



GÖTEBORGS UNIVERSITET
INST FÖR DIDAKTIK OCH PEDAGOGISK PROFESSION

Datorbaserad bedömning i språk

forskningsläge, erfarenheter och
implikationer

AnnaKarin Lindqvist

Examensarbete:	30 hp
Program och/eller kurs:	Examensarbete i ämnesdidaktik
Nivå:	Avancerad nivå
Termin/år	HT 2011
Handledare:	Gudrun Ericson
Examinator:	Shirley Booth
Rapport nr:	HT11-IDPP-04-PDA462

Abstract

Examensarbete:	30 hp
Program och/eller kurs:	Examensarbete i ämnesdidaktik
Nivå:	Avancerad nivå
Termin/år	HT 2011
Handledare:	Gudrun Ericson
Examinator:	Shirley Booth
Rapport nr:	HT11-IDPP-05-PDA462

To support uniform and fair assessment of Swedish students' English language competences, mandatory National tests are provided both for secondary and upper secondary education. The current thesis may be seen as a contribution to the further development and validation of these tests, with regard to the likely introduction of computerized testing methods. The aim is to gain further knowledge from current research in the field, to study and elucidate international experiences of implementation, and to gather and analyze feedback from students and teachers. Based on my findings, some tentative conclusions will be drawn, and recommendations made concerning future developments. During the last few decades technology has become an integrated part of language use in most domains. When reading, writing and communicating in other ways, computers are replacing traditional paper-and-pencil methods. Assessment has to follow and adjust to this (r)evolution. Computer based language materials for teaching and assessment are widely spread and many countries have introduced large-scale computer based language testing. In Sweden this journey has cautiously begun and it is probably more a question of *how*, than *if* the Swedish national tests of English will be computer based in the near future. The study shows that the most important issues in research about computer based assessment concern different aspects of validity. The impact of rubric on the assessed construct is essential, that is the impact of 'the how' on 'the what'. Depending on the extent to which software, innovative formats and the Internet are involved in the assessment process, the test construct must be revised and redefined accordingly. A possible new construct, as well as validity aspects of equity/fairness, practicality, accountability and good practice are emphasized in the literature. Washback-effects at different levels are discussed, and several researchers raise warnings against letting technology drive the development of good assessment towards 'efficiency' and overuse of simple formats. The empirical study undertaken was carried out at six Swedish schools, with more than 200 students and their 12 teachers. Contacts with the University in Bergen made it possible to use national computer based tests of English reading proficiency, already implemented in Norway. The study was performed using a case study approach. Each school was visited twice, for introductory meetings and for observation of the test taking process; field-notes were taken on these occasions. Questionnaires were answered, by students before and after taking the test and by teachers and head teachers in retrospect, with similar questions to allow comparisons. Data were analyzed inductively from both a quantitative and a qualitative perspective, and allow certain triangulation. The results indicate that the participating schools do not (yet) have the required resources to perform national computer based tests in a secure way, and the respondents agree that computers are seldom used in their English classrooms. All respondents are positive to the idea of computer based testing of English and to the actual tests as such, younger students, and boys, being the most motivated. The students also point out motivational aspects concerning, for instance, anxiety and show valuable insight in validity related aspects. The conclusion of the thesis is that the [likely] implementation of computerized testing methods in the Swedish school system must build on research and needs to take into account opinions and suggestions of future users, not least those expressed by students.

Innehållsförteckning

Abstract	2
Innehållsförteckning.....	4
Förord.....	6
1 Inledning och problemavgränsning.....	6
1.1 Uppsatsens bakgrund och sammanhang.....	6
1.2 Övergripande avsikt med uppsatsen	7
1.3 Avgränsningar	7
2 Problemformulering, syfte och frågeställningar	8
2.1 Problemformulering.....	8
2.2 Uppsatsens syfte och frågeställningar	8
3 Begreppslig bakgrund.....	9
3.1 Centrala begrepp	9
3.2 Europarådet och Gemensam europeisk referensram för språk.....	10
3.3 Nationella styrdokument och språksynen i kurs- och ämnesplanerna.....	11
4 Bedömning av språklig kompetens	13
4.1 Teoretiska utgångspunkter	13
4.2 En vidgad syn på bedömning	14
4.3 Olika typer av bedömning.....	15
4.3.1 Vad som avses bedömas.....	15
4.3.2 Olika sätt att genomföra bedömning	16
4.4 Bedömningens användning och konsekvenser.....	17
4.5 Etiska aspekter — Good Practice	18
5 Nationella prov- och bedömningsmaterial i engelska.....	19
5.1 Historik och inramning.....	19
5.2 Motivation och Test Taker Feedback (TTF).....	20
5.3 Ett nytt steg för de svenska nationella proven i språk?.....	22
6 Datorbaserad av bedömningsmaterial i språk.....	23
6.1 Utveckling av datorbaserad bedömning i språk	23
6.2 Incitament till att datorbasera bedömning i språk	24
6.3 Förhållningssätt till datorbaserad av bedömning i språk.....	25
6.4 Validitetsaspekter relaterade till datorbaserad bedömning	26
6.4.1 Aspekter relaterade till rubric, reliabilitet och process.....	27
6.4.2 Aspekter relaterade till autenticitet och interaktivitet	28
6.4.3 Aspekter relaterade till construct och construct validity	29
6.5 Påverkans effekter vid datorbaserad av bedömning i språk	33
6.5.1 Washback.....	33
6.5.2 Erfarenheter av och åsikter om implementering av CBT.....	35
6.6 Likvärdighetsaspekter – fairness, bias och familiaritet.....	37
6.6.1 Likvärdighet och bias	37
6.6.2 Likvärdighet i relation till validitet	38

6.6.3 Jämförande studier – familiaritet	39
6.6.4 Jämförande studier – genusaspekter	40
6.6.5 PISA 2009	41
6.7 Avslutande kommentar: Ett nytt steg för de nationella provmaterialen i språk	42
7 Empirisk studie – metod och design	43
7.1 Metod.....	43
7.1.1 Ansats	43
7.1.2 Undersökningens struktur	43
7.1.3 Forskningsetiska överväganden	44
7.1.4 Urvalsgrupp	44
7.2 Material.....	45
7.2.1 Uppgiftsdomän	45
7.2.2 Provmaterial	45
7.2.3 Provuppgifter.....	45
7.2.4 Enkäter och fältanteckningar.....	46
7.2.5 Bearbetning	47
7.2.6 Reliabilitet, validitet och generaliserbarhet	47
7.3 Genomförande	48
8 Empirisk studie - resultat	50
8.1 Deltagande observation.....	50
8.2 Enkäter till eleverna	50
8.2.1 Elevenkät I	51
8.2.2 Elevenkät II	56
8.2.3 Elevenkät I och II.....	61
8.2.4 Sammanfattande kommentar	62
8.3 Enkäter till lärarna.....	63
8.3.1 Slutna svarsformat.....	63
8.3.2 Öppna kommentarer	65
8.3.3 Sammanfattande kommentar	67
8.4 Enkäter till skollädaarna	67
8.4.1 Slutna svarsformat.....	68
8.4.2 Öppna kommentarer	68
8.4.3 Sammanfattande kommentar	69
8.5 Elevernas resultat på proven	69
9 Sammanfattande diskussion	71
9.1 Design, metod och material	71
9.2 En förändrad syn på språklig kompetens – en didaktisk utmaning	73
9.3 Förhållandet mellan construct/ 'Vad' och rubric/ 'Hur', samt effekter därav, vid bedömning av språklig kompetens via dator.....	74
9.4 Datorbaserade prov i engelska på några svenska skolor — förväntningar, reaktioner, resultat	77
9.5 Implikationer för det svenska nationella provsystemet i engelska	78
Referenser	81
Bilagor	86

Förord

Som lärare i språk började jag under 1990-talet aktivt intressera mig för utvecklings- och bedömningsfrågor, vilket så småningom lett till mitt nuvarande arbete med utveckling av de nationella provmaterialen i engelska för grund- och gymnasieskolan inom projektet *Nationella Prov i Främmande Språk (NAFS)* vid Göteborgs universitet. Ämnet för min uppsats har inspirerats av och har stark anknytning till mitt arbete inom projektet NAFS. Jag vill tacka mina arbetskamrater, som uppmuntrat och uthärdat mitt arbete med denna uppsats och rikta ett alldeles speciellt tack till min handledare, Gudrun Erickson.

1 Inledning och problemavgränsning

I inledningen presenteras studiens bakgrund och sammanhang inom det nationella provsystemet samt det Skolverksuppdrag inom vilket den i uppsatsen aktuella empiriska studien genomfördes. Även den övergripande avsikten med uppsatsen och dess avgränsningar redovisas.

1.1 Uppsatsens bakgrund och sammanhang

Projektet *Nationella Prov i Främmande Språk (NAFS)*, Institutionen för pedagogik och Specialpedagogik vid Göteborgs universitet, arbetar på uppdrag av Skolverket med utveckling av och forskning kring de nationella provmaterialen i engelska och moderna språk för den svenska grund- och gymnasieskolan. Inom projektet finns en lång tradition när det gäller utvecklingsarbete, pilotstudier och storskaliga utprövningar av provmaterial i nära samarbete med yrkesverksamma lärare och inte minst med elever (Erickson, 2006; Lindblad, 1991).

