grupper som genomförde olika uppgifter i det planerade kursprovet. Grupperna genomförde dock uppgifter från flera olika delprov vilket möjliggör en direkt jämförelse mellan uppfattningen om *Usage*, *Reading* och *Listening* inom en och samma elevgrupp (*within-subject design*). Beräkningar visar då att det inte finns någon statistiskt signifikant skillnad mellan elevernas uppfattningar om t.ex. *usage*-uppgiften *Rephrase* och någon av *reading*-uppgifterna. Däremot finns det en signifikant skillnad (Friedman's Two-Way ANOVA by Ranks, $p < .001$) exempelvis mellan elevernas uppfattning om *Rephrase* och *listening*-uppgifterna *Lecture* och *Matching*, i det avseende att *Rephrase* både anses mäta något som är 'viktigare' och ge en mer 'rättvisande bild' av deras nivå i engelska. I jämförelse med *Lecture* ansåg eleverna dessutom att *Rephrase* var en 'bättre' uppgift ($p < .001$).

ATT BEDÖMA SPRÅKLIG KOMPETENS



Som tabellen och diagrammen visar är eleverna generellt sett positiva till *usage*-delen i provet, särskilt när det gäller uppgiften *Rephrase* (med avseende på hur 'bra' den är och huruvida den mäter något som är 'viktigt').¹⁰ I ett relativt perspektiv (i jämförelse med övriga provdelar) upplever de dessutom att *Usage* ger en 'rättvis' bild av deras kunskaper. Utifrån dessa siffror och jämförelsen med elevernas uppfattning om de provdelar som behandlar *Writing*, *Reading* och *Listening* (som också är generellt sett positiv) är det tydligt att denna provdel ur ett elevperspektiv fungerar mycket väl.¹¹ I enkäterna fanns även öppna kommentarfält där eleverna hade möjlighet att specificera hur de upplevde de olika uppgifterna. Eftersom *usage*-uppgiften *Rephrase* är unik för kursprovet i Engelska 7 och därför ny för eleverna är det speciellt intressant att se hur eleverna kommenterar denna. Som tydligt framgår av enkätresultaten är elevgruppen som helhet mycket nöjd med uppgiften både när det gäller kvalitet, relevans och rättvisa:

A very good take on examining. The individuals use of the English language in a grammatical sense.

Great. Better way of testing knowledge than one word gaps.

Great exercise. Different from what ive seen before and entertaining.

A great way to learn something more difficult and useful in English.

¹⁰ I diagrammen återfinns de mest positiva svaren i botten och de mest negativa svaren i toppen av staplarna.

¹¹ För att sätta dessa siffror i en större kontext kan man även jämföra dem med motsvarande undersökningar som genomförts i utprövningar för de nationella proven i Engelska 6. En sammanställning exempelvis från utprövningen vt15 visar då de genomsnittliga resultaten: *Reading* (Good: 3.33 / Important: 3.36 / Fair picture: 3.11), *Listening* (Good: 3.00 / Important: 3.23 / Fair picture: 2.79) och *Mini-gap* (Good: 3.37 / Important: 3.38 / Fair picture: 3.05). Denna jämförelse ger ytterligare stöd för att eleverna generellt sett är mycket positiva till *usage*-delen i kursprovet för Engelska 7.

USAGE I KURSPROVET FÖR ENGELSKA 7

Det är en bra övning, man ser om eleven förstår sammanhanget eller ej.

This was a good exercise because it is important to know how to rephrase things in English.

Bra att man får tänka efter själv och hitta egna lösningar på nya meningar.

Test our grammar skills very well, the task gives a fair picture of our level of English.

De negativa kommentarer som framkommer handlar främst om att man i uppgiften *Rephrase* är begränsad i sin frihet att använda vilka ord och fraser man vill:

I thought the answers could be too wide, too 'open'. It was hard to know what you were looking for exactly.

Dåligt med så många begränsningar att använda sitt språk på. När man ska använda sitt språk i verkligheten är man inte så låst i en mall så detta testar inte ens kunskap för språket på ett bra sätt.

I find that it is very hard to find words while reading. It does not really show how good I am in English.

Using sayings seem kinda weird since I don't really think it shows your English skills.

It was very hard to complete these sentences, because the words that were given are not used so often.

Det kändes som att det fanns för mycket alternativ.

Att en sådan begränsning är inbyggd i uppgiften är förstås direkt kopplat till uppgiftens syfte, eftersom det som provas här inte är det eleverna är mest vana vid (att kommunicera ett budskap så effektivt som möjligt utifrån de resurser man har tillgång till). Istället är syftet i denna uppgift mer inriktat på att pröva elevernas kunskaper när det handlar om en specifik del av kunskapskraven, nämligen den som rör språklig variation och precision. Med avseende på uppgifternas *autenticitet* (Bachman & Palmer, 1996) tycks det med andra ord som att en del elever reagerar negativt, framförallt på en uppgift som *Rephrase*. Det är givetvis också både naturligt och förståeligt att vissa elever reagerar med tveksamhet mot en uppgift som skiljer sig från det de har vant sig vid i språkundervisning och tidigare nationella prov, men samtidigt visar enkätresultaten tydligt att eleverna på det stora hela är mycket positiva till utformningen av denna uppgift.

6. Slutsatser

Enligt elev- och lärarenkäterna i samband med det första kursprovet i Engelska 7 (och utprovningarna inför detta) tycks det finnas ett tydligt stöd bland lärare och elever för att *Usage* utgör en 'bra' provdel, att den prövar något som är 'relevant' och att den ger eleverna en 'rättvis' chans att visa vad de kan när det gäller engelska. Exempelvis ger mer än 85 % av eleverna *usage*-uppgifterna en 3:a, 4:a eller 5:a på skalorna gällande hur bra och hur relevanta uppgifterna är när det gäller att pröva engelsk språkförmåga, och många kommenterar uppgifterna i positiva ordalag i de öppna svaren. De negativa kommentarer som dyker upp i enkäterna rör huvudsakligen det faktum att *usage*-provet inte alltid tillåter att man använder sina tillgängliga språkliga resurser i de fall där kunskaper kring mer avancerade ord och strukturer saknas. Dessa argument är direkt relaterade till Weigles (2002, s. 46ff) distinktion mellan den svaga och den starka formen av *performance assessment*, där enbart den starka formen är direkt knuten till 'task completion' som mått på språklig förmåga. Som diskuterats ovan finns det dock i kursprovet för Engelska 7 anledning att komplettera de konventionella delarna i provet med uppgifter som mer specifikt prövar provtagarnas kunskaper om ord, fraser och grammatiska konstruktioner, eftersom dessa beståndsdelar i den kommunikativa förmågan bedöms utgöra viktiga delar av kursprovets *construct*. Ur ett validitetsperspektiv kan man därför argumentera för *usage*-delens relevans, naturligtvis förutsatt att den har hög reliabilitet med avseende på det den avser att pröva.

Som också diskuterats i tidigare avsnitt relaterar *usage*-uppgifterna inte enbart till muntliga och skriftliga produktiva förmågor utan också till receptiva förmågor. Det är ju dessutom så att *usage*-delen tillsammans med *Listening* och *Reading* ingår i kursprovets receptiva block. I relation till de teoretiska diskussionerna i exempelvis Alderson (2000) och Buck (2001) är det då intressant att fråga sig hur *usage*-kunskaper relaterar till *Listening* och *Reading*. Som analysen av provresultaten visar finns det ett mycket starkt samband mellan de förmågor som kan associeras med läsförståelse, hörförståelse och *Usage*. Sambandet mellan *Usage* och var och en av de andra receptiva delarna är i själva verket ännu starkare än det inbördes sambandet mellan *Listening* och *Reading*, vilket rimligen kan sägas stärka argumenten för att *Usage* bör finnas med som en del av det receptiva blocket i provet.

Ytterligare ett intressant faktum som framkommer i den statistiska analysen är att sambandet mellan *Usage* och *Listening* är ännu starkare än det mellan *Usage*

och *Reading*, trots att samtliga uppgifter inom *usage*-delen till formen är en sorts läsuppgifter. I anslutning till diskussionen i Buck (2001) är det inte orimligt att dra slutsatsen att det starka sambandet mellan *Listening* och *Usage* i kursprovet kan associeras med betydelsen av automatiserat språkprocessande (jfr Bucks distinktion mellan *controlled processes* och *automatic processes*), där en automatiserad förmåga att förstå och tolka ord och fraser har stor betydelse för möjligheten att ta till sig ett komplext innehåll. I någon mån kan man därmed anta att automatiserad språkförmåga eller förmågan att snabbt och effektivt processa ord, fraser och strukturer i många fall kan ha ännu större betydelse i hörförståelse än i läsförståelse, eftersom det där inte finns samma möjlighet att gå tillbaka och använda det som Buck (2001) benämner kontrollerade processer. Även om *usage*-delen inte prövar automatiserade processer på ett direkt sätt (provtagarna kan ju gå tillbaka och läsa ord och meningar i vilken takt de vill), så prövar *Usage* elevernas kunskaper om vokabulär, kollokationer och grammatiska strukturer, vilket kan antas sammanfalla med deras möjlighet att snabbt och effektivt processa och tolka språklig input. Betydelsen av *usage*-relaterade kunskaper för möjligheten att klara en provuppgift kan därmed antas bli särskilt stor i just hörförståelse och i sammanhang där uppfattandet av detaljer är viktigt. Som visats ovan tycks så också vara fallet i kursprovet för Engelska 7, inte minst när uppfattandet av detaljer behövs för att få en helhetsförståelse av ett innehåll.

Förmågan att förstå ett innehåll på specifik snarare än översiktlig nivå tycks alltså ha ett tydligt samband med väl utvecklad *usage*-förmåga. Mot bakgrund av det faktum att vokabulärkunskap och grammatisk kunskap interagerar och inte klart går att separera från varandra (Koda, 2005) kan man konstatera att ett aktivt och automatiserat förråd av ord, fraser och strukturer har mycket stor betydelse för möjligheten att förstå tal eller skrift, inte minst på mer avancerad nivå. Samtidigt visar detta relevansen av att ha med *Usage* i den receptiva delen av kursprovet, eftersom förmågan att förstå och tolka mer avancerat tal och mer avancerade texter utgör tydliga och viktiga delar av centralt innehåll och kunskapskrav i kursen Engelska 7, och *Usage* tycks fånga upp den språkförmåga som behövs för att effektivt kunna processa avancerat språk i såväl tal som skrift.

Det går också att konstatera att resultatet av empiriska data och statistiska beräkningar ligger väl i linje med den generella uppfattningen bland lärare och elever angående *Usage* som en relevant del av kursprovet för Engelska 7, och med avseende på kursprovets *face validity* och *washback*-effekter (Bachman, 1990) är detta naturligtvis både viktigt och tillfredsställande. Som Bachman (1990, s.

288–289) påpekar: ”The bottom line in any language testing situation, in a very practical sense, is whether test-takers will take the test seriously enough to try their best, and whether test users will accept the test and find it useful. For these reasons, test appearance is a very important consideration in test use”.

Slutligen kan *usage*-delen i kursprovet för Engelska 7 relateras till den modell för *test usefulness* som presenteras i Bachman & Palmer (1996). Som tidigare nämnts definieras i modellen ett provs användbarhet utifrån de sex faktorerna *reliability*, *construct validity*, *authenticity*, *interactiveness*, *impact* och *practicality* (1996, s. 17). Föreliggande studie indikerar att *usage*-delen i kursprovet, tillsammans med och som ett komplement till de övriga delarna, har visat sig fungera mycket väl såväl mättekniskt (t.ex. med avseende på reliabilitet och validitet) som vad gäller *interactiveness* (elevernas engagemang i uppgiften), *impact* (värdefulla *washback*-effekter) och *practicality* (genomförbarhet). Även om en minoritet av eleverna reagerar negativt mot att *usage*-delen inte till fullo uppfyller allt man kan önska när det gäller autenticitet, så är slutsatsen ändå att den mer än väl motsvarar kraven på kvalitet, och att provdelen fyller en viktig funktion i kursprovet för Engelska 7.

Referenser

- Alderson, C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Berman, R. (1984). Syntactic components of the foreign language reading process. I J. Alderson och A. Urquhart (Red.), *Reading in a Foreign Language* (s. 139–156). London: Longman.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Council of Europe/Skolverket. (2009). Gemensam europeisk referensram för språk: lärande, undervisning, bedömning.
www.skolverket.se/publikationer?id=2144
- Grabe, W. (2009). *Reading in a second language: moving from theory to practice*. Cambridge: Cambridge University Press.
- International Baccalaureate Organization, Specimen Paper (2013).
http://www.literature.burntmango.org/wp-content/uploads/2013/09/IB_2013_A1_A2_specimen_papers.pdf
- Jeon, E. & Yamashita, J. (2014). L2 reading comprehension and its correlates: a meta analysis. *Language Learning*, 64, 160–212.

USAGE I KURSPROVET FÖR ENGELSKA 7

- Koda, K. (2005). *Insights into Second Language Reading*. Cambridge: Cambridge University Press.
- Shiotsu, T. & Weir, C. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24, 99–128.
- Skolverket (2015). *Ämneskommentar: Om ämnet Engelska*.
https://www.skolverket.se/polopoly_fs/1.243805!/C3%84mneskommentarer%20engelska.pdf
- TOEFL iBT® Test Questions Copyright © (2015).
<https://www.ets.org/Media/Tests/TOEFL/pdf/SampleQuestions.pdf>
- University of Cambridge ESOL Sample Paper (2015).
<http://www.cambridgeenglish.org/exams/advanced/preparation/>
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Vokabulär och bedömning av skriftlig förmåga

Eva Olsson

1 Inledning

Att kunna uttrycka sig i skrift på engelska är en viktig förmåga, inte minst eftersom engelska i allt större utsträckning används för kommunikation i olika sammanhang, både i och utanför Sverige, såväl på fritiden som i (högre) utbildning och arbetsliv (se t.ex. Matsuda, Ortmeier-Hooper & Matsuda, 2009). I de svenska läroplanerna för grund- och gymnasieskolan framhålls vikten av god förmåga att kommunicera skriftligt på engelska och riktlinjer för undervisningen ges (Skolverket, 2011b, 2011c). I kurs- och ämnesplaner anges även kunskapskrav för skriftlig förmåga för olika årskurser och betygsnivåer. Kraven kan till exempel avse förmåga att anpassa texten till mottagaren eller förmåga att uttrycka sig varierat. Till stöd för bedömningen finns också Skolverkets kommentarmaterial till bedömning i engelska där läroplanernas formuleringar förtydligas och exemplifieras (Skolverket, 2011a, 2012).

Eftersom språk till stor del består av ord är elevers ordförråd givetvis av betydelse för deras skriftliga förmåga – utan ord blir det inte någon text. En elev som har tillgång till en rik vokabulär har större möjligheter att variera och nyansera sitt språkbruk, och att uttrycka sig med större tydlighet och precision än en elev med en mer begränsad vokabulär. Många av de faktorer som ska ligga till grund för bedömning av skriftlig förmåga som nämns i kurs- och ämnesplaner för engelska samt i kommentarmaterial till dessa kan kopplas till vokabulär. När *”språkets omfång och bredd”* bedöms i en skriftlig framställning i grundskolan bedöms bland annat i vilken mån *”ett mycket grundläggande förråd av ord och enkla fraser”* används eller om *”en bredare repertoar av ord och satsstrukturer”* används i framställningen (Skolverket, 2012: 23). I kommentarer till kunskapskrav för Engelska 5–7 nämns exempelvis att bedömning av en framställnings grad av anpassning till syfte, mottagare och situation med avseende på interaktion bland annat handlar om hur tydligt eleven *”varierar sitt språk och valet av innehåll så att ord, uttryck, stilnivå och innehåll passar in i olika situationer”* (Skolverket, 2011a).

I denna pilotstudie jämförs elevtexter som skrivits inom ramen för de nationella proven i engelska i grundskolans årskurs 6 och 9 samt i gymnasieskolans kurser Engelska 5 och Engelska 6, och som bedömts uppnå kraven för betygsnivåerna E och A avseende skriftlig förmåga. I studien undersöks såväl gemensamma drag som skillnader i texternas vokabulär mellan årskurser och betygssteg.

2 Bakgrund och syfte

För elever innebär det oftast en större utmaning att skriva på ett främmande språk än att exempelvis läsa och förstå en text på språket (se t.ex. Elgort & Nation, 2010). Den receptiva förmågan föregår den produktiva – man förstår oftast mer än vad man själv kan uttrycka – bland annat därför att den receptiva vokabulären, dvs. de ord som förstås när man läser eller lyssnar, är större än den produktiva vokabulären, dvs. de ord man kan använda i språklig produktion i tal eller skrift (Laufer, 1998; Gass, 2013). Laufer (2005) fann att elever i högre grad använde ord i det receptiva ordförrådet även produktivt om orden var högfrekventa jämfört med om de var lågfrekventa; ju oftare man stöter på ett ord, desto lättare blir det att själv använda det (jfr Nation, 2013). Därav följer att man kan förvänta att elever som studerar engelska som främmande språk i stor utsträckning använder ord som är vanligt förekommande i språket när de börjar skriva för att sedan utöka sin språkliga repertoar efter hand så att även mer lågfrekvent vokabulär används i högre årskurser.

Förutom i skolan använder och lär sig elever i olika omfattning engelska även på sin fritid. Tidigare forskning har visat att mellanstadieelever som ofta ägnar sin fritid åt aktiviteter där engelska används, såsom dataspel, har ett större ordförråd än elever som inte gör det (Sylvén & Sundqvist, 2012). Resultat av en studie bland elever i årskurs 9 visade att de elever som ofta använde engelska på sin fritid tycktes ha tillgång till en rikare språklig repertoar än de elever som mer sällan använde engelska utanför skolan (Olsson, 2012). De elever som frekvent använde engelska på fritiden nyanserade sitt språkbruk i större utsträckning, exempelvis genom att använda förstärkningsord såsom adverb. Vidare varierade de sitt ordval mer än elever som sällan använde engelska på fritiden. Detta var särskilt tydligt i en texttyp där mer vardagligt språk kunde förväntas.

En longitudinell studie av gymnasieelevers argumenterande och utredande texter på gymnasiet visade att de elever som ofta ägnade sig åt engelska på sin fritid använde fler akademiska ord i sina texter än andra elever när de började

gymnasiet (Olsson & Sylvén, 2015). Dock var deras progression vad gäller användning av akademisk vokabulär inte starkare än andra elevers, vilket tyder på att fritidsengelskans effekt på det akademiska ordförrådet kan tänkas plana ut när man nått en viss nivå – troligtvis stöter elever mer sällan på akademiska ord än mer vardaglig vokabulär i samband med sina fritidsaktiviteter.

Vad en elev lär sig i engelska på sin fritid beror på ett antal faktorer, till exempel vad det är för typ av aktiviteter eleven ägnar sig åt, hur ofta och hur länge eleven ägnar sig åt aktiviteterna, vilka språkliga register eleven då kommer i kontakt med och hur eleven använder språket under aktiviteten (jfr Sundqvist & Sylvén, 2016). När elevers skriftliga förmåga bedöms i skolan är det naturligtvis inte bara det de lärt sig i skolan som bedöms utan även det de kan ha lärt sig på andra ställen. Ett större ordförråd tycks kunna vara en fördel som flitig användning av engelskan på fritiden kan ge (jfr Sundqvist, 2009; Olsson, 2016; Sundqvist & Sylvén, 2016)

Att ordförrådets storlek är en faktor som i hög grad bestämmer hur väl skribenter lyckas med sitt skrivande, inte minst när det gäller skribenter som skriver på annat språk än sitt förstaspråk, har bland annat visats i studier av Laufer och Nation (1995). Olika mätbara språkliga drag som är kopplade till vokabulär och som verkar ha relevans för bedömning av textkvalitet har identifierats i en rad studier (se t.ex. Grant & Ginther, 2000; Hinkel, 2011; Crossley, Salsbury & McNamara, 2012). Exempelvis samvarierar textlängd och ordlängd i relativt stor utsträckning med bedömd textkvalitet, dvs. längre texter och texter som i genomsnitt innehåller längre ord bedöms ofta som starkare än kortare texter och texter som innehåller genomsnittligt kortare ord (Grant & Ginther, 2000). Det finns också forskningsresultat som visar att ordvariation och textkvalitet tenderar att samvariera (Crossley & McNamara, 2012). Detta innebär att texter som bedöms hålla hög kvalitet ofta innehåller ett mer varierat ordval än texter som bedöms hålla lägre kvalitet. Emellertid finns det givetvis även korta texter som är utmärkta, liksom det finns långa texter som är mindre bra. Likaså förekommer texter med varierad, men felaktigt använd vokabulär.