Ämnet för min uppsats har valts inom ramen för denna tradition och det fortlöpande forsknings- och utvecklingsarbete som bedrivs. Den empiriska basen består av data från en pilotstudie rörande datorbaserade prov i engelsk läsförståelse som preliminär rapporterades i deskriptiv form till Skolverket i maj 2009 (Lindqvist, 2009). Studien var en del av ett större Skolverksuppdrag med syfte att skapa en kunskapsöversikt över datorbasering av provmaterial inom och utom Norden, i samarbete med Umeå universitet, Institutionen för beteendevetenskapliga mätningar (Skolverket/Nyström, 2010). Pilotstudierna i matematik och engelska i grundskolan hade i Skolverkets uppdragsbeskrivning till syfte att undersöka och kartlägga enligt följande:

Tyngdpunkten ligger på attityder, uppfattningar och reaktioner hos elever och lärare och eventuellt skolledare och har fokus på beredskapen för användning av datormediet som sådant i relation till prov och bedömning, dels på ett allmänt plan och dels med fokus på den datorbaserade formen (Ur Överenskommelse och uppdragsbeskrivning, Skolverket, jan. 2009).

Denna uppsats innehåller en bearbetning och fördjupning av min del av denna rapportering, främst när det gäller vetenskaplig förankring, vidare resultatanalys och diskussion.

1.2 Övergripande avsikt med uppsatsen

En långsiktig och övergripande avsikt med min uppsats är att kunna bidra till utvecklingen av den bedömning av språk som görs på svensk, nationell nivå, framför allt med avseende på eventuell datorbaserad bedömning. Eftersom datorbaserad bedömning förmodligen mer är en fråga om *hur* och *när* än *om*, var den omedelbara anledningen till den empiriska studien, i konsekvens med Skolverksuppdraget, att bidra till underlag för framtida politiska beslut. Det förefaller alltså viktigt att vara förberedd på en (r)evolution när det gäller digital bedömning och därför, inför framtida implementering, också dra lärdom av aktuell forskning och andras erfarenheter, samt – inte minst – av användarna, det vill säga i första hand eleverna, eftersom det är deras verklighet det verkligen gäller och påverkar. Ytterligare en förhoppning är att denna uppsats ska utgöra inspiration till och visst underlag för vidare studier, diskussioner och eventuella framtida beslut inom NAFS-projektet. Det faktum att flera kontakter tagits både inom och utom Norden kan också främja ett vidare erfarenhetsutbyte.

1.3 Avgränsningar

Fokus i föreliggande studie är i huvudsak att belysa datorbaserade språkprovs validitet och praktikalitet i förhållande till användaren och kunskapsobjektet. Flera avgränsningar och inskränkningar har varit nödvändiga, främst givet ämnesområdets komplexitet. En sådan avgränsning handlar om att jag i min forskningsöversikt har koncentrerat mig på datorbaserad bedömning i språk, och framför allt i främmande språk. Inom det svenska nationella provsystemet i språk erbjuds frivilliga provmaterial i tyska, franska, spanska men Sverige har obligatoriska nationella prov endast i engelska för grund- och gymnasieskolan. Mitt fokus avseende den svenska kontexten har därför varit prov i engelska. Detta medför också att jag framför allt avsett studera konsekvenser för bedömning i storskaliga provsammanhang, eftersom uppsatsen bl.a. avser belysa framtida överväganden när det gäller de svenska nationella proven i engelska. Det finns även en stor mängd forskning kring datoranvändning för andra syften och i andra pedagogiska sammanhang, vilket dock inte fallit inom ramen för denna uppsats.

2 Problemformulering, syfte och frågeställningar

Nedan formuleras uppsatsens aktuella utgångspunkt, syfte samt två forskningsfrågor.

2.1 Problemformulering

Det är numera självklart att använda datorer i de flesta sammanhang som har med språk och kommunikation att göra, och det är förmodligen också självklart att de svenska nationella prov- och bedömningsmaterialen i framtiden kommer att vara mer eller mindre datorbaserade. En förändring av *Hur* något görs påverkar emellertid också *Vad* som avses göras, och att utan noggrann eftertanke och analys datorbasera framför allt storskaliga prov- och bedömningsmaterial skulle därför få stora konsekvenser. En god beredskap kan skapa förutsättningar att hantera oönskade och ta till vara önskade konsekvenser av datorbasering av svenska provmaterial.

2.2 Uppsatsens syfte och frågeställningar

Syftet med denna uppsats är att söka erfarenhet i aktuell bedömningsforskning samt i någon mån undersöka den tekniska och pedagogiska beredskapen i svenska skolor att genomföra datorbaserade prov i engelska — i relation till prov och bedömning och med fokus på den datorbaserade formen.

På basis av aktuell bedömningsforskning och de data som samlats in avser jag i min uppsats försöka besvara följande forskningsfrågor:

1. *Vilket är det aktuella forskningsläget avseende relationen mellan construct/'Vad' och rubric/'Hur', samt effekter av detta, när det gäller bedömning av språklig kompetens via dator?*
2. *Hur fungerar datorbaserade prov i engelska på ett urval av svenska skolor — förväntningar, reaktioner, resultat?*

Vidare avser jag, på basis av ovanstående resultat reflektera över vilka implikationer resultaten av forskningsgenomgången och den empiriska studien har för vidare utveckling av det svenska nationella provsystemet i engelska.

3 Begreppslig bakgrund

Nedan presenteras ett antal begrepp som är relevanta för bedömning av språklig kompetens i relation till datorbaserad. Vidare presenteras Europarådets roll för de svenska styrdokumenterna samt språksynen i de svenska styrdokumenterna för språk.

3.1 Centrala begrepp

Grundläggande för bedömningens kvalitet är resultatens 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 (*consequential validity*). *Reliabilitet*, eller tillförlitlighet, handlar om hur "rättvis" och icke slumpmässig bedömningen är avseende till exempel jämförbarhet och likvärdighet (Erickson, 2009a). Att det som prövas innehåller ovidkommande inslag (*construct irrelevant variance*) eller för lite av det som avses (*construct under-representation*) är exempel på sådant som betonas som hot mot validiteten (Messick, 1989). Detta kan leda dels till ogrundade bedömningar, dels till *bias*, det vill säga att (grupper av) testtagare systematiskt gynnas eller missgynnas på felaktiga grunder.

En aspekt med betydelse för bedömningens validitet är dess *autenticitet* och *interaktivitet*, det vill säga i vilken mån uppgiften ger provtagaren möjlighet att visa sina kunskaper. Bachman & Palmer (1996) definierar *autenticitet* som överensstämmelsen mellan det som karakteriserar provuppgiften och provsituationen, och det som kännetecknar den målspråkssituation man avser pröva.

När det gäller implementeringen av ett prov som ett led i undervisningen betonas dess praktikalitet (*practicality/feasibility*), det vill säga dess användbarhet i förhållande till vilka resurser som krävs (Bachman & Palmer, 1996; Hughes, 2003). Olika former av bedömning har eller kan också ha olika typer av effekter på undervisningen och samhället, så kallad *washback* eller *impact* (Gipps, 1994; Messick, 1996).

Vid framtagande och utveckling av prov- och bedömningsmaterial av god kvalitet, är reaktioner och synpunkter från elever, *Test Taker Feedback* (TTF), och lärare, *Teacher Feedback* (TF) av stor betydelse (Erickson, 1998, 2009a). En viktig roll för arbetet med användarrespons spelar analysen av olika typer av motivationsaspekter, som till exempel kapacitetsupplevelse, oro och uppfattningar om bedömningens relevans, vilket diskuteras av bland andra Dörnyei (2001) och Hughes (2003).

Det förekommer en mängd olika termer för olika typer av datorbaserad i samband med prov. Som ett samlande begrepp används generellt *Computer Based Testing/Tests* (CBT) för bedömning som görs och besvaras på datorer, till skillnad från dator-distribuerade prov som skrivs ut och besvaras för hand eller eventuellt muntligt. Motsvarande bedömning i språk benämns ofta *Computer Assisted Language Testing/Tests* (CALT). I förekommande fall benämns traditionella prov som görs med papper och penna ofta PBT (*Paper Based Tests/Testing*).

Datorbaserad utvärdering utgör i sig en del av bedömningens *format*, dvs. det sätt på vilket testtagaren förväntas visa sin kompetens. Format, liksom instruktioner och bedömningsanvisningar utgör de tre viktigaste komponenterna i det som benämns bedömningens *rubric* (OECD, 1999). Fördelar när det gäller konsistens och likvärdighet avseende *rubric* vid datorbaserad utvärdering betonas och diskuteras av Chapelle & Douglas (2006).

En viktig distinktion görs mellan två huvudtyper av prov på dator, beroende på i vilken utsträckning underlaget är datoriserat, tekniken utnyttjas och hur aktivt datorprogrammet är. Termen *Computer Based Tests* (CBT) används, förutom övergripande, också specifikt som beteckning för datorbaserade *linjära* prov då samtliga provtagare gör identiska prov. Dessa prov kan rättas antingen manuellt eller maskinellt, beroende på mjukvarans kapacitet och uppgifternas format. En annan typ benämns *Computer Adaptive Tests* (CAT), det vill säga *adaptiva* prov som i viss mån successivt anpassar sig efter provtagarens prestation och slumpar fram uppgifter på olika svårighetsnivå. Dessa prov består generellt av flervalsformat som rättas maskinellt och avbryts då programmet kan definiera en stabil nivå.

3.2 Europarådet och Gemensam europeisk referensram för språk

Europarådet är en europeisk samarbetsorganisation som bildades 1949 och som i huvudsak arbetar med att främja demokrati, mänskliga rättigheter och rättsstatsutveckling. Europarådets språkpolicy, till vilken Sverige anslutit sig, har enligt ministerrådets rekommendationer som yttersta syfte att uppnå större enighet bland sina medlemmar (Council of Europe, 2001; Skolverket, 2009a). I de nya ämnes- och kursplanerna i språk Gy 2011 och GR 2011 relaterar de svenska styrdokumenterna ännu tydligare än tidigare till denna policy.