I helhetsbedömning av texter måste alltså en rad olika faktorer beaktas. Några av dessa är i fokus i denna pilotstudie där textlängd, ordvariation och ordlängd analyseras i elevtexter för att undersöka likheter och skillnader mellan betygssteg och mellan årskurser. Även frekvensbandsanalyser görs för att undersöka hur frekventa de ord som förekommer i elevtexterna är i det engelska

språket, dvs. i vilken utsträckning elever i olika årskurser och på olika betygsnivåer använder mycket vanliga ord och i vilken mån de även använder mer ovanliga ord.

Studien syftar inte till att hitta genvägar till bedömning av texter genom mekaniska analyser; bedömning av skrivande är mer komplext än så (se t.ex. Wardle & Roozen, 2012). Förhoppningen är att studien ska bidra till att belysa ordförrådets betydelse för förmågan att skriva på engelska och för bedömning av denna förmåga. Genom att jämföra texter från olika årskurser som bedömts uppnå olika betygsnivåer belyses skillnader och likheter i skriftlig förmåga med avseende på vokabulär mellan årskurser och mellan betygssteg. Studien syftar även till att i någon mån åskådliggöra variationen inom en årskurs eller kurs.

3 Textmaterial

De texter som analyseras i studien skrevs inom ramen för de nationella proven i engelska i grundskolans årskurs 6 och 9, samt i gymnasieskolans kurser Engelska 5 och Engelska 6 under åren 2012–2016. I årskurs 6 och 9 är provens skrivuppgifter, i linje med kursplanen, ofta formulerade så att berättande och/eller beskrivande texter kan förväntas, särskilt i årskurs 6, medan gymnasiet provuppgifter oftare uppmanar eleverna att argumentera eller utreda. I instruktionen till skrivuppgiften i årskurs 6 brukar eleverna oftast uppmanas att skriva så mycket de kan, medan man i skrivuppgifterna i årskurs 9 och på gymnasiet brukar ange att texterna exempelvis bör vara mellan 200 och 600 ord, oftast något kortare i årskurs 9 än i gymnasiekurserna, och något längre i Engelska 6 än i Engelska 5. I Bilaga 1–4 ges exempel på provuppgifter och bedömda elevtexter från prov för grundskolans årskurs 9 och gymnasieskolans Engelska 6. För ytterligare exempel på hur de nationella provens skrivuppgifter kan se ut i olika årskurser och kurser hänvisas till [det nationella provprojektets hemsida](#).

Textmaterialet som analyserats i studien består av de elevtexter som använts för att illustrera betygsstegen E respektive A avseende skriftlig förmåga i de bedömningsmaterial till de nationella proven som Skolverket tillhandahöll lärare vid Sveriges skolor som stöd för bedömningen under de aktuella åren (för exempel, se www.nafs.gu.se). En ny betygsskala, F–A, infördes 2011 och därför används i denna studie enbart texter som skrivits därefter och bedömts enligt den nya skalan. De elevtexter som återfinns i Skolverkets material har inför varje nationellt prov sällats ut genom en bedömningsprocess där lärargrupper om ca

15 erfarna lärare från olika skolor i landet ingår tillsammans med provkonstruktörer vid Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet. Inför ett så kallat bedömningsmöte läser och bedömer mötesdeltagarna var och en för sig ett antal (30–50) elevtexter som samlats in vid utprovning av den aktuella skrivuppgiften. Vid mötet jämförs bedömningarna och bland de bedömda texterna görs ett urval av texter som ska illustrera de olika betygsnivåerna. Detta urval tas fram av gruppen gemensamt för att sedan fastställas av Skolverket. Det betyder att de texter som ingår i denna studie inte bara bedömts av en enskild lärare utan att de genomgått en relativt omfattande bedömningsprocess där ett stort antal bedömare deltagit, vilken stärker reliabiliteten i bedömningen (se Erickson, 2010).

När exempeltexter väljs ut till bedömningsmaterialet strävar man oftast efter att välja minst en text som illustrerar var den nedersta gränsen går för betyget, dvs. en text som endast nått och jämnt anses uppfylla kunskapskraven för det aktuella betygssteget. Det kan exempelvis vara viktigt att illustrera var gränsen går mellan F och E med en text som bedöms uppnå kunskapskraven för betyget E med ingen eller liten marginal. Ofta väljer man också en text som illustrerar den övre delen av betygssteget, där betygskraven alltså anses väl uppnådda, men däremot inte kraven för högre betygssteg. När det gäller betyget A finns inget högre betygssteg att avgränsa mot men man strävar i allmänhet efter att välja åtminstone en text som med knapp marginal uppfyller betygskraven för A för att illustrera var B övergår i A.

Eftersom ett syfte med denna studie är att åskådliggöra variationen inom en årskurs valdes texter som bedömts illustrera betygsstegen E och A men inte C, D eller B eftersom ytterligheterna tydligare visar spännvidden inom en årskurs än vad de mellanliggande betygsstegen gör. Eventuell överlappning mellan årskurser kan också synliggöras på ett tydligt sätt när högsta betyg i en lägre årskurs jämförs med lägsta betyg i en högre årskurs. En jämförelse mellan samtliga betygssteg i alla fyra årskurserna skulle riskera att bli för omfattande och detaljerad, och därmed mindre tydlig. Eftersom betygssteget E brukar delas upp i två nivåer i Skolverkets bedömningsmaterial – E nedre då texten bedömts uppfylla kraven med ingen eller knapp marginal respektive E övre då kraven med större marginal bedömts uppfylla – gjordes även vissa statistiska jämförelser mellan dessa två nivåer.

I studien analyseras totalt 71 texter som använts för att illustrera skriftlig förmåga på E- och A-nivå i de nationella proven i engelska i grundskolans årskurs 6 (N=17) och årskurs 9 (N=13), samt i gymnasieskolans kurser Engelska

5 (N=21) och Engelska 6 (N=20). I textmaterialet ingår 42 texter på E-nivå och 29 texter på A-nivå. I Tabell 1 (se avsnitt 5) anges antalet texter på varje betygsnivå per årskurs/kurs. Att antalet texter från grundskolans prov är färre i det använda textmaterialet än antalet texter från gymnasiet beror huvudsakligen på att den nuvarande betygsskalan användes i grundskolans prov först år 2013, medan gymnasieproven använde skalan från 2012; som redan nämnts har enbart texter bedömda efter skalan F–A använts i studien. Antalet exempel som ges för olika betygssteg i Skolverkets bedömningsanvisningar varierar också mellan åren. Vad gäller gymnasieskolans prov i kurserna Engelska 5 och Engelska 6 ges dessa i regel två gånger per år, men i denna studie har enbart vårterminsprov tagits med för att begränsa antalet texter i förhållande till de texter som fanns tillgängliga från grundskolans prov.

4 Analysmetoder

För analysen av textlängd, ordvariation och ordlängd har ett webbaserat verktyg, WordSmith Tools 5.0 (<http://www.lexically.net/wordsmith>), använts. Textlängd anges i textens totala antal ord. Måttet för ordvariation beräknas genom att antalet olika ord (*types*) divideras med det totala antalet ord (*tokens*) i texten. Om exempelvis ordet *telephone* förekommer 5 gånger i en text räknas det som en *type* men som fem *tokens*. För att korrigera för olika textlängd har ordvariationen beräknats efter var femtionde ord i texten, varefter textens medel för ordvariation räknats ut. I WordSmith Tools kallas denna modell för beräkning av ordvariation Standardised Type/Token Ratio, STTR. Ett högre värde visar en större ordvariation än ett lägre. Ordlängd har mätts i antal bokstäver som varje ord består av, varefter den genomsnittliga ordlängden i varje text beräknats.

För analysen av hur frekventa orden i elevtexterna är i det engelska språket har analysverktyget Vocabprofile (<http://www.lextutor.ca>) använts. I denna studie har texternas vokabulär analyserats mot frekvensband som integrerat listor från två omfattande textkorpusar, *The British National Corpus* (BNC; Nation, 2004) och *The Corpus of Contemporary American English* (COCA; Davies, 2012). I BNC ingår texter om totalt 100 miljoner ord och i COCA 450 miljoner ord. Båda korpusarna består av olika typer av texter, till exempel fiktion, tidningstext, talad engelska och akademiska texter. Frekvensband 1 (*K1*) innehåller de 1 000 mest frekventa orden i korpusarna, frekvensband 2 (*K2*) de därpå följande 1 000 vanligaste orden osv. Orden i frekvensband 1 är alltså mer vanligt

förekommande än de i frekvensband 2, medan orden i frekvensband 2 är mer vanligt förekommande än de i frekvensband 3 osv. Exempel på ord i materialet som kategoriseras som KI är *family*, *bad*, *more* och *start*. Inom kategorin K2 förekommer exempelvis *reduce*, *access*, *social* och *effect*, och i kategorin K3 *restrict*, *hence*, *solve* och *reluctant*.

I Vocabprofile kan ord inom 25 frekvensband analyseras. Ord som inte förekommer i dessa 25 frekvensband sorteras som *Off-list*. I analys av elevtexter återfinns i kategorin Off-list ofta felstavade ord eller ord som inte finns på engelska. Exempel på ord i materialet som kategoriserats som Off-list är *birday* (birthday), *cryid* (cried), *wich* (which) och *skola* (sv). I denna studie har inga rättningar av stavfel eller andra fel gjorts. Dock har egennamn, som hamnar i kategorin Off-list, tagits bort innan andelen Off-list ord räknats ut. Mycket ovanliga ord, närmare bestämt ord som inte finns bland de 25 000 vanligaste i den korpus frekvensbanden baseras på, hamnar också i kategorin Off-list. I materialet återfinns exempelvis *wellbeing* och *wrongdoer* i texter skrivna i gymnasiet, men här handlar det om enstaka ord i enstaka texter. Att närmare kategorisera olika typer av ord inom kategorin Off-list eller inom de olika frekvensbanden låg utanför denna pilotstudies ram.

I studien har andelen ord i elevtexterna som återfinns inom de olika frekvensbanden K1 till K25 analyserats, liksom hur stor andel av orden som tillhör kategorin Off-list. Eftersom ord i frekvensband bortom K2 är förhållandevis ovanliga har kategorierna K3 till K25 slagits samman till en kategori (*K3–K25*) för att möjliggöra jämförelser mellan texter; många elevtexter innehåller bara några få sådana ord.

Efter analys av varje texts totala längd, ordvariation, genomsnittliga ordlängd samt andel ord i olika frekvensband beräknades medelvärden för de två betygsstegen E och A i de fyra årskurserna/kurserna (åk 6, åk 9, Engelska 5 och Engelska 6) i syfte att jämföra texter som bedömts uppnå kraven för dessa betygssteg inom och mellan olika årskurser (se Tabell 1 och 2). När medelvärden anges ser man dock inte om det finns stor eller liten variation inom en grupp. För att undersöka om skillnader i värden mellan årskurser eller betygssteg är statistiskt signifikanta har statistiska analyser genomförts: T-test vid jämförelse mellan betygsstegen (E och A), och envägs ANOVA med Tukeys post hoc vid jämförelser mellan årskurser/kurser. Dessa analyser har följts upp med tvåvägs ANOVA för att undersöka eventuella interaktionseffekter mellan betygssteg och årskurs, dvs. i vilken utsträckning skillnaden mellan E- och A-nivå vad

gäller den undersökta språkliga variabelns värde är konstant eller skiljer sig åt mellan årskurser/kurser.

I studien ingår som redan nämnts 71 elevtexter. När dessa delas upp per årskurs och betygsnivå måste det ringa antalet texter som medelvärdena baseras på beaktas. Även i tolkningen av resultat av T-test och ANOVA-analyser manas till försiktighet – inga generella slutsatser kan dras. Däremot kan undersökningens resultat indikera skillnader och likheter mellan eller inom grupper (dvs. årskurser och betygssteg) som kan vara av intresse att undersöka närmare i en fortsatt, mer omfattande studie. Denna studie betraktas, som redan nämnts, som en pilotstudie.

5 Resultat

Inledningsvis presenteras resultat av analyser av textlängd, ordvariation och ordlängd i elevtexterna, följt av resultat av frekvensbandsanalyser, dvs. analyser av hur frekventa de ord som förekommer i elevtexterna är i engelska.

I Tabell 1 visas den genomsnittliga textlängden, ordvariationen och ordlängden per årskurs och betygssteg. Medelvärden anges för texter som bedömts visa betygsstegen E respektive A i varje årskurs/kurs. Antal texter som medelvärdena baseras på framgår i tabellen.

Tabell 1. Textlängd, ordvariation och ordlängd (medeltal)

Årskurs/kurs	Åk 6		Åk 9		Engelska 5		Engelska 6	
	E N=8	A N=9	E N=8	A N=5	E N=13	A N=8	E N=13	A N=7
Textlängd (antal ord)	107.4	238.2	271.8	415.0	307.4	481.4	350.4	495.4
Ordvariation (STTR)	66.4	74.6	72.7	82.9	76.2	78.6	77.4	81.4
Ordlängd (antal bokstäver)	3.4	3.8	3.9	4.2	4.0	4.2	4.1	4.4

Nedan redovisas resultaten av statistiska jämförelser mellan årskurser och betygssteg vad gäller elevtexternas längd, ordvariation och ordlängd. I T-test jämförs värden för de undersökta språkliga aspekterna mellan texter som bedömts visa E-nivå (N=42) och texter som bedömts visa A-nivå (N=29), oavsett årskurs. Resultaten av T-test visar om skillnaderna mellan E- och A-

texter är statistiskt signifikanta eller inte. Genom envägs ANOVA jämförs textlängd, ordvariation och ordlängd i texter skrivna i olika årskurser/kurser: årskurs 6 (N=17), årskurs 9 (N=13), Engelska 5 (N=21) och Engelska 6 (N=20). Resultaten visar om skillnader mellan texter skrivna i olika årskurser/kurser, med avseende på de undersökta språkliga aspekterna, är statistiskt signifikanta eller inte. Dessutom redovisas resultat av tvåvägs ANOVA-analyser som visar om det finns någon interaktionseffekt mellan de båda faktorerna årskurs och betygsnivå, till exempel om skillnader mellan E- och A-nivå vad gäller de undersökta språkliga variabelernas värden är större i vissa årskurser än i andra.

5.1 Textlängd

Tabell 1 visar att den genomsnittliga textlängden mätt i antal ord är kortare i de texter som bedömts visa betygsnivån E än i texter som bedömts illustrera A inom varje årskurs/kurs. En statistisk T-test visar att skillnaden i textlängd mellan A- och E-texter är signifikant ($t= 3,86, p < .001$). Resultatet indikerar, som kan förväntas, att elever som bedöms ha god förmåga att skriva på engelska tenderar att skriva längre texter än elever som inte kommit lika långt i sin skrivutveckling.

Eftersom yngre barn kan förväntas skriva kortare texter än gymnasieelever och uppgiftsinstruktionerna dessutom skiljer sig åt vad gäller angivelse av förväntat antal ord är det svårt att jämföra textlängd mellan olika årskurser. Trots det kan några iakttagelser ändå göras. Tabell 1 visar att de texter som illustrerar betygsnivån A i årskurs 9 i genomsnitt är längre än de texter som illustrerar E i Engelska 5, likaså att A-texter i Engelska 5 i genomsnitt är längre än E-texter i Engelska 6. En viss överlappning mellan årskurser och kurser kan alltså skönjas här, eftersom elever som når målen för A i en lägre årskurs i genomsnitt skriver längre texter än elever som når målen för E i en högre årskurs i några fall.

Om man jämför längden av texter skrivna i de fyra årskurserna/kurserna utan att ta hänsyn till betygsnivå visar en statistisk analys (envägs ANOVA) att det finns vissa signifikanta skillnader i textlängd¹². Texterna i årskurs 6 är, som förväntat, signifikant kortare än texterna i årskurs 9 och än gymnasietexterna (Åk 9 $p= .002$, Eng 5 och Eng 6 $p < .001$). Analysen visar att det i övrigt inte finns några signifikanta skillnader; skillnaderna i textlängd mellan årskurs 9,

¹² $F[3, 67]= 15,91, p < .001$

Engelska 5 och Engelska 6 är alltså inte statistiskt signifikanta. Den överlappning mellan årskurs 9 och gymnasiekurserna som medelvärden i Tabell 1 visar bekräftas i den statistiska signifikansprövningen.

En tvåvägs ANOVA-analys visar att den inte finns någon signifikant interaktionseffekt mellan betygssteg, årskurs och textlängd¹³. Detta betyder att skillnaden i textlängd mellan texter som bedömts visa E- respektive A-nivå är relativt konstant över de olika årskurserna, dvs. skillnaden i textlängd mellan E- och A-texter är inte betydligt större eller mindre i någon årskurs.

5.2 Ordvariation

Tabell 1 visar att ordvariationen följer ett liknande mönster inom årskurserna som det vi såg för textlängd. Ordvariationen är således större i medeltal i texter som bedömts visa betygssteget A jämfört med texter som bedömts uppfylla kraven för E inom årskurserna/kurserna. En statistisk T-test visar att ordvariationen totalt sett är signifikant högre i de texter som bedömts uppnå kraven för A än i de texter som uppnått E ($t=3.65, p=.001$), utan hänsyn tagen till årskurs. Resultatet indikerar att förmåga att variera ordvalet tycks ha betydelse för bedömningen av texten och att goda skribenter, som förväntat, använder en rikare repertoar av ord.

Tabell 1 visar att den genomsnittliga ordvariationen ökar i högre årskurser, men man kan här, liksom när det gällde textlängd, notera flera fall av överlappning. Texterna som illustrerar A-nivå i årskurs 6 har något högre medelvärde avseende ordvariation än de texter som illustrerar betyget E i årskurs 9 trots att tre år passerar mellan årskurs 6 och 9 då elever kan förväntas lära sig en hel del ord. Vidare innehåller A-texter i årskurs 9 i genomsnitt en mer varierad vokabulär än texterna som illustrerar både E- och A-nivåer i Engelska 5 och 6. Redan i årskurs 9 varierar alltså en del elever – de som bedömts uppnå kraven för A – sin vokabulär i samma utsträckning som gymnasieelever som uppnått kraven för A i gymnasiets kurser med avseende på skriftlig kommunikation och interaktion. Några tänkbara förklaringar till dessa något anmärkningsvärda resultat presenteras i detta kapitelns diskussionsavsnitt.

När ordvariationen i texter skrivna i olika årskurser/kurser jämförs, utan hänsyn tagen till betygssteg, visar en envägs ANOVA att det finns vissa statistiskt signifikanta skillnader mellan de fyra grupperna (åk 6, åk 9, Eng 5 och Eng 6)¹⁴.

¹³ $F[3, 63] = .26, p = .854$

¹⁴ $F[3, 67] = 8.58, p < .001$

Det är dock bara skillnaderna i ordvariation mellan årskurs 6 och de tre övriga grupperna som är signifikanta (åk 9 $p = .012$, Eng 5 $p = .001$, Eng 6 $p < .001$). Det finns, som redan antytts, ingen signifikant skillnad i ordvariation mellan texter skrivna i årskurs 9 och de som skrivits i de båda gymnasiestegen.