Under 1990-talet utvecklades på uppdrag av Europarådet en gemensam referensram för språk, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR), och i översättning till svenska 2009 *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning* (GERS), (Council of Europe 2001/Skolverket 2009a). GERS bygger på Europarådets tidigare arbete inom språkforskning, speglar en vid och funktionell språksyn och definierar språkfärdighet i olika nivåer. Den har blivit ett centralt och inflytelserikt dokument, vars syfte är "att övervinna de kommunikationshinder mellan människor som arbetar med moderna språk, som uppstår till följd av de olika utbildningssystemen i Europa" (GERS, s. 1). Referensramen skapar en gemensam grund för att utarbeta till exempel kursplaner, examina och läromedel för språkinläring, samt för att på lokal nivå kunna göra jämförelser och skapa enhetlighet mellan olika undervisningsspråk och på sikt kunna skapa en större samsyn internationellt.

GERS präglas av en handlingsorienterad grundsyn (*action oriented approach*) vilket innebär att man betraktar människor som använder eller lär sig ett språk som sociala aktörer, dvs. samhällsmedborgare vilka, i olika kontexter, måste klara av vissa uppgifter, varav en del är språkrelaterade:

Språkanvändning – vilket inkluderar språkinläring – omfattar de handlingar som människor utför i egenskap av individer och sociala aktörer. Vi utvecklar en rad olika **kompetenser**, både generella och i synnerhet **kommunikativa språkliga kompetenser**. Vi utnyttjar de kompetenser vi förfogar över i olika kontexter och med olika **villkor** och **begränsningar** för att utföra **språkliga aktiviteter** som inbegriper **språkliga processer** där målet är att producera och/eller ta emot **texter** som rör **teman** inom särskilda **domäner**. Det sker genom att vi tillämpar de strategier som verkar passa bäst för de **uppgifter** som ska utföras. De erfarenheter som deltagarna själva gör i samband med detta leder till att deras kompetenser stärks eller förändras (GERS, s. 9).

Den *kommunikativa språkliga kompetensen* definieras som i sig bestående av flera komponenter: den *lingvistiska*, som handlar om lexikala, fonologiska, syntaktiska och liknande kunskaper och färdigheter, den *sociolingvistiska* till vilken räknas bland annat anpassning och medvetenhet om sociala konventioner samt den *pragmatiska*, dvs. den funktionella användningen av språkliga resurser som till exempel diskurskompetens, att producera strukturerat, sammanhängande språk och språk i samspel.

De språkliga aktiviteterna delas vidare in i *reception*, *produktion*, *interaktion* och *mediering*. Dessa aktiviteter används och kontextualiseras sedan inom olika *domäner*, vilka betecknas samhällsdomänen, den personliga domänen, utbildnings- respektive yrkesdomänen. Såvida den inte är automatiserad, kräver vidare all språklig aktivitet och kommunikation att användaren också behärskar ett register av *strategier* för att kunna lösa och hantera de uppgifter han eller hon ställs inför.

För att kunna beskriva olika kvalitativa referensnivåer i språk har GERS en global skala med tre övergripande nivåer: A (Användare på nybörjarnivå), B (Självständig användare) och C (Avancerad användare) med vardera två delnivåer. De olika nivåerna illustreras per kompetens och språklig aktivitet i skalor av skilda slag, bl.a. med så kallade "Man kan"- deskriptorer (*can do statements*).

Referensramen konkretiserar och resonerar också utförligt kring vilka implikationer den handlingsorienterade grundsynen får för språkinläring, språkundervisning och bedömning. Övergripande för den sociolingvistiska språkliga kompetensen är att undervisningen skall främja interkulturell medvetenhet eller interkulturalitet (*intercultural education*). Begreppet flerspråkighet (*plurilingualism*) är centralt och definieras i förlängningen som att en person "bygger upp en kommunikativ kompetens där alla språkkunskaper och språkerfarenheter ingår, och där språken är förbundna med och påverkar varandra" (GERS, s. 4).

3.3 Nationella styrdokument och språksynen i kurs- och ämnesplanerna

Skolans undervisning i språk och synen på vad som är relevanta kunskaper i språk har sin grund i den språksyn som präglar de nationella styrdokument. Dessa styrdokument beror i sin tur av sin samtid, rådande politiska system och aktuell forskning.

Malmberg (2000, 2001) och Andered (2001) beskriver hur språksyn och skrivningar i de svenska styrdokument alltsedan 1970-talet i mycket stor utsträckning präglas av Europarådets

språkpolicy och funktionella språksyn. Språkundervisningen har blivit en integrerad del av den svenska skolans värdegrund och pedagogiska grundsyn, och språk betraktas som verktyg till att vidga användarens perspektiv. Tornberg (2000) vidgar detta resonemang ytterligare och menar att det kommunikativa kompetensbegreppet har betydelse för "språkklassrummet som en möjlig demokratisk mötesplats" (s. 25) och uttrycker i sin avhandling en avsikt att:

Istället för att ställa frågor om hur elever på olika sätt tillgodogör sig ett innehåll eller utvecklar sina språkliga färdigheter [...] rikta fokus på två av språkundervisningens mest prominenta begrepp, nämligen "kommunikation" och "kultur" och knyta dem till relationen mellan språkundervisning och demokrati (sid. 15).

Inom Europarådets språkpolitiska program inleddes under 1970-talet arbetet med framtagande av så kallade *tröskelnivåer* som skulle beskriva och utgå från de uttrycksbehov en språkanvändare ställs inför (van Ek & Trim, 1998). Dessa tröskelnivåer påverkade språkkursplanerna i Lgr80, i vilka utveckling av kommunikativ språklig kompetens kom att prägla skrivningarna (Andered, 2001; Malmberg, 2000). Europarådets arbete under 1980-talet med implementering av *the communicative approach* influerade språkkursplanerna i Lpo 94/Lpf 94 ytterligare och i *Ämnets karaktär och uppbyggnad* betonas helheten och allsidigheten i språkanvändningen (*holistic approach*) samt interkulturell förståelse. Dessa kursplaner genomgick en revidering 2000, vilket motiverades bland annat med att ökad internationalisering och utveckling inom informationsteknologin ansågs påverka arbetet i skolan (Andered, 2001). Vid denna revidering infördes ett system med sju steg som spänner över grund- och gymnasieskolan, och med tydlig inspiration av Europarådets stegmodell (Skolverket, 2009a/GERS).

Under 2011 implementeras åter nya kurs- och ämnesplaner, en ny betygskala i sex steg i grund- och gymnasieskolan, samt betyg från och med årskurs sex. Ett viktigt syfte med de nya kurs- och ämnesplanerna är att genom en ökad tydlighet uppnå större likvärdighet över utbildningssystemet. Tidigare skrivningar kring *Ämnets karaktär och uppbyggnad* har ersatts med *Centralt innehåll* för varje ämne, mot bakgrund av vilket de kunskapskrav som anges för betygsstegen skall tolkas (Skolverket, 2011c, 2011d).

För språkens del innebär revideringen en ytterligare knytning till den gemensamma referensramen och Europarådets språkpolicy och arbetet föregicks av omfattande jämförande studier, på uppdrag av Skolverket, med syfte att belysa överensstämmelsen mellan skrivningarna i GERS och i Kursplaner 2000 (Börjesson, 2009; Erickson, 2009c). Områden som får förstärkt betoning i de nya skrivningarna är till exempel interkulturalitet och behovet av att utveckla strategier för att kunna möta de situationer språkanvändare ställs inför. Även internationellt pågår ett brett implementeringsarbete av den gemensamma referensramen.

4 Bedömning av språklig kompetens

I följande avsnitt redovisas olika bedömningskulturer, dagens vidgade syn på bedömning samt olika typer av bedömning av språklig kompetens. Avsnittet avslutas med en översikt avseende aspekter på bedömningens användning och konsekvenser, samt etiska och motivationsrelaterade faktorer

4.1 Teoretiska utgångspunkter

Synen på bedömning och utformningen av bedömningsmaterial i språk har under senare delen av nittonhundratalet sina rötter i olika bedömningskulturer. Lindblad (1990) och även Weir (2005), utifrån ett brittiskt perspektiv, beskriver tre trender inom språktestning, med utgångspunkt i Spolsky (1978).

Fyrtio- och femtiotalet beskrivs som det första, *förvetenskapliga* skedet och kännetecknades enligt Lindblad av en oskuldfull inställning till bedömningens reliabilitet och validitet. Därefter följde, under sextio- och sjuttioalet, en *strukturellistisk, psykometrisk* period. Synen på bedömning präglades då av strukturalistisk språkvetenskap och behavioristisk psykologi, med statistiska mätningar och bedömningens reliabilitet i centrum (Lindblad, 1990; Weir, 2005). En syntetisk språkfärdighetssyn strävade efter renodlade test i renodlade delfärdigheter och flervalsuppgifter som bedömdes objektivt med rätt eller fel, (*discrete point*). Åttiotalet i svensk bedömartradition präglades av den *psykolingvistisk- sociolingvistiska*, eller postmoderna perioden. Lindblad beskriver hur man talade om språkfärdighet som begrepp och funktioner, inspirerade av Europarådets tröskelnivåer och betonade funktionella, syftesinriktade aspekter av språk. Synen på språkfärdighet blev analytisk, deskriptiv och acceptabilitet blev ett centralt begrepp.

Weir (2005) illustrerar motsvarande perioder utifrån ett brittiskt perspektiv genom att jämföra hur prov- och bedömningskulturen förändrats vid Cambridge sedan *Cambridge Proficiency Examinations* först gavs 1913 och framåt. Weir refererar också till Morrow (1979) som något ironiskt betecknat perioderna ovan *Garden of Eden*, *Vale of Tears* respektive *Promised Land*.

Under nittio- och tjugohundratalet har det kommunikativa paradigmet förstärkts. En handlingorienterad språksyn medför en strävan att kunna dra slutsatser om individens kommunikativa kompetens, vilket förutsätter att också bedömningen i sig präglas av ett kommunikativt synsätt (Skolverket, 2009a/GERS). Denna grundsyn har alltsedan åttiotalet präglat de svenska nationella provmaterialen i engelska och övriga moderna språk.

Kommunikativ språklig kompetens beskrivs, som tidigare nämnts, generellt som sammansatt av lingvistisk, sociolingvistisk och pragmatisk kompetens (Skolverket, 2009a/GERS), och kommunikativ bedömning kännetecknas enligt Davies (1985) av att den är integrativ, direkt och målrelaterad (se nedan 3.3.3). Detta koncept utvecklades ytterligare av Bachman (1990) som beskrev innehållet i ett kommunikativt språkprov som *motivating*, *substantive* (oavhängigt), *integrated* och *interactive*. Bachman anger fyra kriterier för kommunikativ bedömning:

- 1) grammatical discourse, sociolinguistic, and illocutionary as well as strategic competence
- 2) pragmatic – for genuine communication, relate to thoughts and feelings, authentic language to use in a context
- 3) direct as opposed to indirect for content validity
- 4) the learner in a variety of language functions

Ytterligare en definition ges av Douglas (2010) som kontrasterar det kommunikativa paradigmet och det strukturalistiska. Enligt Douglas fokuserade det senare analys av språkliga komponenter utan kontext eller användningsaspekter, och han karakteriserar kommunikativ bedömning som: "assessing the ability to use language for communication in specific contexts, involving productive language either through meaningful input for the test taker to comprehend or interpret, or as a meaningful output generated by the test taker" (s. 69).

4.2 En vidgad syn på bedömning

Gemensam europeisk referensram för språk (Skolverket, 2009a/GERS) liksom åtskillig litteratur i ämnet, som till exempel Hughes (2003), Erickson (2006, 2009a), Douglas (2010) beskriver och diskuterar olika sorters bedömningsformer och provtyper, som har olika syften och delvis olika utgångspunkter.

Under senare årtionden har bedömningsforskning, såväl internationellt som nationellt, fokuserat värdet av en vidare syn på kontinuerlig bedömning för att stödja och utveckla lärande. Synen på bedömningens didaktiska funktion och dess validitet har blivit mer användningscentrerad (Erickson, 2006) och ses i allt större utsträckning som "dels en självklar del av lärande och undervisning, dels som en gemensam angelägenhet för lärare och elever" (Erickson, 2009a). Erickson sammanfattar i tre punkter:

Kunskapsbedömning som didaktiskt redskap [...] kräver förmåga hos aktörerna att

- se det som ska ses,
- kommunicera det som ses och
- finna bra sätt att gå vidare på basis av vad man ser.

När det gäller synen på kunskapsbedömning i ett didaktiskt perspektiv och *varför* en bedömning görs, är en traditionell kontrast den mellan summativ och formativ bedömning, dvs. mellan bedömning av produkt respektive av (och i) process (Douglas, 2010; Hughes, 2003; Skolverket, 2009a/GERS). Man talar om bedömning såväl *av* som *för* lärande (Assessment Reform Group, 2002) och alternativ som till exempel självbedömning, kamratbedömning och portfolio betonas alltmer (Douglas, 2010; Oscarson, 1999; Skolverket, 2009a/GERS). På Europarådets initiativ har *European Language Portfolio* (Little, 2005; Council of Europe) tagits fram som ett verktyg för elevens medvetenhet om sitt lärande och sin kompetens inom flera språk.

Gipps (1994) gör gällande att bedömningskulturen i språkforskningen generellt befinner sig i ett paradigmskifte och beskriver en förskjutning "*beyond testing*" från psykometri mot en mer varierad bedömningskultur i ett sociokulturellt sammanhang, det hon kallar *educational assessment* och *classroom assessment*. Flera forskare menar emellertid, i linje med en vidgad

syn på bedömning, att formativa och summativa bedömningsformer snarare bör ses som komplementära än varandras motsatser (Erickson, 2006, 2009a; Stobart, 2006; Taras, 2005). Såväl Douglas (2010) som Erickson (2009a) resonerar kring olika typer av bedömningar och mätningar som ett kontinuum mellan utvärderingar och test, och varnar också för alltför terminologiska distinktioner på bekostnad av bedömningens validitet i relation till dess syfte. Taras (2005) menar att all meningsfull formativ bedömning måste utgå ifrån en summativ bedömning (*judgement*) av elevens kompetens i förhållande till målen för det som ska bedömas och skriver:

SA [Summative Assessment] has been blamed for many problems related to assessment in our education system and as an obstacle to the growth of FA [Formative Assessment]. Until the centrality and indeed neutrality of SA is acknowledged, the real blossoming of FA will not and can not occur (s. 476).

4.3 Olika typer av bedömning

Det finns således en stor mängd olika bedömningsformer med olika syften. Skillnaderna kan beskrivas med hjälp av grundläggande frågor avseende olika aspekter av bedömningen som betonas i litteraturen, nämligen *vad*, *hur*, *varför* och *vem* (t.ex. Bachman & Palmer, 2010; Erickson, 2009a; Weir, 2005). Svaren på dessa frågor har betydelse för bedömningens validitet.

Avseende *vem* (agenten) som utför bedömningen kan främst följande bedömningstyper urskiljas: själv-, kamrat-, lärar- och extern bedömning. Vikten av att bedömningen görs av olika agenter och även i samarbete mellan dessa, så kallad sambedömning, framhålls allt mer och ett kollaborativt och reciprokt förhållningssätt betonas som väsentligt för god bedömning (Erickson, 2009b).

4.3.1 Vad som avses bedömas

Vad bedömningen avser, det som faktiskt provas, eller eventuellt bör provas, benämns *the construct*. I vid mening när det gäller bedömning av språkfärdighet innebär detta alltså vad det innebär att kunna ett språk enligt rådande språksyn. Att man faktiskt "ser det som ska ses", att man tittar på rätt sätt på rätt saker, är avgörande för bedömningens validitet och förutsätter alltså att prövningen i olika avseenden behandlar och håller sig så nära avsett *construct* som möjligt (Erickson, 2009a, 2010).

När det gäller definitionen av det som bedömningen *relaterar till*, och som därmed definierar det som avses provas, *the construct*, skiljer man vanligen mellan individ-, norm-, och, som till exempel i GERS stegmodell och det nuvarande svenska betygssystemet, mål- och kriterierelaterad bedömning. Dessa har olika utgångspunkter och diskuteras ofta i litteraturen. Hughes (2003) menar till exempel att kriterierelaterade prov sannolikt är att föredra då de, som han ser det, i större utsträckning än andra typer kan ge testtagaren användbar feedback.

Avseende *vad*, vilket *construct*, olika typer av bedömningsmaterial avser pröva, skiljer man vanligen mellan *achievement tests*, som ur ett internt perspektiv prövar ett givet innehåll, och *proficiency tests*, som avser pröva testtagarens allmänna språkfärdighet ur ett externt perspektiv, oavsett hur han eller hon har tillägnat sig den, och i förhållande till en bild av

omvärldens krav. En liknande distinktion kan också göras avseende vad som provas, men också hur det provas, mellan så kallade kunskapsprov (*knowledge tests*) och färdighetstest (*performance tests*). Kunskapsprov avser *indirekt* pröva underliggande förmågor som indikerar en viss kompetens, medan färdighetsprov har som ambition att *direkt* pröva en förmåga i sitt sammanhang, så autentiskt som möjligt (Hughes, 2003; Skolverket, 2009a/GERS). Hughes nämner också att vissa prov betraktas som *semi-direct*, som till exempel vid simulerade samtal med inspelade stimuli.

En annan skillnad gäller *rubric*, det vill säga *hur* bedömningsmaterialet utformas (Chapelle & Douglas, 2006; OECD, 1999). Man skiljer till exempel mellan *discrete point testing*, uppgifter som avser låta varje uppgift pröva en specifik kunskap eller förmåga, och *integrative testing*, integrerade uppgifter. *Discrete point testing* förknippas ofta med dikotom bedömning, som vid flervalsformat som rättas med rätt/fel, medan integrerad bedömning innebär att provtagaren visar flera, integrerade förmågor eller typer av kunskaper i en uppgift, som till exempel vid textskrivning (Hughes, 2003). Enligt Douglas (2010) utmärks dikotoma uppgifter ofta av en svag relation till situation eller syfte, medan integrerade uppgifter förutsätter att provtagaren processar flera språkliga aspekter samtidigt för att komma fram till ett svar eller en lösning, antagande att helheten är mer än summan av delarna. Douglas gör ytterligare en distinktion mellan *integrative* och *integrated tasks*, vilka ännu tydligare prövar flera färdigheter i ett sammanhang, som till exempel när läs- och hörförståelse följs av en skrivuppgift på samma tema. Han menar vidare att varierande uppgiftstyper fyller olika funktioner och med fördel kan förekomma i ett kontinuum, inom samma prov.