En tvåvägs ANOVA-analys visar att det finns en signifikant interaktions-effekt mellan betygsnivå och årskurs vad avser ordvariation¹⁵. Skillnaden i ordvariation mellan texter som bedömts motsvara betygsstegen E och A är större i årskurs 6 och 9 än i gymnasiets kurser Engelska 5 och Engelska 6.

5.3 Ordlängd

Tabell 1 visar att texter som illustrerar betygssteget A i genomsnitt innehåller längre ord än texter som illustrerar E-nivå inom varje årskurs/kurs. T-test visar att orden i A-texter i genomsnitt är signifikant längre än orden i texter som bedömts uppnå kraven för E ($t=2,97$, $p = .004$).

Tabell 1 visar vidare att den genomsnittliga ordlängden i texterna ökar över årskurserna men att det även här finns en viss överlappning mellan årskurs 9 och gymnasiekurserna. De texter som illustrerar A-nivå i årskurs 9 har en genomsnittlig ordlängd på 4.2 bokstäver, vilket är samma värde som för A-texter i gymnasiets Engelska 5, och dessutom något högre värde än för E-texter i gymnasiets Engelska 6, där den genomsnittliga ordlängden är 4.0 bokstäver. Det finns alltså en stor överlappning mellan årskurs 9 och såväl Engelska 5 som Engelska 6 vad avser genomsnittlig ordlängd.

När den genomsnittliga ordlängden jämförs i texter skrivna i de fyra årskurserna/kurserna (åk 6, åk 9, Eng 5 och Eng 6), utan hänsyn tagen till olika betygssteg, visar envägs ANOVA att det finns vissa signifikanta skillnader mellan årskurserna¹⁶. Liksom i analysen av textlängd och ordvariation är det även när det gäller ordlängd dock bara skillnaderna mellan årskurs 6 och övriga årskurser som är statistiskt signifikanta. Ordlängden i texter skrivna i årskurs 6 är signifikant kortare jämfört med de tre övriga grupperna (åk 9 $p < .001$, Eng 5 $p < .001$, Eng 6 $p < .001$). Mellan de övriga gruppernas texter, dvs. mellan årskurs 9, Engelska 5 och 6, finns alltså ingen signifikant skillnad i genomsnittlig ordlängd. Den överlappning som medelvärdena indikerar bekräftas således i den statistiska analysen.

¹⁵ $F[3, 63] = 3.29$, $p = .026$

¹⁶ $F[3,67] = 20.98$, $p < .001$

Resultatet av en tvåvägs ANOVA-analys visar att det inte finns någon signifikant interaktionseffekt mellan betygssteg, årskurs och genomsnittlig ordlängd¹⁷. Det innebär att skillnaden i ordlängd mellan texter som bedömts visa E- respektive A-nivå är tämligen konstant över årskurserna, dvs. den är inte betydligt större eller mindre i någon årskurs.

5.4 Frekvensbandsanalys

Resultaten av frekvensbandsanalysen, dvs. hur frekventa de ord som förekommer i elevtexterna är i målspråket, visas i Tabell 2. Tabellen visar hur stor andel (i procent) av texternas ord som återfinns i olika frekvensband, dvs. om orden är de allra vanligaste i engelskan eller om texterna också innehåller ord utöver de vanligast förekommande. Frekvensband *K 1* visar den andel av texternas ord som återfinns bland de 1 000 mest frekventa orden i engelska. Frekvensband *K 2* visar de ord som följer efter de tusen vanligaste, dvs. ord 2 001–3 000 baserat på ordens frekvens i den korpus listorna grundar sig på. De därpå följande frekvensbandens ord är således ännu mer ovanliga och därför har de här, som redan nämnts, slagits samman till en kategori kallad *K 3–K 25*, eftersom eleverna använde få sådana ord i sina texter. Ord som kategoriserats som *Off-list* finns inte med i de 25 frekvensbanden. Det kan exempelvis röra sig om felstavade ord, eller ord som inte är engelska men som eleverna ändå använt i sina texter. Andelen ord i elevtexterna som tillhör de olika frekvensbanden anges i Tabell 2 per årskurs/kurs och betygsnivå (E och A). Figur 1 (se 5.4.2) och Figur 2 (se 5.4.3) är grafiska illustrationer av resultat som visas i Tabell 2.

Tabell 2. Andel ord inom olika frekvensband (%)

Årskurs/kurs	Åk 6		Åk 9		Engelska 5		Engelska 6	
Betygsnivå	E N=8	A N=9	E N=8	A N=5	E N=13	A N=8	E N=13	A N=7
K 1 (%)	83.7	91.8	87.9	88.6	87.0	88.2	89.0	87.4
K 2 (%)	2.7	2.4	2.1	5.0	3.4	6.1	3.9	5.9
K 3–25 (%)	0.7	0.9	1.8	3.6	2.2	3.5	1.8	4.4
Off-list (%)	13.5	4.9	8.1	2.9	7.3	2.1	4.8	2.2

Nedan redovisas vidare resultaten av signifikanstest (T-test och envägs ANOVA) som visar om skillnader mellan årskurser eller betygssteg, vad gäller förekomst av

¹⁷ $F[3, 63] = .37, p = .775$

ord inom olika frekvensband, är statistiskt signifikanta eller inte. Likaså redovisas resultat av tvåvägs ANOVA som visar om det finns några interaktionseffekter mellan betygssteg, årskurs och andelen ord i olika frekvensband.

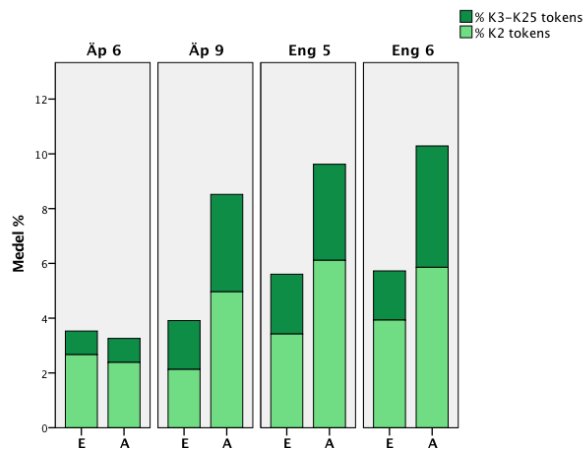
5.4.1 Frekvensband K1

Tabell 2 visar att de ord som till övervägande delen används av eleverna i texterna, i alla årskurser och på alla betygsnivåer, återfinns bland de 1 000 vanligaste orden i engelska; mellan 83.7 % och 91.8 % av texternas vokabulär hör till kategorin K1. Vid en första anblick kan man förvånas över att texter i årskurs 6 som bedömts visa E-nivå innehåller en lägre andel mycket vanliga ord än texter skrivna av elever i årskurs 6 som bedömts visa A-nivå, och även en lägre andel än texter skrivna i högre årskurser. När man avläser tabellen bör man dock notera hur stor andel av orden som kategoriseras som *Off-list*, dvs. hur stor andel av orden som inte förekommer i något av frekvensbanden eftersom de exempelvis är felstavade. Att andelen ord som tillhör K1 är lägre i årskurs 6 än i högre årskurser beror alltså inte på att eleverna där använder fler ovanliga ord utan på att de använder en större andel felstavade ord eller ord som inte finns i engelskan.

Statistiska analyser (T-test och envägs ANOVA) visar att det inte finns någon signifikant skillnad i andelen ord som tillhör kategorin K1 vare sig mellan betygsstegen E och A eller mellan årskurser/kurser. Däremot visar en tvåvägs ANOVA att det finns en signifikant interaktionseffekt mellan betygssteg, årskurs och andelen ord i kategorin K1¹⁸. Analysen visar att det är signifikant större skillnad i andelen sådana ord i årskurs 6 än i högre årskurser, vilket som ovan nämnts tycks hänga samman med den höga andelen ord i kategorin *Off-list* i texter som bedömts visa E-nivå i årskurs 6.

¹⁸ $F[3,63]=3.44, p=.022$

5.4.2 Frekvensband K2 och K3–K25



Figur 1. Andelen ord inom frekvensband K2 och K3–K25

I Tabell 2 liksom i Figur 1 visas att den genomsnittliga andelen ord bortom de 1 000 vanligaste är högre i texter som bedömts motsvara A-nivå än de som bedömts motsvara E-nivå utom i årskurs 6. Man kan misstänka att E-texternas korta längd i årskurs 6 (se Tabell 1) innebär att några få ord som hör till frekvensband K2 ger god utdelning i procentberäkningarna, men det kan naturligtvis finnas andra kvalitetsaspekter som ger utslag. Exempelvis visar Tabell 2 och Figur 2, som redan nämnts, att andelen ord i kategorin Off-list är stor i texterna som illustrerar E-nivå i årskurs 6, vilket kan tänkas dra ner omdömet även om texten även innehåller en del ovanliga ord. T-test visar dock att andelen ord som tillhör kategorierna K2 respektive K3–K25 är signifikant större i texter som bedömts motsvara kraven för A jämfört med i texter som bedömts motsvara E (K2: $t=2.84$, $p=.006$; K3–K25: $t=2,62$, $p=.012$).

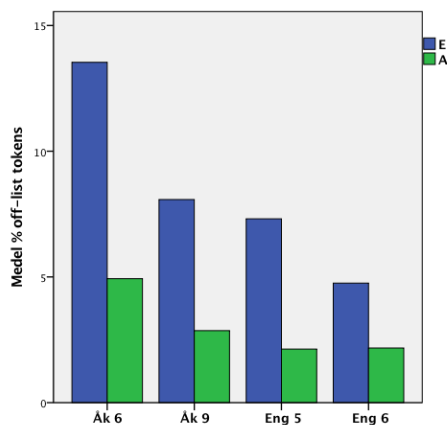
Vidare visar Tabell 2 och Figur 1 att den genomsnittliga andelen ord bortom de tusen vanligaste är större i högre årskurser jämfört med lägre, även om överlappning finns. Exempelvis är den genomsnittliga andelen ord som tillhör kategorierna K2 eller K3–K25 i A-texter i årskurs 9 större än andelen sådana ord i E-texter skrivna i Engelska 5 och Engelska 6. Figur 1 visar som förväntat att störst andel ord bortom de 1 000 vanligaste finns i de texter som bedömts motsvara kraven för A i Engelska 6.

Envägs ANOVA där andelen ord jämförs i texter skrivna i olika årskurser/kurser, utan hänsyn tagen till betygsnivå, visar att det finns vissa signifikanta

skillnader såväl när det gäller andelen ord i kategorierna K2¹⁹ som K3–K25²⁰. Vad avser kategorin K2 innehåller texterna i årskurs 6 en signifikant mindre andel sådana ord än texterna i Engelska 5 och 6 (Eng 5: $p=.041$; Eng 6: $p=.25$). Vad gäller kategorin K3–K25 innehåller texterna i årskurs 6 en signifikant mindre andel av dessa ord än texterna i de andra tre årskurserna (Åk 9: $p=.017$; Eng 5: $p=.004$; Eng 6: $p=.004$). I övrigt finns inga statistiskt signifikanta skillnader mellan årskurser och betygssteg vad avser de nämnda frekvensbanden. Den statistiska analysen bekräftar således den överlappning som medelvärdena indikerade mellan årskurs 9 och gymnasiets kurser vad avser användning av vokabulär utöver de allra vanligaste orden.

Resultaten av tvåvägs ANOVA visar att det inte finns några signifikanta interaktionseffekter mellan betygssteg, årskurs och andelen ord i frekvensband K2²¹ respektive K3–25²². Dock är interaktionseffekten i båda fallen nästan signifikant eftersom skillnaden i andel sådana ord mellan E- och A-texter i årskurs 6 är i det närmaste obefintlig medan skillnaden är större och relativt konstant över högre årskurser.

5.4.3 Kategorin Off-list



Figur 2. Andelen ord i kategorin *Off-list*

¹⁹ $F[3,65]=3.82, p=.014$

²⁰ $F[3,66]=5.79, p=.001$

²¹ $F[3,61]=2.18, p=.100$

²² $F[3,63]=2.24, p=.092$

Tabell 2 och Figur 2 visar att de texter som illustrerar betyget E i genomsnitt innehåller en större andel ord inom kategorin Off-list, vilket oftast (i detta material) innebär ord som är felstavade eller inte finns i engelskan, jämfört med i texter som illustrerar betyget A inom varje årskurs/kurs. T-test bekräftar att andelen sådana ord är signifikant större i E-texter än i A-texter ($t=5.06$ $p < .001$).

Likaså minskar den genomsnittliga andelen ord i kategorin Off-list högre årskurser. Resultat av envägs ANOVA visar att det finns vissa signifikanta skillnader i andelen sådana ord mellan de fyra årskurserna/kurserna²³. Texterna i årskurs 6 innehåller en signifikant större andel ord i kategorin Off-list än texter skrivna i gymnasiets kurs Engelska 6 ($p = .010$), men inte i jämförelse med texter skrivna i de andra årskurserna. I övrigt finns inga signifikanta skillnader i andelen ord som tillhör kategorin Off-list mellan årskurserna. Detta visar att det finns en stor variation i andelen sådana ord inom de olika årskurserna.

En tvåvägs ANOVA-analys visar att det inte finns någon signifikant interaktionseffekt mellan betygssteg, årskurs och texternas andel ord i kategorin Off-list²⁴, vilket indikerar att skillnaden mellan betygsstegen E och A i detta avseende är relativt konstant över årskurserna. Dock tycks skillnaden vara något större i årskurs 6 än i övriga årskurser.

5.5 Jämförelse mellan E nedre och E övre

Som nämndes i den inledande beskrivningen av textmaterialet som analyserats i studien brukar Skolverkets bedömningsanvisningar differentiera betygssteget E i två nivåer, E nedre och E övre, för att illustrera om målen för E är uppnådda med knapp marginal eller om målen är väl uppfyllda²⁵. Bland de 42 texter i materialet som bedömts uppnå kraven för E finns 23 som bedömts visa nivån E nedre och 19 som bedömts visa nivån E övre. T-tester genomfördes för att undersöka om det finns några statistiskt signifikanta skillnader mellan de två nivåerna med avseende på de undersökta språkliga aspekterna, dvs. textlängd, ordvariation, ordlängd och andel vokabulär i olika frekvensband. Inga signifikanta skillnader finns i dessa avseenden mellan de två E-nivåerna.

²³ $F[3,67]=3.67, p = .016$

²⁴ $F[3,63]=1.73, p = .171$

²⁵ I bedömningsanvisningar för årskurs 6 görs inte uppdelningen i E nedre och E övre men däremot beskrivs i ord om målen är uppfyllda med knapp eller god marginal. Här likställs de beskrivningarna med de två nivåerna, E nedre och E övre, som övriga årskurser och kurser använder.

6 Diskussion

Syftet med denna pilotstudie har varit att på några olika sätt belysa ordförrådets betydelse för förmåga att uttrycka sig i skrift på engelska samt för bedömning av denna förmåga. I detta syfte har 71 elevtexter skrivna inom ramen för de nationella proven i engelska jämförts för att undersöka likheter och skillnader mellan texter som bedömts visa betygsstegen E respektive A, liksom mellan texter skrivna i olika årskurser eller kurser, närmare bestämt grundskolans årskurs 6 och 9 samt gymnasieskolans kurser Engelska 5 och Engelska 6. Jämförelser har gjorts av texternas längd, ordvariation och ordlängd; språkliga aspekter som tenderar att samvariera med helhetsbedömning av textkvalitet enligt tidigare forskning (se t.ex. Grant & Ginther, 2000). Vidare har frekvensbandsanalyser gjorts för att undersöka hur frekventa de ord som förekommer i elevtexterna är i det engelska språket, dvs. i vilken utsträckning eleverna enbart använder mycket vanligt förekommande ord eller även mer lågfrekvent vokabulär. Det begränsade antalet texter i studien måste beaktas i tolkningen av resultaten; inga generella slutsatser kan dras. Resultaten kan ändå indikera de undersökta språkliga aspekternas eventuella betydelse för helhetsbedömning av elevtexter. Även ett begränsat material kan bidra till att illustrera likheter och skillnader mellan olika betygsnivåer och årskurser, inte minst vad gäller överlappning mellan årskurser och kurser.

Studiens resultat visar att det finns signifikanta skillnader mellan texter som bedömts uppnå kraven för E respektive A, vilket givetvis inte är förvånande eftersom kraven för att nå A är högre än de för E. Texter som bedömts uppfylla kraven för A var signifikant längre, uppvisade signifikant större ordvariation, innehöll i genomsnitt längre ord och en signifikant större andel ovanliga ord än texter som bedömts uppnå kraven för E. Dessa språkliga drag kan därför ses som utmärkande för de texter i materialet som bedömts visa A-nivå, även om det givetvis finns andra drag som också kan vara utmärkande för texter som bedömts nå kraven för A. Resultaten indikerar att de undersökta språkliga dragen kan ha en viss betydelse vid bedömningen, i linje med vad tidigare forskning visat (se t.ex. Grant & Ginther, 2000; Hinkel, 2011; Crossley, Salsbury & McNamara, 2012). Man kan också notera att skillnaderna i textlängd, ordvariation, ordlängd och användning av mer ovanlig vokabulär inte var signifikanta i jämförelsen mellan texter som bedömts visa nivån E nedre respektive E övre. Inom spannet E-nivå tycks alltså variationen vara sådan att skillnader mellan E nedre och E övre inte är tydliga med avseende på de undersökta

aspekterna, i alla fall inte i ett begränsat textmaterial som det som användes här. I ett större material vore det intressant att närmare undersöka skillnader mellan fler betygssteg inom årskurser och kurser.

Analysen av skillnader i textlängd, ordvariation, ordlängd och användning av mer ovanlig vokabulär mellan årskurser visar att det finns signifikanta skillnader mellan texter skrivna i årskurs 6 och texter skrivna i de andra årskurserna/kurserna, dvs. i årskurs 9 och i gymnasiekurserna Engelska 5 och Engelska 6. Skillnaderna är inte signifikanta mellan texter skrivna i årskurs 9 och i de båda gymnasiekurserna, och inte heller mellan de två gymnasiekurserna. Analysen av interaktionseffekter visade att det var större skillnad i ordvariation mellan E- och A-texter i årskurs 6 än i övriga årskurser, och även att skillnaden i användning av de mest frekventa orden var större där. I övrigt var skillnaderna mellan E och A, vad avser de undersökta språkliga aspekterna, relativt konstanta över årskurserna.

En tänkbar orsak till att skillnaderna tycks vara störst mellan texter skrivna i årskurs 6 och texter skrivna i övriga årskurser skulle kunna kopplas till utveckling över tid. Tidigare forskning har visat att ett års exponering för eller undervisning i ett språk kan betyda mycket för en individs språkutveckling (se t.ex. Muñoz, 2008). Mellan de nationella proven i årskurs 6 och årskurs 9 förflyter tre år av, exempelvis, engelskundervisning i skolan och fritidsaktiviteter där engelska används i större eller mindre utsträckning, medan det bara går ett eller två år mellan de därpå följande proven. Detta är sannolikt en del av förklaringen till de signifikanta skillnaderna mellan årskurs 6 och årskurs 9 – elevers kunskaper hinner utvecklas mycket på tre år. Man kan ändå notera att texter på A-nivå i årskurs 6 i genomsnitt uppvisar lika stor ordvariation som texter som bedömts uppnå kraven för E i årskurs 9 och att skillnaden i genomsnittlig ordvariation (och även textlängd) är stor mellan texter på A- respektive E-nivå i årskurs 6. Resultatet visar att spännvidden inom årskurserna tycks vara stor, och detta redan i årskurs 6. En bidragande orsak till att en del elever redan i årskurs 6 tycks ha utvecklat ett lika varierat ordförråd som elever på E-nivå i årskurs 9 skulle kunna vara att de i stor utsträckning använder engelska på fritiden, vilket tidigare forskning visat kan ha god effekt på just ordförrådet (se t.ex. Sylvén & Sundqvist, 2012). I denna studie undersöktes dock inte elevernas användning av engelska på fritiden eller andra bakgrundsfaktorer, men resultaten visar alltså att det finns stora skillnader mellan elevernas engelskkunskaper redan i årskurs 6.