4.3.2 Olika sätt att genomföra bedömning

När det gäller *Hur*, tillvägagångssättet för att göra en bedömning och hur resultatet förmedlas förekommer å ena sidan holistisk bedömning, som ger en global syntetisk bedömning på en vertikal skala och å andra sidan analytisk bedömning. Den senare görs ofta med hjälp av bedömningsfaktorer och ibland med så kallade matriser, som resulterar i en horisontell bedömning utifrån deskriptorer av delkompetenser, och som sedan kan viktas samman till en nivåbedömning (NAFS-projektet; Skolverket, 2009a/GERS).

Det finns också en mängd tekniker och olika format som används vid bedömning av olika färdigheter och som påverkar typen av bedömning. Dikotoma svarsformat (*discrete point*) i form av flerval bedöms genom så kallad *objektiv* rättning, det vill säga som rätt eller fel, medan integrerade uppgifter kräver en så kallad *subjektiv* bedömning av svarets kvalitet, enligt till exempel *benchmarks*, dvs. givna exempel. Huruvida detta gör bedömningen mer objektiv eller subjektiv i sig är däremot mer komplext, vilket till exempel diskuteras i GERS och av Hughes (2003). Det finns emellertid omfattande studier som visar god interbedömarreliabilitet vid bedömning av de svenska nationella proven i engelska för årskurs nio, avseende delproven reception med stöd av bedömningsanvisningar och skriftlig produktion med stöd av benchmarks. Studien innebar för engelskans del bedömning av 100 slumpvis utvalda delprov i vardera färdigheten. När det gäller delprovet reception visade sig bedömningen där korrelera med över .99 och för skriftlig produktion med mellan .86 och .93 (Erickson, 2009a).

4.4 Bedömningens användning och konsekvenser

Under senare tid har alltmer fokus lagts på vikten av ansvarstagande för hur en bedömning används och hur resultaten av en bedömning tolkas, samt vilka effekter och konsekvenser detta får för såväl individen som i vidare mening för det pedagogiska sammanhang och det samhälle den är en del av. En bedömnings berättigande (*Justification*) och trovärdighet är beroende av hur den används och flera forskare talar om en pedagogisk, etisk och även politisk dimension när det gäller all bedömning (Bachman & Palmer, 2010; Erickson, 2010; Shohamy, 2001; Weir, 2005).

Bachman & Palmer (2010) betonar två fundamentala axiom när det gäller användningen av bedömningsmaterial. Det första är att provutvecklare, provanvändare och beslutsfattare måste ta ansvar (*accountability*) dels inför de individer vars förmåga skall bedömas, dels inför de beslut som fattas på basis av bedömningen. Det andra är att detta ansvarstagande förutsätter förmåga att demonstrera och svara för i vilken utsträckning användningen av en viss typ av bedömning är berättigad, genom väl underbyggd argumentation (s. 92ff).

Bedömningens användning och effekter betraktas som en del av dess validitet och benämns, med hänvisning till Messick (1989) dess konsekvensvaliditet (*consequential validity*). Termen introducerades som en aspekt av en bedömnings överordnade begreppsvaliditet (*construct validity*) och har fått allt större, om än ej oemotsagd, uppmärksamhet i bedömningssammanhang. Messick framhöll emellertid att validitetsbegreppet i grunden bör betraktas som odelat och enhetligt samt att, i enlighet med detta, validitet är intimt förknippat med etiska aspekter:

[...] although there are different sources and mixes of evidence for supporting score-based inferences, validity is a unitary concept. Validity always refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores (s. 13).

En analys av i vilken utsträckning en bedömning är 'riktigt' gjord och bedömer 'rätt' saker benämns validering (*validation*). Bachman & Palmer (2010) framhåller och föredrar emellertid begreppet *justification*. De anser att validering kommit att betyda insamlande av bevis för att stödja förhållandet mellan bedömningens utfall och tolkningen av detsamma. *Justification* avser däremot i högre grad betona att det som undersöks i en sådan process inte bara är tolkningen av bedömningen, utan i vilken utsträckning den avsedda användningen av bedömningen är berättigad. Som ett verktyg för att kunna avgöra i vilken utsträckning en bedömning skall kunna anses relevant och för att man ska kunna rättfärdiga dess avsedda användning sätter Bachman & Palmer (2010) upp fyra principer som bygger på Toulmin (Toulmin, 2003) och benämns *Assessment Use Arguments (AUA)*:

1. Bedömningens konsekvenser ska vara av godo (*beneficial*) för testtagarna.
2. De beslut som tas på basis av resultaten/tolkningen av bedömningen ska beakta samhällets lagar och värderingar (*values*) samt vara opartiska/rättvisa (*equitable*) gentemot testtagarna.
3. Tolkningen av vad (*the ability*) som blir bedömt ska vara meningsfull (meaningful) i förhållande till rådande styrdokument och aktuell forskning; opartisk (*impartial*) i förhållande till olika grupper av testtagare; generaliserbar avseende den målspråksdomän man avser pröva samt relevant och tillräcklig (*sufficient*) för att kunna ta beslut utifrån resultatet.

4. Resultaten ska vara stabila och konsistenta avseende uppgifters inbördes mättekniska egenskaper, bedömningens genomförande och olika grupper av testtagare (s. 103).

4.5 Etiska aspekter — Good Practice

Med anledning av ett ökat fokus på etiska aspekter av bedömning avseende dess berättigande, trovärdighet och effekter, dvs. olika typer av validitetsaspekter, ökar också intresset för och behovet av praxisbeskrivningar och riktlinjer avseende hur bedömning bör hanteras, varför begrepp som 'god bedömning' och '*good practice*' har stor betydelse.

Flera forskare manar, utifrån ett etiskt perspektiv, till generell försiktighet när det gäller tolkning och användning av olika bedömningar och de resultat de ger. Bachman & Palmer (2010) varnar till exempel för missuppfattningar kring, samt övertro och orealistiska förväntningar på språkprov som kan skapa affektiva hinder för användare. Douglas (2010) definierar ett språkprov som ett mätinstrument för att mäta språklig kompetens och framhåller att detta naturligtvis inte är helt oproblematiskt. Genom att göra kontinuerliga mätningar, arbeta med att minimera osäkra variabler och analysera resultaten på ett lämpligt och rättvist sätt, kan rimliga uppskattningar ändå göras. Douglas konstaterar emellertid att "In reality we can not measure language ability at all, we can only observe and measure performance, and on the basis of the performance of our test takers make inferences about their language ability" (s. 10). Douglas (2010) citerar vidare den etiska kod som *The International Language Testing Association* (ILTA) föreskriver och vars första princip anger tonen för det etiska förhållningssättet:

Language testers shall have respect for the humanity and dignity of each of their test takers. They shall provide them with the best possible professional consideration and shall respect all persons' needs, values and cultures in the provision of their language testing service (ILTA – Code of Ethics, 2000).

Ett annat exempel på brett internationellt samarbete är organisationen *European Association for Language Testing and Assessment — EALTA*, som vänder sig såväl till lärare och lärarutbildare som till provutvecklare och forskare inom storskalig bedömning. Inom EALTA har medlemmarna enats kring en praxisbeskrivning för god bedömning, *Guidelines for Good Practice* (EALTA, 2006) som finns översatt till 35 språk och uppmärksammas alltmer såväl i forskning som i lärarutbildning och undervisning i språk, i och utanför Europa. EALTAs riktlinjer för god praxis slår fast ett antal allmänna principer som föreskriver respekt för de studerande/provtagarna, ansvar, rättvisa, validitet och samverkan mellan berörda parter. De rekommenderar vidare dem som ska genomföra bedömning, eller utveckla bedömningsmaterial, ett antal överväganden avseende bland annat bedömningens syfte, relevans, metod, kvalitetskontroll och konsekvens.

Även för datorbaserad och Internetdistribuerad bedömning finns specifika etiska riktlinjer, framtagna till exempel av APA (*American Psychological Association*) och ITC (*The international Test Commission*). Dessa fokuserar fyra nyckelområden: teknologiska aspekter, kvalitetsaspekter, kontroll- och säkerhetsaspekter, vilka i sin tur bryts ner i underrubriker (PsycINFO Database Record (c) 2010 APA).

5 Nationella prov- och bedömningsmaterial i engelska

I följande avsnitt ges en översikt över svenska nationella provmaterial samt ett resonemang kring motivationsaspekter och Test Taker Feedback, dvs. den elevrespons på provmaterial som är en viktig del av underlaget för provutveckling inom Projektet NAFS. Några i sammanhanget relevanta studier redovisas också, främst då den här aktuella empiriska studien bygger på samma tradition.

5.1 Historik och inramning

Sverige har en lång tradition med nationella provmaterial, främst i de så kallade kärn- eller basämnen svenska, matematik och engelska. Under senare år har emellertid flera ämnen tillkommit, framför allt för grundskolan, och det nationella provsystemet utökats (*Nationella prov*).

Fram till läroplanerna 1994, Lpo/Lpf94, gavs standardprov och centrala prov inom ramen för ett normrelaterat betygssystem. I samband med nya styrdokument gav Utbildningsdepartementet Skolverket i uppdrag att utvidga det nationella provsystemet och övergå till det mål- och kriterierelaterade betygssystemet (Erickson & Börjesson, 2001; Erickson & Nihlén, 1994). Syftet var bland annat att tolka och tydliggöra målen i kursplanerna samt bidra till likvärdig bedömning över landet, genom att låta proven stödja lärarens bedömning i relation till de centralt fastställda målen. Obligatoriska prov i engelska infördes för årskurs nio 1998 och för de två första kurserna i gymnasieskolan år 2000. Ämnesproven för åk 5, som introducerats 1996, övergick 2009 från erbjudande till obligatorium.