Vidare visar resultaten att det även i årskurs 9 finns en stor spännvidd mellan elevtexter vad gäller de undersökta språkliga aspekterna. Några texter uppvisar

som nyss nämnts en ordvariation på liknande nivå som i årskurs 6 medan texterna på A-nivå i årskurs 9 uppvisar lika stor ordvariation som texterna på A-nivå i Engelska 6. Eftersom skillnaderna i ordvariation, text- och ordlängd inte är signifikanta i jämförelsen mellan texter skrivna i årskurs 9 och i gymnasiekurserna betyder det att överlappningen är betydande liksom variationen inom årskurserna.

Skillnad i texttyp som olika skrivuppgifter initierat skulle i viss mån kunna förklara varför skillnaden i ordvariation inte är större mellan texter skrivna i årskurs 9 och i gymnasiet. Korpusstudier har nämligen visat att akademiska texter tenderar att ha lägre ordvariation än litterära texter (se t.ex. Nishina, 2007). De nationella provens skrivuppgifter i årskurs 6 och 9 är ofta formulerade så att berättande och/eller beskrivande texter kan förväntas, särskilt i årskurs 6. I gymnasiet provuppgifter uppmanas eleverna oftare att argumentera eller utreda. Alltså skulle den något lägre ordvariationen i gymnasiet jämfört med texterna på A-nivå i årskurs 9 kunna bero på anpassning till texttypen. Dock skulle fördjupade analyser behövas för att närmare undersöka detta vilket ligger utanför vad som är möjligt att göra i denna studie.

Vad gäller textlängd anges ofta, som redan nämnts, en önskad textlängd i provuppgiften i årskurs 9 och i gymnasiet. Därmed visar inte texterna hur långa texter eleverna hade kunnat skriva utan begränsning. När man mäter skriftlig förmåga handlar det givetvis inte bara om flöde och att skriva mycket utan även om att exempelvis kunna anpassa texten till givna förutsättningar. Å andra sidan är det förstås lättare att skriva en lång text om man har ett gott ordförråd än om det är begränsat. Givet de förutsättningar som de olika provuppgifterna satt upp, visar resultaten av denna studie att texter som bedömts uppnå E-nivå är signifikant kortare än texter som bedömts uppnå A-nivå.

Frekvensbandsanalyserna visade att elever på A-nivå tycks ha en större produktiv vokabulär, där även mer ovanliga ord ingår än elever på E-nivå, vilket överensstämmer med vad studiens övriga resultat visat. Analysen visade vidare att det inte fanns signifikanta skillnader i användningen av mer ovanliga ord (K2, K3–K25) mellan årskurs 9 och gymnasiekurserna utan bara mellan årskurs 6 och de övriga årskurserna/kurserna. I linje med ovanstående diskussion om texttypens betydelse för ordvariation hade man kunnat förvänta sig att gymnasietexterna skulle innehålla en större andel mer lågfrekventa ord eftersom akademiska ord, som kan förväntas i utredande och argumenterande texter, är relativt lågfrekventa i engelska och därför ofta återfinns i frekvensband bortom de 2 000 vanligaste orden. Troligtvis är vokabulär bortom de 2 000 vanligaste

orden relativt ovanlig i den engelska elever kommer i kontakt med utanför skolan, även om det förstås beror på vad de gör, men möjligheten att stöta på den typen av ord utanför skolan tillräckligt ofta för att eleven ska tillägna sig orden så väl att eleven själv kan använda orden i skrift tycks begränsad. I skolan förekommer med stor sannolikhet denna typ av ord oftare och goda pedagoger har naturligtvis stora möjligheter att på olika sätt stötta elevers utveckling av produktiv vokabulär bortom de vanligaste orden.

De texter som använts i studien är, som nämnts, texter som illustrerat olika betygsnivåer i Skolverkets bedömningsmaterial. Det betyder att texterna som illustrerar E nedre och många av texterna som illustrerar A-nivå visar en miniminivå för det givna betyget. Det innebär att ett större textmaterial där exempelvis även mycket starka A-texter ingick med stor sannolikhet skulle visa att variationen inom och mellan betygsnivåer är betydligt större än vad som synliggjorts i denna begränsade studie.

7 Slutord

Man kan inte dra generella slutsatser grundade på resultaten av denna begränsade pilotstudie. Däremot bekräftar resultaten vad tidigare studier visat nämligen att ordförrådet tycks ha stor betydelse för elevers förmåga att skriva och därmed också för bedömningen av deras texter. Resultaten indikerar också att skillnader mellan A- och E-nivå bland annat manifesteras i skillnader i textlängd, ordvariation, ordlängd och användning av vokabulär som är lågfrekvent i engelska.

Man kan fundera på om nuvarande kunskapskrav och betygsnivåer, vilka är kopplade till de olika årskurserna, är de optimala för att beskriva elevers kunskapsnivå i engelska, med tanke på den spännvidd som verkar finnas inom en årskurs och den överlappning som finns mellan årskurser. I Skolverkets statistik över de nationella provens resultat kan man utläsa att en stor andel elever (t.ex. över 20 % läsåret 2015/16) uppnår betyget A i årskurs 6 och årskurs 9 (<https://www.skolverket.se/statistik-och-utvardering/>). Provresultaten ger däremot inte upplysning om elever besitter kunskaper på ännu högre nivå än just kraven för A för den aktuella årskursen/kursen.

Resultaten kan också ge anledning till reflektioner kring den utmanande uppgiften att i klasser där spännvidden är mycket stor bedriva undervisning som stimulerar och motiverar alla elever. Att exempelvis i en klass i årskurs 9 undervisa elever som precis eller inte alls uppnår kraven för E tillsammans med elever som troligtvis skulle klara att följa gymnasiekurser i engelska innebär en stor

utmaning för läraren. Det finns dock forskning som visar på att lärande och undervisning i heterogena grupper tycks ha gynnsamma effekter vad gäller såväl jämlikhet som kunskapsresultat (t.ex. Hanushek & Wössman, 2006). Att elever i mycket olika omfattning ägnar sig åt engelska på sin fritid kan innebära att spridningen blir större i engelska än i andra ämnen. Å andra sidan är det naturligtvis en stor förmån att undervisa i ett språk som de flesta elever anser är mycket viktigt att kunna och som många i stor utsträckning frivilligt använder på sin fritid. Det kan dock innebära att en del elever upplever att de lär sig mer engelska utanför skolan än i skolan och att motivationen minskar att anstränga sig på engelsklektionerna i skolan (se t.ex. Henry, 2014). Det finns anledning att fundera över hur man på bästa sätt kan organisera undervisning i och bedömning av engelska i svensk skola så att alla elever stimuleras att utveckla sina språkkunskaper oavsett förkunskaper. Man kan även reflektera över hur bedömning skulle kunna ske så att de kunskaper en elev faktiskt har synliggörs i bedömningen, även om eleven befinner sig på en högre nivå än förväntat för en viss ålder.

Referenser

- Cobb, T. (u.å.) *Web Vocabprofile*. (Hämtad 2017-02-11, från <http://www.lextutor.ca/vp/>). An adaptation of A. Heatley, P. Nation & A. Coxhead (2002). *Range and frequency*.
- Davies, M. (2012). *Corpus of contemporary American English (1990–2012)*. Hämtad 2016-08-24, från <http://corpus.byu.edu/coca/>.
- Elgort, I., & Nation, I.S.P. (2010). Vocabulary learning in a second language: Familiar answers to new questions. In P. Seedhouse, S. Walsh & C. Jenks (Eds.), *Conceptualizing learning in applied linguistics* (pp. 89–104). Houndmills: Macmillan.
- Erickson, G. (2010). Good Practice in Language Testing and Assessment – A Matter of Responsibility and Respect. In Kao, T. and Lin Y. (Eds.), *A New Look at Teaching and Testing: English as Subject and Vehicle* (pp. 237–258). Taiwan: Bookman Books Ltd
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Researching in Reading*, 75(2), 115–135.
- Crossley, S.A., Salsbury, T., & McNamara, D.S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243–263.
- Gass, S. (2013). *Second language acquisition: An introductory course*. New York: Routledge.

- Grant, L., & Ginther, A. (2000). Using computer tagged linguistic features to describe L2 writing differences. *Journal of second language writing*, 9 (2), 123–145.
- Hanushek, E.A. & Wössman, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences in Differences Evidence Across Countries. *The Economic Journal* 116(510), C63–C76.
- Henry, A. (2014). Swedish students' beliefs about learning English in and outside of school. In D. Lasagabaster, A. Doiz and J M Sierra (Eds). *Motivation and Foreign Language Learning: From Theory to Practice* (pp. 93–116). Amsterdam: John Benjamins
- Hinkel, E. (2011). What research on second language tells us and what it doesn't. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning. Vol. 2*. New York: Routledge.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19(2), 255–271.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook*, 5, 223–250.
- Laufer, B., & Nation, I.P.S. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322.
- Matsuda, P.K., Ortmeier-Hooper, C., & Matsuda, A. (2009). The expansion of second language writing. In R. Beard (Ed.), *Sage handbook of writing development* (pp. 457–471). Los Angeles: Sage.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578–596.
- Nation, I.S.P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Philadelphia: John Benjamins.
- Nation, I.S.P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nishina, Y. (2007). A Corpus Driven Approach to Genre Analysis: The Reinvestigation of Academic, Newspaper and Literary Texts. *ELR Journal*, 1 (2).
- Olsson, E. (2012). "Everything I read on the Internet is in English". *On the impact of extramural English on Swedish 16-year-old pupils' writing proficiency*. Göteborg: Göteborgs Universitet.
- Olsson, E. (2016). *On the impact of extramural English and CLIL on productive vocabulary*. (Doktorsavhandling). Göteborg: Göteborgs Universitet.
- Olsson, E., & Sylvén, L. K. (2015). Extramural English and academic vocabulary. A longitudinal study of CLIL and non-CLIL students in Sweden. *Apples – Journal of Applied Language Studies*, 9(2), 77–103.

VOKABULÄR OCH BEDÖMNING AV SKRIFTLIG FÖRMÅGA

- Scott, M. (2008). *WordSmith Tools*, version 5. Lexical Analysis Software. Hämtad 2017-02-11, från www.lexically.net/wordsmith.
- Skolverket (2011a). *Kommentarer till gymnasieskolans ämnesplan i engelska*. Stockholm: Skolverket.
- Skolverket (2011b). *Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*. Stockholm: Skolverket.
- Skolverket (2011c). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011 (Reviderad 2016)*. Stockholm: Skolverket.
- Skolverket (2012). *Kommentarmaterial till kunskapskraven i engelska*. Stockholm: Skolverket.
- Skolverket (2017). Statistik och utvärdering. Hämtad 2017-05-15, från <https://www.skolverket.se/statistik-och-utvardering/>
- Sundqvist, P. (2009). *Extramural English matters: Out of school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. (Doktorsavhandling). Karlstad: Karlstads Universitet.
- Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in Teaching and Learning: From Theory and Research to Practice*. Springer.
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24, 302–321.
- Wardle, E., & Roozen, K. (2012). Addressing the complexity of writing development: Toward an ecological model of assessment. *Assessing Writing*, 17(2), 106–119.

Bilaga 2 Exempel på bedömda elevtexter åk 9 (grundskolan)

Bedömning: E

Hey my name is N N and im 14 years old and I live in Sweden.

I want to join this camp beacuse, I want to meet peoples from other part of the world. I think this should be an good experience for me. that not have left Sweden. One other thing, my dad thinks that I is outside to little.

I have a suggest for some activities that we coul'de do at the camp. We coul'de play football outdoors, I think that should be interesting to learn how children from other part of the world play football. An indoor activite shoulde be learning other laugese I think that should be interesting and meaningful. If I for example went to France it should be good that I have learned some Franch. At the day time I think we should be outdoors a lot beacus it is good with sun shing. At the night we should learn some history from other part of the world that would be lerning, interesting, and meaningful.

And I think we should discuss war beatwen some countris, that is an large issue in the whole world. I think we should discuss that exampel. whay they are fighting, if some countri win what will happen after that, those questions wold I want to have ansers for. An other question I want to have anser for is whay, are allomost every countri speaking different laugish. That would be a great discussen I think.

I think this camp would be very inportent in the future, beacus at this camp do you learn something from other countris. And this could lead to world peas if the kids at the campe talk to the peoples in the countrie, and explen whay the other counti is fighting.

I think this campe should be very very inportent in the furture.

Sow this is my letter for you I hope that I meet you at campe connect if I get in.

to: intcamp@connect.ch

Bedömning: A

Dear Amanda and Josef

My name is N N and I am 15 years old. I come from a small town named X. X lies in the middle of Sweden and is a very nice town. My interests are sports, music, art and of course meeting new people. I want to participate in The International Youth Camp because I want to learn more about other countries and cultures, most of all I want to meet new people and represent Sweden. I think that it would be a great experience because i don't know how it is to live outside Sweden and I think that I would grow as a person.

Something I want to do is celebrate new traditions and learn more about different cultures. that's why I think that it would be a great idea that everyone gets to decide what we will do one day and show something from their country. When it's my turn I think that I would celebrate what we in Sweden call "Midsommar". We would dance "Little frogs" and eat typical Swedish food. One thing I want to discuss is why many kids in my age dont know anything about other cultures and how we can change that. I know that we need to make people aware of that there are many more cultures and way of life then in Sweden. I think that it is very important to unite more countries with eachother so that we can start making the world a better place togheter. I think that The International Youth Camp is a great idea and I believe that when I come home I am gonna be more open to other cultures and my friends will notice that, and they will do the same. I would keep in touch with my new friends and we could learn more about eachother over the internet or by letters. It wont change the world right away but if a few people goes to "IntCamp" every year we could change the way of living and then our children would have more respect from the start and their children even more. Of course we can change the world today but only if we get more people to understand and that is difficult. The best way I can think of is International Camps and thats why I want to a part of it.

I look forward to meeting you N

Bilaga 4 Exempel på bedömda elevtexter Engelska 6 (gymnasieskolan)

Bedömning: E

Innocent Fun or Dangerous Possibilities?

Your first cigarett, Your first drink, First time you gambled. We have all been throw that part of our lifetime. And even if you were strong enough to say No to this temptations You've been there, seen that. And the Drugs, that's a bit more unusual I think, but it's been there in the background, waiting for a weak youth to get it. And some of us have taken it, why? Yeah, why not may some of us think, But the sadest part of this temptation is that you get stuck in it, sooner or later, because drugs are just that way. You test it once because you're best friend told you that one time more or less isn't such an' big deal. And then one more time because you liked the Feeling. Suddenly you are stuck in the middle of the addiction. There you are, stuck in the middle of it and wondering: "What went wrong, how did I got into this mess?" You were a youth, couldn't say no, because it gave you a "cool" mark in your forehead, but how "cool" is it when you are growing up, may have miss a big part of your life, because of that thought "one time, more or less, why not"...

The main issue in todays society is that drugs, alcohol, cigarettes & models, they all over the world. In a bigger or a smaller piece. The peer pressure, "she's got that shirt for a 100 box, I must have it". "He has done that and that and he says its not a dangerous part to do", "it's a party tonight, sober? Your'e such a freak"!! That's the most common scentences you hear in todays society.

She take one drink and one cigarette, why cant you? And she take another Drink and another cigarette, and you do the same. Cool you think. Day after that you wake up wondering what you did, and you have not an' single clue what happened. How cool is that on a scale? Sure, following the crowd isnt always an' bad idea, depends on what the crowds are up to, what plans they have. Some youths' are really working to get good scales and then they have a lots of fun on the weekends. SOBER nights, no cigarettes, but they have really good time they to, like the other gangs who doesn't care which scales they got. Just talking about the weekend, How much they have to drink to get drunk and which people they are going to invite to get a cool party! Cool? Sure. If you think that school is for freaks and that you are expecting to get money from "soc" all of your lifetime, really really cool. Don't think that you will look back on your

VOKABULÄR OCH BEDÖMNING AV SKRIFTLIG FÖRMÅGA

youths with luck, when you one day wakes up on the hospital with an overdose heroin in your blood.

Sure, its easy to say so when you been throw that part and seeing “Friends” ending up like that. Its more difficult to see whats happening, or what would happen when youre in middle of the whole mess.

Think about this following words when something is about to get to a mess!

YOU ARE GOOD THE WAY YOU ARE!

(and no one are better than you to say NO to this)

Bedömning: A

Innocent Fun or Dangerous Possibilities.

A person once said “The best way of getting rid off your temptations is to follow them”. Think about that for a moment and imagine the world in front of you. What would it look like?

In today’s society young people are faced with all kinds of temptations; drinking, gambling, drugs etc. None of these examples are good or healthy in any way. So why do we still do these things? I would have to say like Roman historian Publicus Cornelius Tacitus once said “Things forbidden have a secret charm” and also the so common peer pressure plays an important role.

When you are young and uncertain of yourself it’s easier to be manipulated, and media knows this. A lot of advertisements often tells you to buy this or that and you will look just like him or her. And the fear of being alone and different from everyone often makes you buy things you don’t really want to. You could call this following the crowd. This is where your conscience is very important because it is your conscience that tells you wether it’s good or bad.

A typical example of this is smoking. Maybe you want to be part of this cool group of kids in school but they are all smoking. So you are thinking – to be one of them I also have to smoke. In this example following the crowd is bad, but there are situations where it can be usefull. I will get back to that later.

Dealing with your temptations is definitely a part of growing up. It is a way of testing your limits. This is where good role models are so important. I believe the lack of role models is one of the biggest causes to a lot of these problems. Good upbringing and good parents are very important. It is part of raising a child – telling him or her what to do and not to do. The parents can easily turn off the TV when something inappropriate is on, even if the children complain. When the children are getting older they need to understand what is good and what is bad, hopefully the parent already has taught the basics. Parents and role models I believe are the most important persons keeping young people away from these bad temptations such as smoking.

When you are older you maybe stop listening to your parents and your friends become more important. Maybe you then try to smoke and drink some beer, it's okay I believe and hopefully your upbringing has taught you something. As long as you can get yourself back again after you have let go. Often you learn from a new experience. Younger people can't deal with temptations in a responsible way.

However there are times when peer pressure can be useful. If you have a goal you want to reach really bad, a group or a crowd can support you reaching it. So you can't say that all peer pressure is bad and following a crowd can help too sometimes. It is how you follow it, use your own head and don't let the temptations lead you into bad peer pressure. My final words are: Start with being a good role model yourself and maybe some day everyone will be one as well. And No, the world would not look very good if every one followed their temptations.

Proven i praktiken – om nationella prov i engelska

Ulrika Arvidsson, Lena Asp, Jerker Brorsson

1 Inledning – syfte

Den andra halvan av vårterminen är i svenska grund- och gymnasieskolor en period präglad av nationella prov, bland annat i engelska. Proven upptar en stor del av lärares och elevers tid. Ett av provens syften är att ge oss lärare stöd för en likvärdig och rättvis bedömning.