Under 2011 införs en ny skollagstiftning samt nya kurs- och ämnesplaner för grund- och gymnasieskolan och med en ny betygsskala från F till A. För de nationella proven konkretiseras syftestexten framför allt när det gäller likvärdighets- och utvärderingsaspekter (Skolverket, *Nationella prov*). Vidare förstärks ytterligare den handlingorienterade språksynen och relationen till Europarådets språkpolicy som beskrivs i GERS i de nationella styrdokumenterna för språk, vilket uttrycks i Kommentarmaterialet till kursplanen i engelska enligt följande (Skolverket, 2011a, s. 6):

- Grundskolans nya kursplan ingår liksom tidigare, tillsammans med gymnasieskolan, i ett gemensamt system med generella och påbyggbara språknivåer, så kallade steg. Utgångspunkten för detta system är Europarådets "Gemensam europeisk referensram för språk, lärande, undervisning och bedömning" (GERS). Detta är ett vedertaget europeiskt system med generella språknivåer.
- Den nya kursplanen i engelska har ytterligare harmoniserats gentemot den europeiska referensramen. Bland annat omfattar det centrala innehållet de texttyper och kontexter för språkanvändning som förekommer i den europeiska referensramen.

De svenska nationella prov- och bedömningsmaterialen för engelska och främmande språk är så kallade *proficiency tests* som syftar till att pröva och ge en generaliserbar bild av testtagarens allmänna språkfärdighet, oavsett hur han eller hon har tillägnat sig den. Den språkliga

kompetensen delas in i delfärdigheter som avspeglas i olika delprov som fokuserar muntlig respektive skriftlig produktion och interaktion, samt receptiv förmåga avseende lyssna och läsa. Denna indelning i fyra delfärdigheter återfinns också i GERS. Resultaten på samtliga delprov aggregeras sedan med hjälp av en sammanvägningsmodell som gör att de fyra färdigheterna får lika stor tyngd i det sammanlagda provbetyget. Samtliga prov- och bedömningsmaterial inom NAFS vilar på ett antal gemensamma principer (*NAFS-projektet, Gemensamma principer*) som har sin grund i såväl nationella styrdokument som aktuell nationell och internationell forskning (Erickson, 1999, 2006; Erickson & Börjesson, 2001).

Jämförande studier indikerar att de svenska provmaterialen i engelska väl kalibrerar i förhållande till GERS stegmodell när det gäller avsedda nivåer. En studie innebar att tolv experter på referensramen från tolv olika länder oberoende av varandra ombads göra en bedömning, utifrån GERS, av svårighetsgrad och domäntäckning avseende det engelska ämnesprovet för årskurs nio (Erickson, 2009c, 2011) och en annan handlade, som tidigare nämnts, om förberedelsearbete inför kursplanerevideringen 2011 (Börjesson, 2009).

5.2 Motivation och Test Taker Feedback (TTF)

Om bedömningen också skall kunna stödja och stimulera lärande är användarens attityd till och upplevelse av provningen viktig. Uttryck för ansträngning och tillfredsställelse (*effort, cost, pleasantness, novelty*) samt för oro (*anxiety, fear of failure*) när det gäller såväl inre som yttre motivation lyfts fram av Dörnyei (2001) och Giota (2002) som betydande när det gäller elevers (och lärares) motivation. I samband med likvärdighetsperspektiv, genomförbarhet och praktikalitet (Hughes, 2003) nämns också provanvändarnas kapacitetsupplevelse och förväntan (*self-efficacy, expectancy-value, attribution*), vilka kan upplevas påverka genomförbarheten eller de egna möjligheterna (Dörnyei, 2001; Eklöf, 2006).

Giota (2002) anför att de elever som upplever skolan positivt får ett slags positiv vaccination att ta med sig ut i vuxenlivet. Utifrån ett interaktionistiskt perspektiv definierar hon motivation som ett flerdimensionellt begrepp som innefattar kognitiva, sociala och emotionella mål samtidigt. Självbild och självkänsla skapas av hur eleven blir bedömd och bemött i skolan och huruvida man litar på sin förmåga påverkar faktorer som ansträngning, uthållighet, engagemang och ängslan, samt de val man gör senare i livet: "Många prestationer handlar inte i första hand om att kunna utföra en handling utan att vilja orka genomföra den" (s. 286).

Weir (2005) skriver:

A test taker's interest or motivation may affect the way a task is dealt with. Affect can help or hinder performance. Given that we wish in most cases to 'test for best', steps must be taken to make test events as positive as we can in the full knowledge that some stress is perhaps unavoidable (s.53).

Weir menar att provtagarens förutsättningar är det första som måste beaktas för att kunna uppnå en rättvis bedömning. Han diskuterar i termer av olika typer av anpassningar, affektiva aspekter och familiaritet med frågeformat eller medium, som t.ex. datorvana vid CBT. Alderson (2000b) lyfter fram liknande aspekter i sina studier och menar att uppfattad oro inför en uppgift

(*state anxiety*) tydligt påverkar elevens inre motivation och har visat sig sänka resultaten vid prov i läsning.

Test Taker Feedback (TTF)

Att aktiva lärare och inte minst elever, är ovärderliga samarbetspartners i kvalitetsarbetet kring såväl processer som produkter när det gäller utveckling av provmaterial, bland annat när det gäller motivationsaspekter, är väl belagt i litteraturen (Bachman & Palmer, 1996, 2010; Dörnyei, 2001; Erickson, 1998, 2006; Erickson & Gustafsson, 2005) och en hörnsten för arbetet inom NAFS-projektet. Synpunkter har samlats in, som en del av valideringsprocessen och med hänsyn till etiska överväganden, sedan mitten av 1990-talet i samband med storskaliga utprovningar. Elever, och lärare, ger vid utprovningarna respons på enskilda uppgifter dels genom öppna kommentarer och dels genom att markera bland annat uppfattad svårighetsgrad, relevans och egen prestation på femgradiga så kallade Likert-skalor (Erickson, 1998). Dessa synpunkter beaktas vid kompositionen av prov men har också direkt påverkan på konstruktion av olika uppgifter under processen.

Ett flertal studier inom NAFS-projektet har direkt eller indirekt involverat elevers uppfattningar om de nationella proven i engelska. Analyser och sammanställningar av TTF (Erickson, 1999, 2006) visar att elever i grundskolan mest uppskattar uppgifter som ger dem möjlighet att kommunicera utifrån ett givet ämne, som i de produktiva delproven. De är också positiva till uppgifter som verkar autentiska och pedagogiska när det gäller innehåll och metod, samt är lätta att förstå, trevliga och utmanande att göra samt upplevs rimliga i svårighet. De negativa kommentarerna handlar ofta om att uppgifter upplevs som onyttiga, svåra, stressiga eller alltför triviala. Analyser av korrelationer mellan elevernas uppfattningar visar att en positiv inställning till en uppgift betydligt starkare korrelerar med uppfattning om en uppgifts nytta och relevans än med upplevelsen av uppgiften som lätt eller svår.

En deskriptiv redovisning av en större enkätstudie från Skolverket avseende lärares och elevers uppfattningar om proven i gymnasieskolan (Skolverket, 2005) visar överlag att acceptansen för, framför allt de engelska, nationella proven är god. Respondenterna upplever både proven i sig som positiva och motiverande, samt betraktar deras roll att stödja en likvärdig bedömning som viktig. Resultaten i denna Skolverksstudie bekräftas också av en mindre intervjuundersökning med tio gymnasieelever (Åhs, 2005). Studier har även gjorts avseende ämnesproven i engelska för årskurs fem. Dessa visar att flertalet elever tycker att det är roligt och värdefullt att få visa vad de kan och att även så unga elever har värdefulla insikter om bedömning (Velling Pedersen, 2004).

Inom ramen för ett större nätverksprojekt (*European Network for Language Testing and Assessment*) har även större europeiska enkätstudier med fokus på elevers, och lärares, uppfattningar om bedömning och prövning av engelsk språkfärdighet gjorts. Analysen av cirka 1400 elevers enkätsvar från tio europeiska länder indikerar här att god bedömning, enligt eleverna, karakteriseras av domäntäckning och variation, kommunikativ nytta, möjlighet att lära sig något, samt klarhet, tydlighet, rättvisa och tillräckligt med tid (Erickson & Gustafsson, 2005). Liknande resultat rapporteras också av Shohamy (2001) som funnit att grunden till testtagares acceptans av provmaterial är att dessa upplevs realistiska, erbjuder ett inläringstillfälle samt är trevliga och inte stressande att göra.

Elevens åsikter om vad de upplever som positivt respektive negativt när det gäller språkprov stämmer väl överens med gängse definitioner av validitet. Erickson (2010) visar med autentiska exempel från en internationell studie hur elever, i konsekvens med Messick (1989), ofta uttrycker att "dåliga" prov är för ensidiga och inte ger dem utrymme att visa sin förmåga (dvs. ev. *construct under-representation*), eller att deras prestation hindrats av otydligheter, tidsbrist och liknande (dvs. ev. *construct-irrelevant variance*).