Följande text, om hur de nationella proven i engelska planeras, genomförs och hur resultaten används, är författad av tre grundskole- och gymnasielärare med vardera 20–25 års erfarenhet av att undervisa i engelska. Utgångspunkten är våra personliga erfarenheter av nationella prov i engelska på högstadiet och gymnasiet. Vår text grundar sig på empiri snarare än teori. Vi är väl medvetna om att redogörelsen är subjektiv och att tillvägagångssätt och erfarenheter kan se annorlunda ut på andra skolor. Vår förhoppning är dock att redogörelsen ska väcka några tankar om och ge viss inblick i hur proven kan hanteras på en skola. Under de senaste åren har brister i hanteringen ute på skolor uppmärksammats då provuppgifter har spridits via sociala medier när handhavandet inte har varit korrekt. Vi ser därför att det finns ett behov av att belysa de olika momenten som ingår i planering, genomförande och efterarbete av proven. Det finns många likheter gällande genomförandet av nationella prov i engelska mellan stadierna men i de fall skillnader förekommer, betonar vi detta specifikt. Vår ambition är att i texten både generalisera utifrån våra erfarenheter men också ge mer konkreta exempel på hur vi har arbetat genom åren.

Målgruppen för texten är lärarstudenter, lärare eller personer som generellt är intresserad av att få en bild av hur arbetet med de nationella proven kan gå till på en skola. Dessutom ser vi att det finns ett behov bland skolledare att bli mer insatta i det omfattande arbetet med provens olika delar, varför vi också vänder oss till dessa. Skolledare med kunskaper om proven bör kunna säkerställa handhavande och organisation korrekt och effektivt.

2 Översikt av delprov

Obligatoriska nationella prov i engelska ges i grundskolan som ämnesprov för elever i årskurs 6 och 9. I gymnasiet finns kursprov i engelska 5 på alla gymnasieprogram och i engelska 6 för elever på alla högskoleförberedande program. Kursproven i engelska 5 och 6 ges både höst- och vårtermin. Skolan väljer det kursprov som ligger närmast slutet på kursen. Kursproven ges även i den kommunala vuxenutbildningen; detta kommer dock inte att beröras här.

De nationella proven i engelska är så kallade *proficiency tests* vilket innebär att proven har som syfte att pröva elevernas allmänna språkfärdighet i engelska oavsett när, var och hur eleverna har erövrat dessa färdigheter. Proven består av tre delprov som fokuserar på färdigheterna tala/samtala, läsa och lyssna samt skriva. Delprov A, den muntliga delen, genomförs sedan 2015 i slutet av höstterminen i engelska (samt svenska och matematik) för årskurs 6 och 9 och för gymnasiet under vårterminen fr.o.m. vecka 11. Delprov B, prövar elevernas receptiva förmåga i läs- och hörförståelse och genomförs på fasta provdagar under vårterminen. I delprov C, som prövar elevernas skrivförmåga, genomför eleverna en skrivuppgift i okänt ämne på en fast provdag.

3 Allmän förberedelse inför proven

När arbetslag och skollledning planerar läsårets kalendarium utgår man delvis från de dagar då de nationella proven äger rum. Särskilt i grundskolan kan några veckor på vårterminen vara svårorganiserade med tre eller fyra delprov per vecka i olika ämnen. För att skapa lediga dagar och frigöra lärare för genomförande och bedömning av proven lägger därför en del grundskolor en properiod för elever i andra årskurser än årskurs 6 och 9 under dessa veckor. Vidare bör en sammanställning göras över vilka elever som eventuellt är i behov av anpassningar vid provtillfället. Detta är något som den undervisande läraren ofta gör i samråd med skolans speciallärare eller specialpedagog och rektor.

Ytterligare en tidig förberedelse är beställning av proven, vilket är rektors ansvar. Ofta är det en fördel om även den undervisande läraren är inblandad eftersom rektor inte alltid är insatt i de behov som finns i klassen och vilket material som finns att beställa.

4 Delprov A – muntlig produktion och interaktion

Delprov A prövar elevernas förmåga att uttrycka sig och interagera i tal på engelska.

4.1 Delprov A – Förberedelse

När provmaterialet kommer till skolan är det dags för lärare att bekanta sig med det muntliga delprovet. Ett viktigt moment är att lyssna på de inspelningar av elevsamtal som bifogas. Detta görs med fördel inom engelsklärargruppen och det är ett bra tillfälle att träna bedömarkompetensen tillsammans. Det kan vara ett utmärkt tillfälle för gemensam fortbildning som inte kostar skolan mer än lärarnas konferenstid.

Här följer ett exempel på hur vi har jobbat på våra respektive skolor. Under konferenstid eller egen planeringstid, lyssnar vi till de bifogade exempelsamtalen utan att veta hur elevprestationen har betygsatts. Därefter görs individuell bedömning. Sedan diskuterar och resonerar vi kring förtjänster och brister i exempelsamtalen. Vidare läser vi de medföljande kommentarerna och bedömningarna och gör jämförelser med den egna diskussionen. Detta är en bra förberedelse och fortbildning som ger en god start på arbetet med det muntliga delprovet. Dessutom kalibreras vår bedömningsförmåga.

Det finns dock fortsatt många skolor som inte har möjlighet till denna typ av fortbildning då det ibland kan vara svårt att hitta gemensam tid. I så fall får förberedelser ske individuellt.

Även eleverna behöver förberedas inför det muntliga provet. Detta kan ske i form av undervisning där fokus ligger på muntlig produktion. Här kan lärare och elever tillsammans konkretisera vad bedömningsfaktorer och kunskapskrav i styrdokumentet innebär. I normalfallet är detta inte något nytt för eleverna utan något man även tidigare arbetat aktivt med. Under förberedelsen inför delprov A poängteras också vikten av att eleverna i samtalet ska hitta en balans mellan att ta initiativ och driva samtalet framåt, samtidigt som de ska vara inlyssnande och fånga trådar i partners inlägg. Dessutom lyfts vikten av att talutrymmet fördelas jämlikt under provtillfället.

Ytterligare en förberedelse inför det muntliga delprovet är att säkerställa inspelningsmöjlighet med tillräckligt god kvalitet. I takt med teknikens utveckling har detta blivit allt lättare och idag använder många sin lärardator eller

ATT BEDÖMA SPRÅKLIG KOMPETENS

annan digital enhet för inspelning. Eftersom muntliga övningar numera är vanligt förekommande i undervisningen är eleverna oftast relativt bekväma med att genomföra muntliga samtal samt att dessa spelas in.

Det muntliga provet ska enligt instruktionerna ske i par. Denna indelning är en grannlaga uppgift. Ofta är det bra att kombinera elever på samma språkliga nivå. Dessutom bör de känna sig trygga med sin partner. Därför är det lämpligt att tillsammans med eleverna resonera kring grundvalen för indelningen. Läraren kan även poängtera att det kan vara bra att inte känna sin partner ”alltför väl” då det kan ge goda förutsättningar för en naturlig konversation i provsituationen. Eleverna bör få möjlighet att meddela om det är någon de själva tycker sig vara mindre bekväma att samtala med. Denna information, i kombination med lärarens kunskaper om elevernas färdigheter och tidigare prestationer, blir underlag för indelningen.

I gymnasiet, till skillnad från grundskolan, ska läraren ungefär en vecka före provdagen ge eleverna en kort presentation av provets tema samt en kort ämnesintroduktion. Detta delges i lärarinformationen.

4.2 Delprov A – Genomförande

4.2.1 I grundskolan

Det muntliga delprovet genomförs under en sexveckors period med början i november. Varje samtal tar ca 15–25 minuter att genomföra. Dessutom tillkommer en stund för instruktioner samt tid för bedömning och nedskrivning av kommentarer efteråt. Vissa skolor låter undervisande lärare själv genomföra proven på lektionstid vilket medför att övriga elever arbetar självständigt med andra uppgifter.

Här följer ett exempel på hur vi har organiserat genomförandet. Under samtalen finns två lärare närvarande som gör bedömningen på plats. Detta möjliggörs då lärarna frigörs från undervisningen. Att samtalet bedöms av två lärare, en som undervisar klassen i ämnet och en ämneskollega som inte har gruppen, ökar bedömningens tillförlitlighet och stärker bedömningskompetensen. Tillsammans reflekterar vi kort efter samtalet över elevernas prestation i relation till kunskapskrav och exempelsamtal och elevprestationerna betygssätts preliminärt. Eftersom inspelning sker finns ytterligare möjligheter till kompletterande medbedömning i efterhand.

Medbedömning ökar tryggheten för både lärare och elever samt stärker bedömningens reliabilitet, men är inte alltid praktiskt genomförbar på alla skolor.

PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA

I många fall genomför läraren på egen hand alla samtal själv i sin undervisningsgrupp. Då är inspelning en förutsättning för medbedömning i efterhand.

Under samtalet kan vissa elever inledningsvis vara något spända och avvaktande men blir vanligen intresserade av uppgiften och diskuterar frågorna och påståendena aktivt. Enligt vår erfarenhet tycks de allra flesta elever känna sig nöjda med delprovet när det är avslutat och de upplever sig ha fått möjlighet att visa vad de kan och att det dessutom var roligt.

4.2.2 I gymnasiet

I gymnasiet genomförs delprov A en bit in på vårterminen. Det finns olika sätt att organisera genomförandet. Vissa skolor genomför det muntliga provet under lektionstid medan andra skolor använder sig av schemabrytande dagar och genomför delprovet koncentrerat under dessa dagar. Vid provtillfället får eleven 15 minuter att förbereda sig individuellt och därefter genomförs själva provet under ca 15 minuter.

Ett välfungerande scenario kan se ut på följande sätt: en resursperson, exempelvis en elevassistent, finns tillgänglig för att turvis ta emot elevpar och ge eleverna instruktioner, elevmaterial, anteckningspapper och penna. Resurspersonen ansvarar för att inga hjälpmedel används och att förberedelsen sker individuellt. Efter elevförberedelserna finns, i anslutande lokaler, två engelsklärare som löpande tar emot var sitt elevpar för genomförande av samtalet. Ett sådant ”rullande” upplägg innebär att vi hinner bedöma upp till 24 elevers samtal under ett för- eller eftermiddagspass. Detta tillvägagångssätt ger oss en mer renodlad bedömarroll, då vi inte behöver tänka på instruktioner och förberedelsetid till eleverna eftersom detta sköts av resurspersonen. I de fall det finns ytterligare engelsklärare tillgängliga kan förstås fler samtal ske parallellt. Bedömning av samtalet görs direkt på plats och vi använder noteringsunderlaget från lärarinformationen där bedömningsfaktorerna finns angivna. I mån av resurser försöker vi även att låta två bedömare närvara vid vissa samtal för att stärka reliabiliteten. Det är även en trygghet att spela in samtalen för att vid behov kunna gå tillbaka för att lyssna på delar igen och få medbedömning av kollegor.

En förutsättning för att kunna organisera på detta sätt är att arbetslaget tillsammans kan bryta upp schemat för att frigöra både lärare och elever. Tillvägagångssättet gör att antalet salar som tas i bruk även minimeras. Om läraren enskilt skulle genomföra provet hade minst fyra lokaler behövts – en för själva

provet, en för varje elev som förbereder sig individuellt och en för övriga klassen. Eftersom eleverna aldrig lämnas ensamma med materialet i vårt scenario ovan riskerar vi inte att eleverna samarbetar eller att materialet sprids.

Det förekommer att lärare låter elever genomföra det muntliga delprovet utan lärare, vilket läraranvisningarna säger att man endast i undantagsfall kan göra, med inspelning och därefter överlämnar den till lärare för bedömning. Dock kan det vara problematiskt att grunda sin bedömning enbart på en inspelning, utan att ha sett eleverna under samtalet då man missar gester, miner och kroppsspråk vilket är en del av kommunikationen. Vidare kan även delprovets sekretess äventyras när eleverna själva hanterar material och inspelade filer, varför tillvägagångssättet bör undvikas, om inte mycket speciella omständigheter föreligger.

5 Delprov B – receptiva förmågor

Delprov B prövar elevernas förmåga att förstå talad och skriven engelska.

5.1 Förberedelse

All undervisning som inriktas mot läs- och hörförståelse kan ses som en förberedelse inför delprovet. Det är dock bra att låta eleverna bekanta sig med olika uppgiftsformat som kan förekomma. Detta görs lämpligen med exempeluppgifter från NAFS hemsida²⁶ eller frisläppta prov. I undervisningen ingår också att resonera kring strategier som kan användas generellt i kommunikation inklusive provsituationen. Hur kan man tänka om man inte kan ett ord? Hur kan man anteckna på ett bra sätt? Hur kan man mentalt förbereda sig om man “tappar bort sig” i hörövningen?

På ett organisatoriskt plan måste tillgång till lämpliga salar och vaktande personal säkerställas inför provtillfället. Vad gäller salars lämplighet är det inte en självklarhet att den sal där undervisningen normalt sker är mest lämplig i en provsituation. Det kan handla om att eleverna sitter alltför tätt, att akustiken inte är lämplig för hörförståelse eller att ljudanläggningen inte är fullgod.

I god tid före provdagen är det också viktigt att planera provets genomförande för de elever som behöver anpassningar av olika slag. Detta är individuella anpassningar som undervisande lärare i samråd med kollegor, och i vissa fall med eleven, resonerar sig fram till. Det kan handla om att elever behöver

²⁶ Projektet Nationella prov i främmande språk, Göteborgs universitet. [NAFS](#)

PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA

sitta i mindre grupp eller ha utökad provtid, varför ytterligare personal och fler salar kan komma att behövas. Anpassningar kan också handla om att förstora text, kopiera på färgat papper eller att eleven muntligt ger svar, som skrivs ner av läraren. Denna typ av anpassningar får dock tidigast ske dagen före provet och därefter måste proven låsas in igen. Denna hantering av anpassningar kan äventyra sekretessen om inte reglerna följs fullt ut. I klasser utan behov av anpassningar öppnas provets försegling först på provdagens morgon. För elever med funktionsnedsättning finns ett standardiserat lyssningsstöd för den långa lästexten vilket måste beställas separat.

5.2 Delprov B – Genomförande

Delprov B, den receptiva delen, genomförs för både grundskola och gymnasium på liknande sätt. Provet ges samma datum i hela landet för respektive årskurs/kurs. Tidsåtgången för delprov B i årskurs 6 är 50 minuter för läsförståelsedelen och runt 45 minuter för hörförståelsedelen. I årskurs 9, engelska 5 och 6 är tidsåtgången 90 respektive 50 minuter. En rast på 20–30 minuter planeras också in mellan de två provdelarna.

Enligt vår erfarenhet brukar eleverna uppleva att de har tillräckligt med tid för att genomföra delprovet. Skolverket rekommenderar att alla elever stannar kvar i salen under hela provtiden. Elever kan exempelvis rekommenderas att ta med en bok att läsa i under resterande provtid. I vissa fall kan det dock vara nödvändigt att frångå detta för att ge arbetsro till de elever som ännu inte är färdiga.

Det är inte ovanligt att vaktande lärare under provet inte är engelsklärare. Det är då viktigt att engelsklärare, som är förtrogna med förutsättningarna och provmaterialet, har informerat övriga berörda om vad som gäller under provet och kan säkerställa att alla elever får samma instruktioner. Det kan vara lämpligt att försöka se till att någon med god kunskap om provets handhavande finns tillgänglig i början av provtillfället eller allra helst under hela provet. Detta så att man snabbt och enkelt kan lösa praktiska frågor och problem som, trots god förberedelse, kan uppkomma. Ytterligare en viktig aspekt är att säkerställa att elever inte har tillgång till exempelvis mobiltelefoner för att inte riskera spridning av provuppgifter eller fusk.

6 Delprov C – skriftlig produktion och interaktion

Delprov C prövar elevernas förmåga att uttrycka sig och interagera i skrift på engelska.

6.1 Förberedelser

Som förberedelse inför delprov C behandlas i undervisningen bedömningsfaktorererna för skriftlig produktion och interaktion och dessa konkretiseras. Det är något som sker kontinuerligt då lärare och elever tillsammans belyser kunskapskraven och aktivt arbetar med att exemplifiera värdeorden utifrån styrdokument och kommentarmaterial. Elever har, oavsett språklig nivå, arbetat med att skriva texter i olika genrer och med olika syften i undervisningen.

I vår undervisning brukar vi, inför delprov C, låta eleverna titta igenom sin skriftliga produktion och reflektera över den lärar- och kamratrespons de fått och vad de kan dra för lärdom av denna. Eleverna har fått träna på att ge respons på varandras texter utifrån kunskapskrav och exempeltexter från tidigare års frisläppta prov och andra elevproducerade texter. Syftet är att medvetandegöra dem om styrkor och förbättringsområden i sin egen skrift. Vi påminner eleverna om att noga läsa vad själva uppgiften faktiskt innebär – vilken typ av text de förväntas skriva samt att utnyttja den hjälp provmaterialet ger via t.ex. stödpunkter. Vidare påpekas att eleverna förväntas hålla sig inom ett visst ordspann.

Under de senaste åren har alltmer av elevernas skriftliga produktion och interaktion i undervisningen kommit att ske digitalt. Därför välkomnas förstås införandet av möjligheten att skriva delprov C digitalt både av elever och lärare. Det gör att provet blir mer autentiskt. Vidare möjliggörs viss textredigering för eleverna, vilket de är vana vid.

När delprovet genomförs digitalt får inte eleverna ha tillgång till stavnings- eller grammatikkontroll. Dessutom måste det säkerställas att elevernas digitala enheter inte kommunicerar med internet och på så vis äventyrar sekretessen eller möjliggör fusk. För att möta dessa krav kan skolor skapa provkonton som blockerar internetuppkoppling samt endast erbjuder ett enklare ordbehandlingsprogram. Skolor kan även köpa färdiga provplattformar. Andra sätt att tillgodose kraven är att ha ett flertal vakter i varje provrum vilket organisatoriskt kan vara svårt att hantera.

Frågor av mer teknisk och organisatorisk karaktär kan också behöva beaktas. Det kan exempelvis handla om trasiga elevdatorer och behov av extra enheter för utlåning. Man kan också behöva ha teknisk support vid provtillfället. Hur

PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA

elevs texter förs över till bedömande lärare måste hanteras på säkert sätt. Det är också viktigt att tänka igenom eventuella anpassningar för de elever som kan behöva det. Elever med funktionsnedsättning kan ges möjlighet att återge den text de producerat för läraren eller att få den uppläst för att kunna redigera.

För engelska 6 i gymnasiet finns ett förberedelsematerial som ska behandlas och bearbetas innan delprovet skrivs. Cirka två veckor innan det skriftliga provet delas förberedelsematerialet ut, vilket levererats tidigare under våren. Det kan ges som hemuppgift att till följande lektion läsa in sig på materialet. Eleverna kan stryka under och kommentera i häftet som ett stöd för kommande diskussioner. Därefter följer en förberedelselektion som ofta inleds med att låta eleverna diskutera sina tankar utifrån materialet i små grupper. Det brukar bli ett bra samtal kring olika texttyper och hur man ska förhålla sig till källor. Det varierar hur väl eleverna förbereder sig med hjälp av detta material, såväl innehållsmässigt som språkligt. Vissa elever ser potentialen att få stöd av förberedelsematerialet och utnyttjar även möjligheten att erövra vokabulär som de kan tänkas behöva, medan andra av olika skäl inte har läst igenom materialet särskilt grundligt. Efter lektionen lämnar eleverna in materialet, då detta inte ska användas under själva provtillfället.

6.2 Delprov C – Genomförande

Liksom delprov B ligger delprov C på ett fast datum och har varierande tidsomfattning beroende på årskurs; från 45 minuter i årskurs 6, till 80 minuter i årskurs 9 och engelska 5 samt för engelska 6, 100 minuter.

De allra flesta elever kommer snabbt igång med sin text och hittar någon ingång som passar den egna personen och förmågan. Varje år finns det några elever som skriver långa texter och upplever sig begränsade av ordspannet de förväntas hålla sig inom. I enstaka fall hinner de inte avsluta sina texter på det sätt de hade önskat.

7 Efterarbete

7.1 Bedömning

När så eleverna har gjort sitt är det dags för läraren att ta sig an bedömningen av de olika delproven. Att hitta det mest rättvisa, likvärdiga och praktiska sättet för bedömning av elevernas prestationer kan vara en svår uppgift.

ATT BEDÖMA SPRÅKLIG KOMPETENS

För delprov A handlar det om att komplettera anteckningarna man gjort under samtalets gång och med hjälp av kunskapskrav och bedömningsfaktorer fastställa elevens delprovsbetyg. Inspelningarna är till god hjälp när man behöver lyssna på vissa samtal ytterligare en gång eller när man behöver ta hjälp av en kollega för medbedömning i de fall då man varit ensam bedömare.

Delprov B är den provdel, vars bedömning är enklast och snabbast att genomföra, då upp emot hälften av uppgifterna är s.k. slutna format där enbart ett alternativ är rätt. Även öppna svar exemplifieras tydligt i bedömningsanvisningarna. Oftast bedömer respektive lärare proven från sin klass och de öppna svar som man anser svårbedömda tar man upp med kollegorna vid en ämnesträff och jämför med bedömningsanvisningarna. Dessa möten brukar ofta lösa de problem eller funderingar som eventuellt kan finnas och det är givande att tillsammans resonera kring olika varianter på elevlösningar.

Delprov C är det delprov som oftast kräver mest bedömningstid. Liksom inför bedömningen av delprov A är det viktigt att studera de bifogade elevexemplen från bedömningsmaterialet, kunskapskraven och kommentarmaterialet. Då man vid det här laget känner sina egna elever väl, kan det vara värdefullt att man som undervisande lärare inte ensam bedömer sina elevers texter. Detta för att undvika att påverkas av tidigare kunskaper om elevens skrivförmåga. Ett välfungerande sätt vi har använt oss av i gymnasiet är att tre lärare samarbetar och byter texter enligt ett cirkelschema. Vi gör individuellt en grundlig bedömning med motivering samt betyg och återlämnar därefter texterna till undervisande lärare. Denne gör i sin tur en snabb bedömning utan att ta del av förste bedömarens kommentarer innan den egna bedömningen görs. Undervisande lärare jämför sedan de två bedömningarna. Därefter har vi ett möte där vi lyfter de texter vars bedömning skiljer sig åt. Den tredje bedömarens får nu möjlighet att läsa denna text och vi diskuterar fram en gemensam bedömning. Undervisande lärare har dock det slutgiltiga beslutet.

Generellt är det bra att sträva efter att byta elevtexter med lärare på andra skolor eller program för att undvika att man vänjer sig vid en viss typ av elevtexter. Detta kan medföra att lärare skapar ett ”internt betygssystem” och att nivån sätts utifrån det elevunderlaget man möter på den egna skolan.

Eftersom bedömningen av delprov C tar mycket tid i anspråk hanteras det ur arbetstidshänseende på olika sätt. Vi har erfarenheter av grundskolor som använder sig av en schablontid för att räkna ut rättningstid av de olika proven. Det kan till exempel handla om 30 minuter per elev i engelska. Tyvärr täcker

PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA

inte denna tid in den faktiska tidsåtgång som krävs för all bedömning och rapportering av elevernas resultat. Om undervisande lärare har flera klasser som genomför nationella prov i engelska och i andra ämnen innebär detta en tung arbetsbelastning och i dessa fall kan proven fördelas ut för bedömning och rättning bland övriga behöriga lärare även om de inte undervisar klassen. Det kan dock se väldigt olika ut mellan skolor och stadier.

För att frigöra bedömningstid har vissa grundskolor s.k. ”rättningsdagar” när några undervisningsgrupper har praio-period eller annan schemabrytande aktivitet vilket medför att färre vikarier behövs. Det kan även vara positivt att involvera lärare från årskurs 4 och 5 i genomförande och rättning av några delprov för årskurs 6 vilket kan göra att de blir mer förtrogna med kunskapskraven. Detta kan ses som fortbildning både i ämnet och i bedömning.

Vi har erfarenheter från kommuner som använder sig av så kallad ”extern bedömning” där alla delprov för årskurs 6 och 9 samlas in och fördelas på några få rättande lärare. Det innebär att alla skriftliga delar av de nationella proven, bedöms av ett antal ämnesbehöriga lärare. Bedömningen äger rum under cirka fem arbetsdagar i en lokal som oftast inte ligger i anslutning till någon av skolorna för att lärarna i lugn och ro ska kunna fokusera på bedömningen. Konkret kan det gå till så att gruppen delas in så att hälften bedömer läsförståelse och den andra hälften bedömer hörförståelse. Vid delprov C, Writing, bedöms varje elevtext av två lärare som var för sig gör en bedömning. Är man helt överens fastslås betyget. Om bedömningarna skiljer sig åt marginellt resonerar man sig oftast fram till ett gemensamt beslut. I de fall där uppfattningarna skiljer sig ett helt betygssteg görs en tredje bedömning av ytterligare en lärare för att man ska bli ense om ett betyg. I enstaka, mycket svårbedömda fall kan omdömen från ytterligare en eller två bedömare komma i fråga.

7.2 Frånvarande elever

Om en elev är frånvarande vid ordinarie provtillfälle är det upp till rektor att besluta om eleven ska få lov att göra provet eller inte i efterhand. Detta är dock oftast något som sker i samråd med undervisande lärare som ofta ser en vinst med att låta eleven göra något delprov i efterhand för att bidra till en mer rättvis betygssättning.

En praktisk lösning kan vara att man mot terminens slut samlar ihop de elever som missat något delprov av alla de nationella prov som gjorts och genomför ett provtillfälle där eleverna har olika prov och provtider. Oftast är det då

årets prov som används vilket man bör ta hänsyn till när man beaktar elevens resultat, eftersom elever kan ha fått information om provets innehåll.

7.3 Elevrespons

När all bedömning är färdig får eleven information om sitt resultat från sina respektive lärare. Vanligen sker det under en engelsklektion där eleverna arbetar individuellt och en elev i taget får sitt resultat från läraren. De allra flesta elever är intresserade av att få se sitt prov och diskutera resultatet och hur de kan utvecklas vidare. Ytterst få elever väljer att inte delta. Då detta samtal oftast sker med enbart några veckor kvar av terminen, har läraren i de flesta fall hunnit göra en översyn av allt eleven visat av sina kunskaper och kan dessutom meddela slut-/kursbetyg.

7.4 Rapportering

De allra flesta skolor använder sig av någon typ av lärplattform där läraren lägger in elevens resultat på varje delprov tillsammans med en skriftlig kommentar. Flera system tillåter även att man kopplar resultatet till kunskapskraven. Via lärplattformen kan både vårdnadshavare och elev ta del av resultatet på de nationella proven. Vidare måste ansvarig ämneslärare lägga in resultatet för varje enskild elev på varje delprov för direktrapportering till Statistiska centralbyrån.

Därutöver behöver man som lärare stämma av vilka elevers prov som ska in till Göteborgs universitet för resultatuppföljning, kopiera dessa och skicka in. Avslutningsvis ska lärarenkäten fyllas i där vi som lärare får ge respons på årets provdelar.

8 Systematiskt kvalitetsarbete

Skolor arbetar aktivt med det systematiska kvalitetsarbetet och i detta ingår en analys av betyg och elevernas resultat på de nationella proven. Resultaten på de nationella proven kommer in som en naturlig och självklar del, då ett av huvudsyftena med proven är att ge underlag för att se i vilken mån kunskapskraven nås på huvudmannanivå.

Vi har erfarenheter från skolor där ansvarig ämneslärare först gör en reflektion kring elevernas resultat i sitt ämne och sedan tar med detta till läsårsutvärderingen. Till detta tillfälle sammanställer rektor skolans resultat på de nationella proven och skolans betygsstatistik för varje årskurs/kursbetyg. Undervisande

PROVEN I PRAKTIKEN – OM NATIONELLA PROV I ENGELSKA

lärare samlas i mindre grupper för att analysera underlaget. Utifrån dessa analyser kommer vi fram till de mål som skolan behöver arbeta extra mycket med framåt. Detta kan även vara lämpligt att analysera ämnesövergripande för att upptäcka tendenser som inte enbart är ämnesberoende. Dessa analyser kan även belysa behov av lärarfortbildning och utvecklingsområden som finns vilket är av stor vikt för skolledning och huvudmän i utvecklingsarbetet.

9 Vad diskuteras bland lärare?

Under läsårets gång är resonemang och diskussioner om de nationella proven ofta förekommande bland lärare. Nedan följer ett urval av denna typ av frågor.

En av läsårets första frågor som rektorer ofta tar upp i information till vårdnadshavare handlar om ledighet i samband med nationella prov. Här upplyser skolledningen om att inga ledigheter ges i samband med de datum proven ligger. Detta kan i vissa fall leda till diskussioner med vårdnadshavare. Av rent organisatoriska skäl samt likvärdighetsaspekter är dock många lärare nöjda med den restriktiva ledighetspolicyn.

Sedan höstterminen 2015 äger delprov A, den muntliga delen, rum i grundskolan redan i slutet av höstterminen medan övriga provdelar genomförs mot slutet av vårterminen. Många lärare upplever att en muntlig bedömning redan på hösten kan vara problematisk eftersom eleven har hela vårterminen på sig att utveckla sina muntliga färdigheter ytterligare och att elevens resultat på provdel A hade kunnat se annorlunda ut om provet genomförts på vårterminen. Denna specifika provdel skulle på så sätt kunna dra ner hela provbetyget. Därför är det viktigt att betygssättande lärare bedömer elevens samlade kunskaper vid terminsslut och inte enbart utgår från det nationella provbetyget. Detta understryker ytterligare att det sammanvägda provbetyget kan uppvisa diskrepans i relation till slutbetyget i årskurs 6 och 9, ett fenomen som givetvis också gäller gymnasieskolan.

Ytterligare ett område som diskuteras är bedömningsskalan i de nationella provdelarna i engelska gällande årskurs 9 samt engelska 5 och 6. Detta vållar stundtals diskussion, både att stegen D och B finns med, trots att de inte är verbaliserade i styrdokumentet, och att en ytterligare uppdelning görs. Medan vissa lärare uppskattar den större finmaskigheten i modellen, anser andra att den vållar problem.

ATT BEDÖMA SPRÅKLIG KOMPETENS

När provdelarna är bedömda ska en sammanvägning av resultaten från varje delprov ske till ett provbetyg. Denna sammanvägningsprocedur är kompensatorisk. Det innebär att en elev som fått ett F på ett av delproven ändå kan få ett E i provbetyg om de övriga provdelarna når upp till E. En elev som har goda receptiva förmågor kan i provbetygssammanhang låta dessa kompensera för andra mindre utvecklade förmågor. För vissa lärare kan detta tillvägagångssätt kännas motsägelsefullt i relation till betygssättning i slutet av terminen/kursen där samtliga kunskapskrav för ett visst betygssteg ska vara uppfyllda för att detta betyg ska kunna sättas. Det finns dock ingen sådan regel för betygssättningen på nationella prov.

De nationella ämnes- och kursproven utgår från kurs- och ämnesplanerna, som är uppbyggda kring syfte med förmågor, centralt innehåll och kunskapskrav, men proven täcker inte allt inom dessa styrdokument. De nationella proven ska utgöra en del av lärarens betygsunderlag. Vid betygssättning ska all tillgänglig information användas om elevens kunskaper i förhållande till de nationella kunskapskraven och det ska göras en allsidig bedömning av elevens kunskaper. Det kan med all rimlighet därför vara så att provbetyget skiljer sig från ämnesbetyget/kursbetyget.

I samband med de nationella proven i engelska önskar många lärare att få tid för sam- eller medbedömning av olika provdelar. Detta är något som är tidskrävande och måste planeras av skolledningen. Tyvärr kan det dock få stryka på foten då den gemensamma tiden för ämneslärare är begränsad och konkurrensen om tiden är stor. Exempelvis kan mycket tid tas i anspråk av aktiviteter av mer övergripande karaktär, såsom läslyftet och mattelyftet.

Den vikt som proven får, både från samhälle, skola och eleven själv, kan för några elever skapa en viss nervositet. Elever kan ställa frågan ”Vad händer om jag inte klarar nationella provet?” och de kan ha en tendens att skrämman upp varandra. Detta är något som lärare får bemöta och hantera regelbundet genom att avdramatisera provet för vissa elever medan andra elever behöver proven som motivationshöjande faktor. Eftersom läraren i undervisningen kontinuerligt arbetar med det proven mäter och förbereder eleverna genom olika moment bör en elevs resultat inte komma som en fullständig överraskning.

Det har på senare år förekommit diskussioner kring formativ undervisning och kritiska röster har hörts gällande de nationella provens summativa karaktär. Trots att provet i sin natur i första hand är summativt anser vi att provet kan bidra med formativ respons på hur lärarens undervisning har fungerat. Kanske märker läraren att resultaten på gruppnivå är sämre för hörförståelse vilket kan

PROVEN I PRAKTIKEN
– OM NATIONELLA PROV I ENGELSKA

ge en indikation på kommande utvecklingsområden för den enskilde läraren. Vidare kan den enskilde eleven bli medveten om styrkor och svagheter och ta med dessa insikter i sina vidare studier.

10 Avslutning

Sammanfattningsvis kan konstateras att de nationella proven i engelska är en inarbetad och naturlig del av vardagen i skolan. Vi upplever att de i allmänhet är väl mottagna och att lärare ser proven som ett stöd sin bedömning. Vidare anser vi och många med oss att proven är den tydligaste tillgängliga konkretisering av kurs- och ämnesplaner samt kunskapskrav. Detta ger oss i yrkesrollen en trygghet vid betygssättning. Vår erfarenhet, efter att ha arbetat i skolans värld i många år, är att såväl lärare som elever menar att provbetygen i relativt hög utsträckning stämmer med den tidigare bedömningen av elevernas kunskaper. Ofta finns det också överensstämmelse med elevernas egen självskattning. Således är proven, trots den tid som genomförande och bedömning tar, en uppskattad del i lärarens vardag.

ATT BEDÖMA SPRÅKLIG KOMPETENS

Språkprov på dator – vad händer då?

AnnaKarin Lindqvist, Ulrika Arvidsson, Sofia Nilsson

Nedanstående text behandlar utvecklingsmöjligheter och utmaningar vid datorbaserad av språkprov. Den inleds med en bakgrund och en belysning av olika validitetsaspekter i relation till aktuell forskning och tar sedan upp två exempel – prövning av läsförståelse respektive skriftlig produktion och interaktion som pappers- eller datorbaserade prov.

Bakgrund

De nationella proven kommer att datorbaseras och genomföras helt eller delvis digitalt inom en relativt snar framtid och ett omfattande förberedelsearbete pågår, både inom Skolverket och på de olika lärosäten som utvecklar de nationella provmaterialen.

Vid ett regeringssammanträde den 1 april 2015 beslutades om en *Översyn av de nationella proven för grund- och gymnasieskolan* (Kommittédirektiv 2015:36). Denna översyn behandlar det nationella provsystemet i Sverige i stort och en planerad digitalisering som en del av detta. Direktiven resulterade i en enmansutredning: *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning* (SOU 2016:25). Utredningen föreslår i korthet att en digitalisering av de nationella proven ska påbörjas och ske stegvis, och att en försöksverksamhet med datorbaserade prov ska inledas 2018 med målsättningen att nya digitala prov kan börja användas i full skala 2022. Det slutgiltiga uppdraget från utbildningsdepartementet till Statens Skolverk kom under hösten 2017 och beaktar utredningens förslag.

Inom projektet NAFS, Nationella prov i främmande språk, vid Göteborgs universitet, har det sedan 2009 kontinuerligt genomförts utprövningar av och studier kring möjliga lösningar för datorbaserad av proven i engelska, flera av dessa i samarbete med Skolverket (t.ex. Skolverket, 2010; Göteborgs universitet, 2014). Det har bl.a. handlat om rent tekniska utprövningar av skolors förutsättningar att genomföra språkprov på dator men framför allt analyserande studier kring huruvida resultaten på datorprov speglar provtagarens förmåga att förstå och producera målspråket, och i vilken utsträckning datorprov, avsiktligt

eller oavsiktligt, också prövar något annat – förmågan att använda datorn på det avsedda sättet.

Inledning

Alla prov och bedömningsmaterial består av flera komponenter som samverkar i det som prövas. Ett prov har vanligtvis som främsta fokus att ge möjlighet för provtagaren att visa en viss förmåga eller kunskap, men sättet detta ska visas på blir oundvikligen också något som prövas, som exempelvis förmågan att hantera den angivna tiden samt de anvisade verktygen och eventuella hjälpmedel. Datoranvändning utgör i sig ett sådant verktyg och är en del av bedömningens *format*, dvs. det sätt på vilket eleven förväntas visa sin kompetens. Format, instruktioner och bedömningsanvisningar utgör de tre viktigaste komponenterna i det som ofta benämns bedömningens *rubric* (Chapelle & Douglas, 2006).

För olika typer av datorbaserade provformat förekommer en mängd olika termer och som ett samlande begrepp används generellt *Computer-Based Tests – CBT*. Termen CBT används också specifikt som beteckning för datorbaserade linjära prov då samtliga provtagare gör identiska prov, vilket är den variant som är aktuell för de nationella provmaterialen i dagsläget²⁷. CBT kan rättas och bedömas antingen manuellt eller maskinellt, beroende på mjukvarans kapacitet och uppgifternas format. Vid jämförelser benämns prov som görs med papper och penna ofta *PBT – Paper-Based Tests* (ETS, 2011).

Att numera använda dator vid språkliga aktiviteter som inbegriper att läsa och skriva är naturligt och självklart i många sammanhang. Vissa saker blir effektivare, som till exempel att kunna redigera i en text när man skriver eller att kunna navigera och söka digitalt i en stor textmängd vid läsning.

Beträffande prov är det viktigt att veta vad provet avser pröva och om detta är vad som faktiskt prövas. Eftersom den här aktuella kontexten är nationella prov i *engelska*, såsom mål och kriterier för ämnet beskrivs i 2011 års kurs- och ämnesplaner (Skolverket, 2011a, 2011b), är en viktig fråga i vilken mån datoranvändning och digitala format är relevant eller eventuellt kan störa elevens möjlighet att visa (eller möjligheten att pröva) kunskaper i just *engelska*. Denna och liknande frågor har med ett provs eller en bedömnings validitet att göra.

²⁷ Ett annat datorbaserat provformat benämns CAT (*Computer Adaptive Testing*). Dessa prov har ofta ett gemensamt startblock och anpassar sedan svårighetsgraden efter provtagarens prestation.

Olika sätt att se på CBT

En fråga som betonas vid CBT är i vilken mån det som faktiskt är föremål för bedömningen, det så kallade *kunskapsobjektet* eller *the construct*, förändras som en konsekvens av datoranvändningen och om detta i så fall är önskvärt eller ett hot mot bedömningens validitet. Att *rubric*, dvs. tillvägagångssättet, kan ha betydelse för bedömningens validitet och resultat benämns ofta metodeffekt. Bachman (1990) menar att:

[...] the specific operations we use for making the construct observable reflect both our theoretical definition of the construct and what we believe to be the context of language use. [...] Performance on language tests thus varies as a function both of an individual's language ability and of the characteristics of the test method. (s. 113)

Sedan 1980-talet har CBT diskuterats frekvent inom bedömningsforskningen i språk. Chappelle & Douglas (2006) identifierar tre olika förhållningssätt till teknologins roll vid språkprovning, vilka de menar har stor betydelse för utvecklingen i relation till CBT. Det första antagandet karakteriseras av en syn på teknifiering som en ren effektivisering. Det gäller till exempel användning av stordatorer för distribution och automatisk rättning, varvid utgångspunkten blir att på ett snabbare, lättare och effektivare sätt 'göra samma'. Det andra antagandet baseras på ett jämförande och problematiserande perspektiv som utgår från att pappersprov är det 'normala' och det tredje, ett mer innovativt förhållningssätt till datoranvändning, ser tekniken som en naturlig del av ett förändrat kunskapsobjekt för språklig kompetens.