5.3 Ett nytt steg för de svenska nationella proven i språk?

En långsiktig och övergripande avsikt med denna uppsats är, som tidigare nämnts, att kunna bidra till utvecklingen av svenska nationella bedömnings- och provmaterial i språk, framför allt med avseende på eventuell datorbaserad. Den här aktuella empiriska studien genomfördes också initialt inom ramen för ett större uppdrag av Skolverket, inom språk och matematik, att skapa underlag för framtida diskussioner i frågan (Skolverket/ Nyström, 2010).

När det gäller NAFS-projektet tillhandahålls en mängd provmaterial i främmande språk i en så kallad Provbank och formativa material, liksom en mängd exempel på uppgiftstyper erbjuds via hemsidan (NAFS-projektet), men inga av dessa uppgifter är i sig digitala. Det har dock funnits planer från Skolverkets sida att datorbasera provmaterial i engelska för de yngre åldrarna samt skapa en digital bedömningsplattform för provutveckling, men detta har av olika skäl inte realiserats. Inom NAFS-projektet pågår en mindre studie avseende effekter relaterade till skrivande på dator vid skriftlig produktion i engelska för gymnasieskolan, vilket enkäter visar är relativt vanligt ute på skolorna. I dagsläget är det, som tidigare nämnts, emellertid endast de nationella proven i svenska för invandrare (Skolverket. *Sfi - Svenskundervisning för invandrare.*) som i Sverige erbjuds i såväl pappers- som digitaliserad form från och med 2010.

Innan nästa steg tas inom NAFS-projektet mot mer digitaliserade bedömningsformer och eventuella datorbaserade nationella provdelar i språk, föreföll det väsentligt att studera aktuell forskning inom området datorbaserad språkbedömning samt ta del av andra länders erfarenheter. I ljuset av detta samt som komplement till den teoretiska bakgrund som givits i avsnitt 3-5 ovan, följer därför i avsnitt 6, en sådan genomgång. Dessa avsnitt avser vetenskapligt förankra, skapa bakgrund till och ytterligare belysa resultatet av den efterföljande empiriska studien genom att ge underlag för att kunna svara på uppsatsens första forskningsfråga: *Vilket är det aktuella forskningsläget avseende relationen mellan construct/'Vad' och rubric/'Hur', samt effekter av detta, när det gäller bedömning av språklig kompetens via dator?*

6 Datorbaserad av bedömningsmaterial i språk

Detta avsnitt behandlar datorbaserad bedömning av språklig kompetens under de senare decennierna enligt aktuell forskningslitteratur. Först ges en bakgrund kring utveckling, incitament och förhållningssätt avseende datorbaserad av bedömningsmaterial i språk. Därefter belyses ett antal validitetsaspekter, utifrån kriterier föreslagna av Bachman & Palmer (1996, 2010), med fokus på aspekter rörande HUR?, *rubric*, och VAD?, *construct*, då dessa betraktas som avgörande när det gäller datorbaserad och anses få effekter för bedömningen. Vidare redovisas påverkans effekter, erfarenheter av implementering och likvärdighetsaspekter, också betonade av Bachman & Palmer som viktiga för en bedömnings rättfärdigande. Relativt genomgående används, främst av utrymmesskäl, förkortningen CBT (*Computer Based Testing*) för datorbaserad bedömning.

6.1 Utveckling av datorbaserad bedömning i språk

I vår tid blir olika former av digital testning allt vanligare. Chapelle & Douglas (2006) diskuterar huruvida datoranvändning börjar bli så nära förknippad med språkanvändning att själva språkbegreppet, *the construct*, står inför en förändring, och de menar att det snarare är fråga om evolution än revolution att betrakta datorisering som naturlig i språkbedömning i framtiden. Det finns emellertid, enligt Chapelle & Douglas, förvånande lite forskning kring effekterna av CBT, med tanke på att sådan förekommit i cirka tjugo år, och de konstaterar att:

Computer technology may in the future radically change research and practice in language assessment but doing so will require the type of research that engages with the complexity of the issues, crossing the boundaries between assessment, language, and technology for the purpose of developing paths that work toward the goals of applied linguists (s. 117).

Det finns emellertid redan en rik flora av digitala bedömningsinstrument för språksammanhang som till exempel DIALANG (*Diagnosis of Language*) och TOEFL (*Test of English as a Foreign Language*), (Alderson, 2000b; Huhta m.fl., 2002; Moe, Carlsen & Hasselgren, 2006) och även test av *high-stakes*-karaktär som t.ex. *placement-tests*. Under hösten 2008 gjorde projektgruppen, inom vilken den i uppsatsen aktuella empiriska studien genomfördes, bland annat flera studiebesök i andra europeiska länder med erfarenhet av storskalig datorbaserad testning. Det bedrivs till exempel forskning kring och utveckling av storskalig datorbaserad, adaptiv testning (CAT) vid University of Durham i Storbritannien samt vid Cito, *Institute for Educational Measurement* i Holland (Skolverket/Nyström, 2010). I Danmark har man försökt få till stånd CAT för ungdomsskolan i flera ämnen (Wandall, 2009), Finland har ett flertal datorbaserade provmaterial och Universitetet i Bergen har sedan 2002 lett arbetet med att utveckla digitala nationella prov i engelska för grundskolan i Norge (Moe, 2009; Moe, Carlsen & Hasselgren, 2006).

Historiskt sett har dock CBT sin vagg i USA. Upprinnelsen var enligt Fulcher (2000) att man konstruerade de lämplighetstester, så kallade Alpha-tester, som gavs inom den amerikanska armén under första världskriget, i ett för tiden modernt provformat med flervalfrågor. Denna typ av flervalstest började också något senare användas storskaligt i utbildningssammanhang. För att effektivisera det omfattande rättningsarbetet designades en maskin speciellt för ändamålet, en så kallad *Test Scoring Machine*, som var mer exakt och tio gånger snabbare än

manuell rättning. Denna IBM, modell 805, (*International Business Machines*) blev tillgänglig på marknaden 1935 och betraktas som datorns föregångare inom språktestning. Provtypen var effektiv att producera och administrera, och automatisk, maskinell rättning bidrog ytterligare till att flervälsformat än idag dominerar i storskalig testning (Chapelle & Douglas, 2006; Fulcher, 2000). Utvecklingen av datorteknologin har sedan gjort att stordatorer (*Mainframe Computers*) har använts frekvent i språkbedömning sedan 60-talet, för rättning, statistisk analys och lagring. På senare tid har spridningen av persondatorer (*Personal Computers*) underlättat användning av ordbehandlingsmjukvara för till exempel testkonstruktion, distribution online och för att ge omedelbar feedback. Datorer har kommit att spela en omfattande roll i provsammanhang och används därmed i hela processen vid alltifrån design, konstruktion, distribution, rättning, statistisk analys och lagring. Fulcher (2000) menar dock att CBT fram till 2000 mest handlat om ”teknikifiering” av tillhandahållande och efterbearbetning. Han tror att vidare utveckling möter mer konceptuella än tekniska hinder och förutspår att vad som prövas (*test construct*) vid CBT kommer att dominera diskussionen och forskningen inom datorbaserad språkbedömning under början av 2000-talet.

Chapelle & Douglas (2006) konstaterar att olika traditionella språkfärdigheters påverkan av datoriserad prövning diskuteras relativt flitigt i forskningen. Man hänvisar till Buck (2001) angående hörförståelse med inslag av multimedia, Weigle (2002) avseende digital bedömning av och teknologins effekter på skriftlig produktion och Alderson (2000b) som problematiserar datorbaserad när det gäller läsning. De ger även exempel på hur man med multimedia kan iscensätta testsituationer för att virtuellt simulera autentiska situationer, och menar att: ”Taken together, the strands of the technology thread point to an important change in the fabric of language assessment: the comprehensive introduction of technology” (Chapelle & Douglas, 2006, s. 1).

6.2 Incitament till att datorbasera bedömning i språk

I sina inledningsanföranden vid en konferens i Reykavíjk på Island 2009, anordnad av Europarådet och *The Joint Research Centre (JRC); The Centre for Research on Lifelong Learning (CRELL)*, redogör Scheuermann (2009) och Bjerkestrand (2009) för Europarådets satsningar när det gäller forskning om CBT och dess politiska kontext. De, i likhet med Kozma (2009) från *Intel, Microsoft, and Cisco Education Taskforce*, menar att de europeiska utbildningssystemen måste förändras för att kunna möta 2000-talets sociala och ekonomiska behov. Kozma betonar samtidigt, och framtidens, behov av IKT-kompetens (informations- och kommunikationsteknologisk kompetens), *ICT Literacy*, och ’21st century skills’ och därmed nödvändigheten av att transformera bedömning av språklig kompetens, eftersom: ”Assessment is the means by which society determines what students have learned and what they can do next” (Kozma, 2009).

Bjerkestrand (2009) beskriver hur datorbaserad i utbildnings- och testsammanhang i Europa både är ett mål och ett medel. Europaparlamentet och Europarådet utfärdade 2006, i den så kallade Lissabonstrategin, rekommendationer för nyckelkompetenser och en gemensam referensram (*The European Coherent Framework of Indicators and Benchmarks*). De nyckelkompetenser för livslångt lärande som anges är:

1. Kommunikation på modersmålet.
2. Kommunikation på främmande språk.
3. Matematiskt kunnande och grundläggande vetenskaplig och teknisk kompetens.
4. Digital kompetens.
5. Lära att lära.
6. Social och medborgerlig kompetens.
7. Initiativförmåga och företaganda.
8. Kulturell medvetenhet och kulturella uttrycksformer.