När det gäller en övergång för nationella prov från PBT till CBT är ett jämförande perspektiv nödvändigt, med tanke på till exempel likvärdighets- och stabilitetsaspekter. Douglas (2010) skriver:

The question for language test developers and users is not *Should these techniques be used for language testing?* They are being used and will continue to be. Rather, the questions we need to ask include these:

- What are the effects of the use of technology in language assessment on test-taker attitudes such as anxiety and motivation?
- How does language performance differ with different technologies?
- How does the use of technology affect the definition of the language ability construct we are attempting to measure?

- How does technology affect the nature of assessment tasks that can be developed?
- What are the limits of technology for scoring constructed response assessment tasks and how does automated scoring differ from human scoring? (s. 116–117)

Validitetsaspekter avseende CBT i språk

Avgörande för ett provs eller en bedömnings kvalitet är dess validitet och reliabilitet. Bedömningens *validitet* beror av hur 'rätt' den görs, att det bedömningsmaterial man använder faktiskt prövar och skapar underlag för bedömning av det man avser bedöma (*the construct*) och att bedömningen används så att den får rimliga och relevanta konsekvenser (jfr Messick, 1989; *consequential validity*). Bachman & Palmer (1996 bearb. 2010) har föreslagit sex kriterier för utvärdering av en bedömnings kvalitet, lämplighet och användbarhet, nämligen: reliabilitet, begreppsvaliditet, autenticitet, interaktivitet samt påverkan och praktikalitet. Underrubrikerna nedan relaterar till dessa kriterier.

Reliabilitet

Reliabilitet, eller tillförlitlighet, handlar om hur 'rättvis' och icke slumpmässig bedömningen är avseende till exempel jämförbarhet och likvärdighet (Erickson, 2009). Bedömningens reliabilitet benämns stundtals *scoring validity* och betraktas alltmer som en delaspekt av dess validitet, eftersom god validitet förutsätter att resultaten är reliabla och konsistenta (t.ex. Weir, 2005). Chapelle & Douglas (2006) betonar vinster vid CBT i form av enhetlighet, opartiskhet och därmed likvärdighet. Det faktum att instruktioner, eventuella hjälpfunktioner, process, tid och eventuell rättning (dvs. *rubric*) styrs via dator som en oskiljaktig del av provet innebär att mycket av den variation som kan omgärda PBT undviks, vilket kan medföra högre reliabilitet. Chapelle & Douglas menar också att exempelmaterial med syfte att göra provtagare bekanta med formatet lättare kan göras enhetliga, relevanta och tillgängliga digitalt, vilket därmed kan öka likvärdigheten.

När det gäller maskinell rättning av CBT betonas ofta vinster i form av exakthet, objektivitet och effektivitet. Historiskt sett förutsatte maskinell rättning enbart flervalformat, men de tekniska möjligheterna har blivit något bredare och viss bedömning av öppna format har utvecklats, till exempel när det gäller skriftlig och muntlig produktion. Trots detta kan begränsningar i svarsformat och datorns kapacitet att rätta vara ett konservativt element i CBT som

innebär att överväganden kring effektivitet och mätteknik tillåts överskugga eller begränsa andra sätt att bedöma och pröva kommunikativ språklig kompetens (t.ex. Douglas, 2000). Detta skulle kunna leda till att man frestas göra det lätt mätbara till det viktiga (Erickson, 2006) och att provets kunskapsobjekt eventuellt snävas in så att validiteten hotas av så kallad *construct under-representation*, vilket generellt är en fara vid bedömning (Messick, 1989).

Begreppsvaliditet

Synen på begreppsvaliditet relaterar till synen på CBT och dess eventuella metodeffekter. I mycket stora drag handlar detta antingen om att se (all) datoranvändning som en naturlig och nödvändig del av språklig kompetens, eller om att se (vissa) metodeffekter som störande och irrelevanta för kunskapsobjektet och det som avses prövas, så kallad *construct-irrelevant variance* (Messick, 1989). Diskussionen speglar förmodligen också en viss kronologi i den utveckling som skett under 2000-talet. Ett första, relativt obekymrat, antagande är att CBT mer effektivt prövar samma kunskapsobjekt/-begrepp (*construct*) som motsvarande pappersprov, dvs. att datorbaserad egentligen inte har, eller ska ha, någon betydelse för bedömningens validitet. Detta problematiseras emellertid alltmer i takt med utbredningen av CBT.

Chapelle & Douglas (2006) ser möjligheter att med CBT pröva och synliggöra nya språkliga förmågor. De diskuterar i termer av olika kunskapsobjekt – *the construct of interest* respektive *the test construct*, och betonar vikten av att klargöra vad som avses prövas respektive faktiskt prövas i ett aktuellt prov. Om provtagaren vid hörförståelse till exempel kan gå tillbaka och lyssna flera gånger eller vid skriftlig produktion kan använda online-hjälp eller språkkontroller, mäter provet implicit även dessa strategiska förmågor, vilket de menar innebär att dessa måste betraktas som del av provets kunskapsobjekt.

Enligt Chapelle & Douglas och senare Chapelle (2010) är ett paradigm som förutsätter att språkprov ska mäta språkliga kompetenser oberoende av kontext för språkanvändningen förmodligen orimligt. De menar att en förutsättning, för att kunna se både möjligheter och nackdelar med metodeffekter, är att begreppet validitet omvärderas för CBT och att andra avsiktliga dimensioner än korrelationen med traditionella pappersprov belyses. De betonar emellertid att en validering av CBT måste baseras på överväganden av i vilken utsträckning datorkompetens, eller datoranvändningen i sig, faktiskt påverkar utfallet. Exempel på hot mot validiteten skulle vara irrelevanta eller oavsiktliga metodeffekter

när det gäller annorlunda provprestationer, som orsakas av oro eller ovana vid formatet, ensidighet i uppgiftsurval eller uppgiftstyper, eller otillräcklig bedömning vid automatisk rättning, till exempel vid öppna svarsformat.

När det gäller exempelvis prövning av läsförståelse på dator diskuteras om läsning på dator innebär att det traditionella kunskapsobjektet breddas och i vilken mån detta i så fall är för begränsat. Chappelle & Douglas (2006) menar att kunskapsobjektet för språklig kompetens i till exempel en akademisk miljö inte kan särskiljas från förmågan att använda dator: “communicative language ability needs to be conceived in view of the joint role that language and technology play in the process of communication” (s. 107–108). Detta skulle då innebära en delning av begreppet språklig kompetens i olika kontextberoende delkompetenser, vilka i några situationer är förbundna med datorer. En liknande syn på språklig kompetens återfinns i OECDs skrivningar om literacitet (*Reading Framework*) avseende PISA (OECD, 2009, 2013). I denna definition av läsning inbegrips läsning av såväl handskriven, tryckt som elektronisk text. Det senare innefattar också olika strategier kring att till exempel skrolla och hantera hyperlänkar, med motiveringen att detta är en viktig del av läsförmåga inom flera språkliga domäner. Från 2009 prövas både traditionell och digital läsning i PISA och 2015 prövades även den traditionella läsningen via dator.

Praktikalitet och påverkan

I vilken mån ett prov är användarvänligt och möjligt att tillämpa i praktiken benämns dess praktikalitet. Bachman & Palmer (1996, 2010) definierar en testmetods praktikalitet som skillnaden mellan tillgängliga resurser och krävda resurser, och syftar då på såväl mänskliga och materiella resurser som resurs i form av tid. Litteraturen kring CBT påpekar generellt att CBT kan innebära vissa säkerhetsrisker, jämfört med pappersprov, och att implementeringsfasen medför ett ökat behov av support. Man understryker vikten av att gå stegvis fram och undersöka aspekter som kostnader, påverkans effekter på inlärning och även familiaritet med och oro inför CBT. Meijer (2009) påpekar att nationell bedömning är del av pedagogiska och socio-politiska system, som genom att mäta vad som pågår i dessa system oundvikligen förändrar desamma. Han betonar att nationella bedömningsinstrument upplevs som markörer dels för vad som är viktigt att lära, dels för vad som utgör kompetens inom ett område (så kallad *feedback effect*).

Autenticitet, interaktivitet och motivation

Autenticitet och interaktivitet betraktas som en förutsättning för meningsfullhet och generaliserbarhet, med betydelse för bedömningens validitet. Douglas (2000, 2010) och Chapelle & Douglas (2006) återkommer ofta till datorns kapacitet att mediera 'äkta' interaktivitet på målspråket samt att simulera verklighetslika kontexter. De varnar dock samtidigt för övertro på autenticitet och påminner om att mycket som har med språklig kompetens att göra inte alls innefattar datoranvändning.

Om bedömningen skall kunna stödja lärande och uppfattas som autentisk, relevant och meningsfull, är användarens attityd till och upplevelse av provningen viktig. Uttryck för ansträngning och tillfredsställelse samt för oro lyfts fram av exempelvis Dörnyei (2001) som betydande för såväl inre som yttre motivation. I samband med likvärdighetsperspektiv och praktikalitet nämns också att provtagarnas kapacitetsuppfattning och förväntan kan upplevas påverka provets genomförbarhet och de egna möjligheterna.

Att aktiva lärare och inte minst elever, är ovärderliga samarbetspartners i kvalitetsarbetet när det gäller utveckling av provmaterial är väl belagt i litteraturen (Bachman & Palmer, 2010; Dörnyei, 2001; Erickson, 1998, 2006; Erickson & Gustafsson, 2005) och en hörnsten för arbetet inom NAFS-projektet. Elever och lärare ger sedan mitten av 1990-talet respons på enskilda uppgifter avseende bland annat uppfattad svårighetsgrad och relevans samt egen prestationsupplevelse. Under de studier som gjorts inom projektet NAFS avseende CBT på senare år har en stor mängd sådana synpunkter samlats in och det övergripande intrycket är att elever generellt är mycket positiva utifrån en mängd motivations- och nyttoaspekter, men att de också har viktiga synpunkter av mer kritisk art. Nedanstående är smakprov från en elevenkät från grundskolan (Lindqvist, 2012):

– Jag tycker att det var as bra! Och jag ser fram till det i framtiden

– Det kan bli tröttsamt att sitta med penna och stirra på ett papper. Jag tycker defenitvt att vi ska ha prov på datorn.

– Jag tycker att det är lite mer intressant att göra proven i datorn och det är lättare. Men i så fall borde skolan byta datorskärmarna.

– Efter ett tag blev det lite jobbigt att titta på skärmen. Det blev lite flimrigt. På papper blir det inte det☺

– Detta va ett bra sätt men allt krångla mycket så allt skulle ha gått mycket snabbare med papper och penna.

Att pröva läsförståelse i engelska på dator eller på papper

Att läsa olika typer av text på någon form av skärm är numera en del av vardagen för många. Även i skolans värld introduceras medvetet och kontinuerligt läsning i digitala format. När det gäller att pröva läsförståelse via dator bör emellertid ett flertal aspekter studeras och övervägas.

De nationella proven i engelska består i dagsläget av tre delprov som prövar muntlig och skriftlig produktion och interaktion samt reception, dvs. läs- och hörförståelse. Det receptiva delprovet innehåller tre eller fyra läsförståelseuppgifter baserade på både så kallade kontinuerliga och icke-kontinuerliga texttyper. De ingående texttyperna kräver olika slags läsning och prövar i viss mån olika typer av lässtrategier. Bredd, variation och autenticitet eftersträvas i utformningen av uppgifterna så att så många elever som möjligt skall ges så stor chans som möjligt att visa vad de kan.

Delprovet innehåller ett flertal olika svarsformat med såväl öppna som slutna svar. Det finns generellt en längre narrativ text som kräver ett djupare, kontinuerligt läsande för att tillägna sig innehåll, se sammanhang och dra slutsatser samt flera kortare läsförståelseuppgifter. De kortare läsförståelseuppgifterna kan t.ex. vara texter av informations-, nyhets- eller populärvetenskaplig karaktär. Slutna svarsformat förekommer i samband med till exempel flervalsfrågor eller matchning, dvs. översiktsläsning av kortare texter som kombineras med passande fras.

Nulägesbeskrivning

De nationella proven i engelska avser att pröva traditionell läsning. God läsförståelse kan översiktligt beskrivas som förmågan att innehållsligt tillgodogöra sig texter av många olika slag (Skolverket, 2014a) samt förutsätter strategier för att anpassa läsningen efter textens syfte, innehåll, form och kontext. Förkunskaper, såväl språkliga som innehållsmässiga, i form av t.ex. ordförråd, kunskaper i andra språk eller allmän omvärldskunskap, kan bidra till förståelsen.

Digital läsförståelse anses bygga på traditionell läsförståelse och vara ”sammansatt av olika förmågor och strategier för att tillägna sig datormedierad information” (Skolverket, 2014a, s. 3). Digital läsning skiljer sig från traditionell läsning

t.ex. genom att den oftast är icke-narrativ och mer kontextbunden med fokus på informationssökning samt oftast icke linjär, icke-sekventiell och varierad beroende på webbsidans uppbyggnad. Texten består generellt av flera lager som läsaren kan klicka sig vidare i och läses ofta mer fragmentariskt (Skolverket, 2014a).

Aspekter kopplade till digital läsning lyfts fram i kurs- och ämnesplanerna för engelska på alla nivåer i form av skrivningar kring att söka, värdera, välja och tillägna sig innehåll från olika källor (Skolverket, 2011a, 2011b), men digitala sök- och läskompetenser prövas inte i de nationella proven.

Empirisk undersökning: en utprovning av läsförståelse i CBT-format

Datorbaserade utprovningar av uppgifter i engelsk läsning genomfördes inom projektet NAFS första gången 2009, med uppgifter som lånats från de norska nationella proven i engelska (Skolverket, 2010; Lindqvist, 2012). Sedan 2015 har utprovningar av uppgifter i CBT-format som avser pröva läsförståelse återupptagits, denna gång med fokus på längre narrativa textunderlag. Inom ramen för dessa utprovningar har elever i grund- och gymnasieskolan genomfört datorbaserade versioner av sådana uppgiftstyper.

I den här aktuella studien, som genomfördes under läsåret 2016–2017, fick eleverna läsa en längre narrativ text uppdelad i tre eller fyra avsnitt, och besvara frågor i såväl öppna som slutna format. De prövade olika grafiska utformningar och tekniska lösningar. Utöver den långa lästexten utprovades andra vanligt förekommande uppgiftstyper i de nationella proven. Skolorna som ingick i utprovningen använde olika typer av teknisk utrustning (datorer, 'paddor' etc.) och olika stora skärmar. Eleverna hade också delvis olika erfarenhet av arbete med digitala verktyg i klassrummet. I direkt anslutning till genomförandet besvarade eleverna en enkät kring bl.a. hur läsförståelsen fungerade rent tekniskt, om upplevelsen av att göra digitala prov och om sin egen digitala vana.

Analyser av utfallet visar att tendensen är att eleverna ställer sig positiva till digitala prov i läsförståelse. Gymnasieeleverna är än mer positiva än eleverna i årskurs 9. Några gymnasieelever kommenterar upplevelsen så här:

Gillar konceptet mycket!!!!
Snälla lös så alla prov blir på datorer. Ärligt:)

ATT BEDÖMA SPRÅKLIG KOMPETENS

Elever med god tillgång till och vana vid att använda digitala hjälpmedel i skol-sammanhang tycks mer positiva till digitala prov. Exempel på orsaker som anges är att det är lättare att skriva och ändra i sina svar, och att det går snabbare:

För mig brukar det gå åt mycket tid för att skriva och sudda när man skriver för hand. Jag får ofta skrivkramp och sitter och spänner mig i hela kroppen. Med datorn kunde man slappna av mer och det var enklare och skriva och ändra sina svar.

(elev, gymnasiet)

De som är negativa till digitala prov menar t.ex. att det är jobbigt att läsa på skärm, att det känns säkrare att ha det på papper ”så att man kan vara säker på vad man lämnar in” men också att ”om man har en penna och papper kan man stryka under i texten och enklare gå tillbaka för att kontrollera svar”. Dessa elevsynpunkter stämmer väl överens med vad som framkom i de första studier-na 2009 (Lindqvist, 2012).

Även om det finns en generellt positiv inställning till att göra uppgifter av den typ som ingått i studien digitalt, så är gymnasieeleverna minst positiva till just den långa lästexten. Format av lucktexttyp och korta texter med flervals-frågor tas emot mer entydigt positivt.

Fler utprovningar av både teknisk och innehållsmässig karaktär kommer att ske de närmsta åren. Det är i dagsläget rimligt att anta att det är en större utma-ning att utforma uppgifter, som på ett lämpligt sätt prövar traditionell läsför-ståelse i datorbaserad form på högre nivåer, där längre och mer komplexa texter används som underlag.

Att gå från papper till dator – vad händer då?

Projektet NAFS har ett pågående uppdrag att påbörja utveckling av dator-baserade prov (CBT) som mäter traditionell läsförståelse. Avsikten initialt är att så långt som möjligt pröva samma förmågor som i pappersbaserade prov (PBT). Det finns dock anledning att anta att själva överförandet till dator innebär en viss förändring av kunskapsobjektet. Om eleven, för att visa sin läsförståelse, behöver kunna hantera en dator, t.ex. skrolla och klicka sig fram, och detta i sig inte utgör en del av kunskapsobjektet finns risk för metodeffekter som bör kontrolleras för att inte påverka validiteten negativt. På samma sätt kan skrivande för hand också skapa metodeffekter i ett provsammanhang, när provtagare sällan skriver för hand annars. Chappelle & Douglas (2006) diskuterar om språklig kompetens i en akademisk miljö kan särskiljas från förmågan att använda dator och menar att

traditionella pappersprov istället skulle kunna sägas innebära negativa metod-effekter och vara ett exempel på ett alltför insnävat kunskapsobjekt.

Olika texttyper och olika svarsformat i de pappersbaserade nationella proven lämpar sig förmodligen mer eller mindre väl för CBT. Inom projektet NAFS pågår ett arbete med att analysera existerande format, om och hur kunskapsobjektet förändras vid en övergång till CBT samt i vilken mån det finns risk för oavsiktliga metodeffekter och ett alltför insnävat (eller utökat) kunskapsobjekt. Vidare studeras vilka nya möjligheter och utmaningar som uppstår vid utformning av datorbaserade läsförståelseprov. I sammanhanget kan noteras att Skolverket i en rapport, utifrån den svenska läsförståelsen i PISA 2009, konstaterar att korrelationen där är stark mellan resultaten på de digitala och traditionella läsförståelsetesten (Skolverket, 2013).

Det finns flera existerande format som troligtvis gynnas av en datorbaserad. Detta gäller t.ex. så kallade lucktexter där flervalsoalternativen presenteras i form av en 'rullgardin', vilket möjliggör för eleven att se det valda alternativet i sin kontext på ett mer autentiskt sätt. I frågeformat där det handlar om att hitta och identifiera ett ord eller en passage i den lästa texten, skulle eleven i ett datorbaserat format kunna markera ett avsnitt av texten för att avge ett svar, utan att producera egen text.

Med en datorbaserad av läsförståelsedelen i de nationella proven finns det troligtvis också möjlighet att utveckla nya frågeformat som underlättar provningen av vissa förmågor. *Drag and drop*-funktioner skulle exempelvis ytterligare möjliggöra provning av förmågan att förstå struktur och sammanhang i en text. Förståelsen av instruktioner och beskrivningar skulle kunna provas med hjälp av klickbara bilder.

Det finns också utmaningar med att övergå till att pröva läsförståelse via dator. Tekniska förutsättningar, som t.ex. skärmstorlek, skiljer sig mellan olika skolor. De prov som utvecklas måste kunna fungera för en stor variation av digitala lösningar och skärmstorlekar. Den långa läsförståelseuppgift som ingår i dagens prov ryms inte inom de ramar som vanligt förekommande provplattformar erbjuder. Det är en utmaning att ge eleven möjlighet att, som i de pappersbaserade proven, se frågor och text samtidigt. Likaså ryms inte mer omfattande lucktexter, utan dessa behöver delas upp i flera separata avsnitt. Uppgifter av detta slag avser att pröva bl.a. förmåga att förstå sammanhang och att språkligt binda ihop innehåll, vilket kräver att eleven har tillgång till hela texten och kan gå fram och tillbaka. Itemiserade provformat bör däremot fungera väl.

ATT BEDÖMA SPRÅKLIG KOMPETENS

De begränsningar som idag upplevs vid utformningen av datorbaserade prov, t.ex. möjligheten att få plats med både text och frågor, och arbeta med estetiskt tilltalande layout, kan förhoppningsvis delvis överkommas med teknisk och designmässig utveckling. Provplattformens utformning och möjligheter, liksom provtagarens digitala verktyg vid provtillfället, kommer dock att utgöra en begränsning. På samma sätt är naturligtvis dagens prov i sin utformning i viss mån styrda av A4-formatet, vars begränsningar vi oftast inte ifrågasätter.

Hur skulle då bedömningen av ett datorbaserat prov i engelsk läsförståelse inom det nationella provsystemet kunna se ut? De frågor som redan idag är lättträttade, dvs. flervalfrågor och frågor där enstaka ord efterfrågas, kommer i stor utsträckning att kunna automaträttas. Korta svar om ett eller några få ord kommer delvis att kunna automaträttas. De flesta provplattformer ger möjlighet att lägga in ett visst antal acceptabla svar och utifrån de storskaliga utprovningar som föregår ett nationellt prov finns goda möjligheter att på förhand lägga in vanligt förekommande alternativa svar, och i viss mån även accepterade felstavningar. Bedömande lärare får sedan gå igenom, och utifrån bedömningsanvisningar ta ställning till, de elevsvar som systemet inte kan hantera automatiskt.

Det kan finnas skäl till att ännu inte fullt ut överge pappersformatet och en kombination av papper och datorbaserad utformning kan vara ett alternativ. Texter skulle kunna distribueras i pappersform medan frågor och svar datorbaseras. Det skulle möjliggöra fortsatt användning av de längre lästexterna som används idag. Än så länge ger förmodligen pappersbaserade prov bäst möjlighet att använda sig av lässtrategier som att stryka under, ringa in och på andra sätt markera nyckelord eller väsentlig information.

Det finns andra skäl till att faktiskt överge papperet, men det bör göras med medvetenhet om förändringens positiva och negativa effekter. Wästlund (2007) lyfter fram att läsning på skärm tycks innebära en större ansträngning, *mental workload*, än läsning på papper och drar slutsatsen att:

[...] more importantly, in order to optimize the information processing ability during consumption of information of onscreen presented material, it is time to abandon the high-tech recycling of the ancient scroll and to break free of the traditions of printing, i.e. displaying written material in the standard A4 format and instead adapt the page layout to fit its intended medium. (s. 35)

Slutligen kan konstateras att prövning av läsförståelse på papper om/när läsning på skärm blir det mest vedertagna, troligtvis skulle inverka negativt på provets *face validity*, dvs. i vilken grad provet subjektivt uppfattas mäta det som det avser mäta, i det här fallet läsförståelse.

Att pröva skriftlig produktion och interaktion i engelska på dator eller på papper

Att använda datorn för skrivande är något som idag introduceras tidigt i en elevs skolvardag. Det kan därför tyckas otidsenligt att delprovet i skriftlig produktion och interaktion, *Focus: Writing*, i relativt hög utsträckning fortfarande skrivs med penna och papper. I övergångsprocessen till att genomföra provet på dator finns dock ett flertal aspekter som bör beaktas. Nedan följer en nulägesbeskrivning samt en sammanfattning av en studie som gjorts inom projektet NAFS (Göteborgs universitet, 2014) och till vilken det även refereras i tidigare nämnda SOU 2016:25, *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. Därefter följer ett resonemang kring vilka konsekvenser och eventuella metodeffekter en datorbaserad av det skriftliga delprovet kan medföra.

Nulägesbeskrivning

Fram till och med vårterminen 2018 har delprovet *Focus: Writing* i stor utsträckning skrivits för hand. Det har dock varit möjligt för skolan att låta eleverna skriva på dator om eleverna i ordinarie undervisning är vana vid att skriva på dator samt om skolan kan tillhandahålla datorer till alla elever. Skolan måste ta hänsyn till säkerhetsaspekter, som att uppgifter och elevtexter inte sprids via internet samt att elever inte har tillgång till stavnings- och grammatikkontroller. Att vissa elever har fått skriva på dator och andra inte kan förstås uppfattas orättvist av både elever och lärare, vilket också framkommer i SOU 2016:25. Även bland de elever som skriver delprovet på dator förekommer dock stora skillnader i förutsättningar när det gäller exempelvis programvara, skärmstorlek och tangentbord. Att många elever skrivit delprovet för hand har också genererat frågeställningar gällande validiteten i form av autenticitet, då det i en icke provsituation vore en självklarhet både att skriva på dator och att ha tillgång till stavnings- och grammatikkontroll.

Empirisk undersökning

I en empirisk undersökning som genomfördes under 2014 inom NAFS-projektet, studerades specifikt hur användandet av dator med stavnings- och grammatikkontroll påverkar elevers prestationer vid genomförandet av den skriftliga delen i de nationella proven, samt elevers och lärares upplevelse av detta (Göteborgs universitet, 2014). En storskalig utprovning genomfördes i

årskurs 9 samt i kurserna engelska 5 och 6 på gymnasiet med sammanlagt 574 elever. Vid utprövningen använde de deltagande skolorna den programvara och de hjälpfunktioner som respektive skola hade tillgång till, varför dessa varierade mycket såväl på skol- som på individnivå.

En del av denna studie innebar att 30 texter slumpmässigt valdes från respektive årskurs/kurs utifrån resultat på en ankaruppgift²⁸. Med hjälp av resultatet på ankaruppgiften delades elevtexterna in i tre kategorier. Dessutom gjordes på individnivå, i en klass per utprövningsgrupp (åk 9, eng 5, eng 6), en jämförelse mellan 2014 års ordinarie nationella prov, vilket skrevs för hand, och utprövningstexten som skrevs på dator och med stavnings- och grammatikkontroll.

Tendensen avseende samtliga grupper i den storskaliga utprövningen är att de flesta elever tycks utnyttja stavningskontrollen väl och gör relativt få stavfel. Elevgruppen med lägst poäng på ankaruppgiften tenderar att göra flest stavfel. Dessa stavfel utgörs främst av felaktiga ordval som ligger nära det tänkta ordet, t.ex. *cloths* i stället för *clothes*, *sad* istället för *said*. Dessa misstag kan förmodligen vara exempel på när stavningskontrollen ger förslag till stavning av ordet men eleven väljer fel. Bland elever med högst poäng på ankaruppgiften förekommer mycket få stavfel. Det är dock svårt att avgöra om detta beror på att dessa elever hanterar stavningsfunktionen väl, eller om de generellt sett behärskar stavning relativt väl.

Grammatikkontrollen tycks vara svårare för eleverna att använda. I gruppen med lägst resultat på ankaruppgiften förekommer ett flertal grammatikfel, vilket kan antyda att eleverna inte är förtrogna med verktyget men också att de har bristande grammatiska kunskaper. Dock uttrycker eleverna, oavsett språklig nivå, att upplevelsen av att ha tillgång till dessa båda hjälpfunktioner är positiv och att en känsla av trygghet infinner sig. Detta gör att eleverna upplever sig kunna fokusera på andra aspekter av sitt skrivande än grammatik och stavning.

Att eleverna har gott stöd av stavningsfunktionen bekräftas i den individbaserade delen av studien som jämför en elevs text skriven för hand med samma elevs text skriven på dator med stavnings- och grammatikkontroll. Majoriteten elever har färre antal stavfel i den datorskrivna texten, vilket gäller såväl årskurs 9 som engelska 5 och 6. Den typ av stavfel som ändå förekommer mest frekvent i de datorskrivna texterna omfattar homofoner, dvs. ord som uttalas liknande,

²⁸ Ankaruppgiften används i samtliga utprövningar och ger en indikation om grupper och individers språkfärdighetsnivå i engelska.

eller ord som ligger nära det tänkta ordet, exempelvis *to – too; see – sea; then – than*. Eftersom stavningskontrollen inte reagerar på korrekt stavade ord, oavsett sammanhanget, markeras inte dessa som fel och uppmärksammas inte av eleven. En annan typ av stavfel som förekommer är när eleven har fått ordet markerat och därefter väljer fel alternativ i menyn som kommer upp. I dessa fall kan felet ibland bli förstörande. Följande exempel illustrerar detta: *I fell wonderful when I have a lot to do at the same time*.

Att grammatikkontrollen är svårare att hantera än stavningskontrollen bekräftas även i den jämförande studien. En tendens tycks vara att elever på språkligt lägre nivå har svårast att använda funktionen. Bland eleverna med ett mer utvecklat språk tycks antalet kongruensfel och tempusfel ha minskat något, medan exempelvis fel som har med prepositioner och meningsbyggnad att göra kvarstår.

I studien ombeds deltagande elever och lärare besvara en enkät. Majoriteten av lärarna är generellt positiva till att eleverna ska skriva delprovet på dator och då använda stavnings- och grammatikkontroll. Att genomföra provet på ett korrekt sätt medförde dock tekniska svårigheter och krävde en hel del förberedelse. Som nämnts tidigare uttrycker en majoritet av eleverna att de föredrar att skriva på dator. De anser också att det är relativt lätt att använda stavnings- och grammatikkontroll och att dessa funktioner hjälper dem att producera bättre texter. Följande citat illustrerar olika aspekter av detta:

- Jag tycker att det sparar mycket tid när man till exempel är osäker på hur ett ord stavas. Man behöver då inte fastna vid att försöka komma på hur det stavas.
- Att ha stavningskontroll hjälpte och jag tycker att man lär sig av det också. Om man t.ex. skrev *whit* och ändrade till *with* så kom man ihåg det till nästa gång man skrev ordet.
- Det underlättar i och med att jag slapp fundera om jag skrivit rätt eller inte. Det minskar oron helt enkelt.

Dock finns det också elever som problematiserar dels att de tekniska förutsättningarna ser mycket olika ut på grund av tillgången till olika mjukvaror och datorer, dels att elever som redan har goda kunskaper gällande grammatik och stavning kan komma att missgynnas.

- If we do the national tests with our computers it will be hard to know how good the students are at this. It also feels kind of unfair to those who have practiced on these skills a lot.

ATT BEDÖMA SPRÅKLIG KOMPETENS

– There are no English grammar/spell check on [typ av mjukvara]. So it was impossible to use.

Sammantaget är emellertid elevernas upplevelse av att skriva på dator i hög grad positiv. Enkätsvaren visar att 80–90 % av eleverna från såväl årskurs 9 som engelska 5 och 6 föredrar att använda datorn vid skriftlig produktion och interaktion. Följande citat från undersökningen visar på elevernas positiva inställning:

– Det blir en bättre styckeindelning och jag känner själv att det gör mitt arbete bättre. Det går snabbare för mig att skriva på dator och jag sparar på så sätt tid att tänka på vad jag vill skriva.

– I seriously need a computer to write decent texts. My handwriting is horrendous and hinders me in writing.

Att gå från papper till dator – vad händer då?

Det som vid en första anblick ser lätt och självklart ut – att skriva text på dator i provsammanhang – för med sig metodeffekter och konsekvenser. Vid övergången till datorbaserad skrivning kommer kunskapsobjektet för provet och provspecifikationerna att förändras, vilket behöver hanteras i relation till provningens validitet.

I enlighet med nuvarande styrdokument avser delprovet *Focus: Writing* att pröva och ge möjlighet att bedöma språkliga kvaliteter som tydlighet, ledighet, variation och struktur. I detta inbegrips faktorer som grammatiska strukturer, fraseologi, idiomatik, ordförråd, syntax, sammanhang och stavning. I den stund dator används i provsituationen kan, utöver elevens förmåga att uttrycka sig på engelska i skrift, själva hanteringen av tekniken komma att bedömas. Utifrån ett likvärdighetsperspektiv skulle följaktligen eventuella hjälpfunktioner behöva standardiseras, t.ex. i form av ett verktygsfält som ingår i provet och elever kan bli bekanta med genom exempelmaterial. Vilka hjälpfunktioner som kan vara relevanta, och i vilken utsträckning dessa ska användas vid provsituationen, behöver noggrant övervägas och analyseras med hänsyn till vad delprovet avser pröva.

I vilken mån får då ett datorbaserat prov ett förändrat *construct*, dvs. att ytterligare något mäts än det som mätts tidigare? Kan god användning av stavnings- och grammatikfunktioner ses som en språklig strategi och på så sätt vidga kunskapsobjektet för provet? Kommer kraven på korrekthet vid textskrivande att öka om eleverna har tillgång till hjälpfunktioner? Handlar det då om att skriva engelsk text på dator snarare än att skriva text på engelska?

En del av gällande kunskapskrav för skriftlig produktion, som i dag inte bedöms i de nationella proven, innefattar bearbetning av text. Genom ett införande av datorbaserade prov skulle viss bearbetning med ordbehandlings- och textredigeringsfunktioner eventuellt kunna införlivas i det som avses prövas, och följaktligen vidga kunskapsobjektet för delprovet.

En ytterligare aspekt av att genomföra det skriftliga delprovet på dator kan vara möjligheten att distribuera elevtexter till medbedömare och externa bedömare samt att smidigt avidentifiera elevtexter. Medbedömning påverkar bedömningens reliabilitet positivt, vilket lyfts fram i Skolverkets rapport *Sambedömning i skolan* (Skolverket, 2014b). I sammanhanget kan noteras att system för automatiserad rättning av skriftlig produktion (*automated essay scoring – AES*) utvecklas allt mer och att en kombination av denna typ av rättning och lärarbedömning torde kunna bli aktuell i framtiden (t.ex. Weigle, 2013).

Slutord

Utvecklingen mot datorbaserade nationella prov i engelska är både välkommen och angelägen. Som framkommit finns det såväl möjligheter som utmaningar i denna process. För att övergången ska kunna ske på ett så tryggt och säkert sätt som möjligt behövs ett fortsatt varsamt, men ändå tydligt framåtriktat analys- och utvecklingsarbete, med fokus på aspekter i relation till reliabilitet och validitet. Detta inte minst med tanke på de nationella provens syfte att stödja en likvärdig och rättvis bedömning och betygssättning.

Referenser

- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Chapelle, C. (2010). Computer Assisted Teaching and Testing. I Long, M.H. & Doughty, C.J. (Red.), *The Handbook of Language Teaching* (s. 628–644). Chichester: Wiley Blackwell.
- Chapelle, C. & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University press.

ATT BEDÖMA SPRÅKLIG KOMPETENS

- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder Education.
- Dörnyei, Z. (2001). *Teaching and Researching Motivation*. Harlow: Longman.
- Educational Testing Service (2011). *Practical Considerations in Computer-Based Testing*, Tim Davey, Research Director, ETS Research & Development Division. Hämtad 2017-01-12 från https://www.ets.org/Media/Research/pdf/CBT_2011.pdf
- Erickson, G. (1998). Test-taker feedback – Elevers synpunkter som del av arbetet med nationella prov i engelska. I B. Ljung och A. Pettersson (red.), *Perspektiv på bedömning av kunskap*, (s. 79–100). Lärarhögskolan i Stockholm, Institutionen för pedagogik.
- Erickson, G. (2006). Bedömning av och för lärande. En kollaborativ ansats i arbetet med nationella prov i språk. I U. Tornberg (Red.), *Mångkulturella aspekter på språkundervisningens kommunikativa praktiker. En konferensrapport*. Örebro: Örebro universitet, Pedagogiska Institutionen.
- Erickson, G. (2009). "Att bäras åt" – Om den goda bedömningens flersidighet och ömsesidighet. I U. Tornberg, et al. *Språkdiraktiska perspektiv. Om lärande och undervisning i främmande språk* (s. 159–174). Stockholm: Liber.
- Erickson, G. & Gustafsson, J-E. (2005). *Some European Students' and Teachers' Views on Language Testing and Assessment. A report on a questionnaire survey*. European Association of Language Testing and Assessment. Hämtat 2016-12-09 från <http://www.ealta.eu.org/resources.htm>
- Göteborgs universitet. (2014). *Rapport av uppdrag avseende utprovning av Focus: Writing med stavnings- och grammatikkontroll*. Arvidsson, U., Axelsson, P., Brorson, J. & Lindqvist, AK. Intern rapport till Skolverket.
- Lindqvist, AK. (2012). *Datorbaserad bedömning i språk – forskningsläge, erfarenheter och implikationer*. Masteruppsats i ämnesdidaktik. Göteborg: Göteborgs universitet.
- Meier, R. (2009). *Transition to Computer-based Assessment: Motivations and Considerations*. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 104–107). JRC Scientific and Technical Reports.
- Messick, S. A. (1989). Validity. I Linn, R.L. (Red.), *Educational Measurement*. (Third edition), (s. 13–103). New York: American Council on Education/Macmillan
- OECD. (2009). *PISA 2009. Reading Framework*. Hämtat 2016-12-12 från <http://www.oecd-ilibrary.org>
- OECD. (2013). *PISA 2015. Draft reading literacy framework*. Hämtat 2016-12-11 från <https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%200Reading%20Framework%20.pdf>
- Skolverket. (2010). *Adaptiva och andra datorbaserade prov – en kunskapsöversikt*. Peter Nyström (Red.). Stockholm: Skolverket.
- Skolverket. (2011a). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket. Hämtad 2017-06-27 från

- <https://www.skolverket.se/publikationer?id=2575>
Skolverket. (2011b). *Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*. Stockholm: Skolverket. Hämtad 2017-06-27
<https://www.skolverket.se/publikationer?id=2705>
- Skolverket. (2013). *Digital och traditionell läsning. Analys av olika elevgruppers läsning utifrån PISA 2009*. Skolverkets aktuella analys 2013. Stockholm: Skolverket.
- Skolverket. (2014a). *Digital läsning i engelska*. Bedömningsstöd Engelska. Bjerklund Larsson, A. & Lindqvist, AK. Stockholm: Skolverket.
- Skolverket. (2014b). *Sambedömning i skolan*. Stödmaterial. Thornberg, P. & Jönsson, A. Stockholm: Skolverket. Hämtad 2017-06-26 från
<https://www.skolverket.se/publikationer?id=3172>
- Statens offentliga utredningar. (2016). *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. (SOU 2016:25). Betänkande av Utredningen om nationella prov. Stockholm 2016. Hämtad 2017-06-27 från
<http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2016/03/sou-201625/>
- Sveriges Riksdag. (2015). *Översyn av de nationella proven för grund- och gymnasieskolan*. Kommittédirektiv 2015:36. Stockholm 2015. Hämtad 2017-06-27 från
<https://www.riksdagen.se/sv/dokument-lagar/dokument/kommittedirektiv/oversyn-av-de-nationella-proven-for-grund-och-h3b136>
- Weigle, S. (2013). English as a Second Language Writing and Automated Essay Evaluation. I Shermis, M.D. & Burstein, J. (Red.). *Handbook of Automated Essay Evaluation Current Applications and New Directions* (s. 36–54). London: Routledge.
- Weir, C. J. (2005). *Language Testing and Validation. An Evidence Based Approach*. Oxford: Palgrave Macmillan.
- Wästlund, E. (2007). *Experimental Studies of Human Computer Interaction: Working memory and mental workload in complex cognition*. Doktorsavhandling i psykologi. Psykologiska institutionen, Göteborgs universitet.