Bjerkestrand redogör vidare för hur JRC och CRELL, med stöd från Europarådet, genomför en mängd helt eller delvis datorbaserade forskningsprojekt med anknytning till dessa nyckelkompetenser, som till exempel *European Survey on Language Competences 2011*, som redovisas 2012, och PISA (*OECD Programme for International Student Assessment*), som planeras vara helt datorbaserat 2015. Bjerkestrand betonar också att datorbaserad skulle underlätta för alla, från huvudmän till testtagare, eftersom vi står inför en allt ökande mängd datainsamlingar i form av europeiska studier och mätningar.

6.3 Förhållningssätt till datorbaserad av bedömning i språk

Sedan 1980-talet har frågan om datorbaserad i språksammanhang diskuterats frekvent inom bedömningsforskningen. Chapelle & Douglas (2006) identifierar tre olika förhållningssätt till och antaganden om teknologins roll vid språkprovning (s. 116), som de menar har stor betydelse för utvecklingen i relation till CBT. Nedan används dessa som rubriker för att spegla ett antal forskares syn, och eventuellt kan de även avspegla en viss kronologi i hur diskussionen har böljat i takt med teknikens utveckling, olika intressenters intåg på marknaden och en förändrad syn på språkanvändning.

1) Tunnelseende –"It is an efficiency."

Det första antagandet om teknologins roll vid språkprovning karakteriseras av Chapelle & Douglas (2006) som en form av tunnelseende och syn på teknifiering som en ren effektivisering. Det gäller till exempel användning av stordatorer för distribution och automatisk rättning (*scoring*) varvid utgångspunkten blir att på ett snabbare, billigare och effektivare sätt 'göra samma'. Även möjligheter att göra prov när och var som helst och omedelbart få ett resultat, samt tekniska möjligheter att få inblick i testprocessen (Alderson, 2000a) kan i viss mån betraktas som effektiviseringar.

2) Jämförelse –"It should be considered suspect."

Detta andra sätt att förhålla sig utgår enligt Chapelle & Douglas från ett jämförande perspektiv. Enligt författarna präglas förhållningssättet av viss misstänksamhet och att "the non-technology condition is the normal one" (s. 116). Flera forskare problematiserar vilka effekter datorbaserad får, men ser också metodologiska och pedagogiska möjligheter med CBT jämfört med pappersprov, till exempel när det gäller multimedia (Alderson, 1990, 2000a; Douglas, 2000; Fulcher, 2000). Detta förhållningssätt benämns också (senare) i litteraturen *the Migratory approach* (t.ex. Ripley, 2009).

Alderson (1990) skyller en viss skepsis i skolorna mot datorbaserade språkmaterial på de språklaboratorier som var vanliga under 1960-talet, men ser stora och goda möjligheter till pedagogisk utveckling genom ytterligare implementering av CBT och CALL (*Computer Assisted Language Learning*). Med syftning på liknande tendenser till tröghet i utbildningssammanhang görs, nästan tjugo år senare, nedanstående jämförelse av Ripley (2009) som har erfarenhet av utveckling och implementering av CBT både i England och internationellt:

In England it took around 25 years from the introduction of calculators on a wide-scale until they were first expected to be used in school examinations. In 1994 and 1995 examination setters began to set mathematics tests which required students to use a calculator (alongside second papers which prohibited their use). Within a very few months, mathematics teachers began to teach students the skills of using a calculator. A 25-year gestation period for the calculator does not augur well for more radical innovative assessments (s. 98).

3) Innovation – "It should be considered a resource."

Ett innovativt förhållningssätt till datorbaserad ger, enligt Chapelle och Douglas (2006), anledning att tänka i nya banor. Författarna efterlyser forskning som tar in tekniken och en innovativ agenda, men betonar att "the agenda needs to be driven by the concerns of applied linguists for assessment" (s. 116). Sådan forskning bör, enligt Chapelle (2010), inte isolera effekterna av teknologi utan istället sträcka sig "beyond comparison" och fokusera på design, samt användning och effekter av specifika teknikbaserade pedagogiska praktiker. Hon menar också att synen på *construct* behöver förändras för att teknikbaserad språkinläring och språkprovning ska kunna studeras och analyseras på ett fruktbart sätt (se vidare under 6.4.3). En innovativ agenda förutsätter därmed, enligt Chapelle, att provutvecklare och forskares antaganden om teknikens roll avslöjas, reflekteras över och kanske moderniseras. Detta förhållningssätt benämns också i litteraturen *the Transformational approach* (Ripley, 2009).

6.4 Validitetsaspekter relaterade till datorbaserad bedömning

Oavsett förhållningssätt till teknikens roll när det gäller bedömning, är bedömningens kvalitet beroende av dess validitet samt, som en nödvändig delaspekt av detta, resultatens reliabilitet. Ett noggrant övervägande av dessa kvalitetsaspekter är, som tidigare nämnts, en förutsättning för att bedömningen i sig och de antaganden som görs utifrån dess resultat sedan skall kunna användas på ett riktigt och likvärdigt sätt. Det är också genomgående frågor som är relaterade till aspekter av validitetsbegreppet när det gäller CBT som litteraturen återkommer till.

Chapelle & Douglas (2006) samt Chapelle (2010) refererar i sin forskning angående CBT till sex kriterier som föreslogs av Bachman & Palmer 1996 för utvärdering av en bedömnings lämplighet och användbarhet, nämligen: reliabilitet, begreppsvaliditet (*Construct validity*), autenticitet, interaktivitet samt påverkan (*Impact*) och praktikalitet.

De första fyra av ovanstående validitetsaspekter behandlas därför i följande avsnitt (6.4), avseende vad som bör beaktas när det gäller datorbaserad provning i synnerhet. Därefter fokuseras speciellt olika påverkans-, praktikalitets- och implementeringseffekter (6.5) samt likvärdighetsaspekter (6.6) i relation till CBT, eftersom (som tidigare nämnts under 4.4) Bachman & Palmer (2010) har vidareutvecklat sin modell i det att man ytterligare betonar just

konsekvenser, likvärdighets- och användningsaspekter av bedömningen som centrala och utgår ifrån ett än mer enhetligt validitetsbegrepp för en bedömnings rättfärdigande (*justification*). Dessa aspekter förekommer även i den här aktuella empiriska studien, för att då speciellt belysa ett utsnitt av svenska provanvändares syn och förhållanden. Bachman & Palmers utvärderingsaspekter från 1996 återfinns i något ny skepnad i de krav (*claims*) som författarna 2010 menar bör ställas i olika skeden av ett bedömningsförlopp (s. 434):

1. Claim: consequences are beneficial
2. Claim: decisions are values sensitive, equitable
3. Claim: interpretations are meaningful, impartial, generalizable, relevant, sufficient
4. Claim: assessment records are consistent

6.4.1 Aspekter relaterade till rubric, reliabilitet och process

Bedömningens reliabilitet betraktas alltmer som en delaspekt av dess validitet, eftersom god validitet, bland annat i enlighet med de krav som ställs av Bachman & Palmer, 2010 (ovan samt 4.4), förutsätter att resultaten är reliabla och konsistenta. Weir (2005) väljer till exempel att relatera till aspekten reliabilitet som *scoring validity*.

Chapelle & Douglas (2006) betonar vinster i form av enhetlighet, konsistens, opartiskhet och därmed likvärdighet avseende *rubric* vid CBT. Det faktum att instruktioner, eventuella hjälpfunktioner, process, tid och ev. rättning styrs via dator som en oskiljaktig del av bedömningen innebär, enligt författarna, att mycket av den variation som vid en traditionell provningsprocess kan omgärda *rubric* undviks och medför att bedömningen blir mer reliabel. De menar också att exempelmaterial med syfte att uppnå familiaritet med format hos testtagaren kan göras enhetliga, relevanta och lättillgängliga för användaren om de distribueras som en del av den aktuella bedömningen, eller till exempel online, vilket därmed kan öka likvärdigheten.

När det gäller rättning av datorbaserade bedömningsmaterial, betonas ofta vinster i form av exakthet, objektivitet och effektivitet. Historiskt sett förutsatte användningen av stordatorer för rättning (*scoring*) flervalformat och byggde på det statistiska konceptet *reliabilitet* som utvecklades under början av 1900-talet (Chapelle & Douglas, 2006; Fulcher, 2000). Fulcher talar om en, för tiden, ny typ av kunskapstest som än idag tenderar att dominera i utbildningssammanhang. Han betonar dock distinktionen mellan reliabilitet och *scorability* och anser att flervalssuppgifter i sig inte visat sig mer reliabla än andra format, utan möjliga att lätt rätta maskinellt och därför praktiska. Flera forskare (t.ex. Hughes, 2003) framhåller emellertid fördelar ur framför allt likvärdighetsperspektiv med denna så kallade objektiva rättning, medan andra (t.ex. Lindblad, 1990) påminner om att den objektiva rättningen föregåtts av ett subjektivt förfarande vid frågekonstruktionen.

På senare tid har emellertid de tekniska möjligheterna blivit något bredare och viss bedömning av öppna format har utvecklats till exempel när det gäller skriftlig och muntlig produktion (Alderson, 2000b), som till exempel e-rater (ETS – About e-rater) och PhonePass (Chun, 2006). Trots detta ser en del forskare att begränsningar i svarsformat, liksom datorns kapacitet att rätta, kan vara ett konservativt element i CBT, som skulle kunna innebära att överväganden baserade på psykometriskt och teknologiskt tänkande tillåts överskugga andra sätt att

bedöma, eller begränsa bedömning av kommunikativa värden i språket (Alderson, 2000a; Douglas, 2000; Fulcher, 2000). Detta skulle kunna leda till att man frestas göra det lätt mätbara till det viktiga, vilket ofta betonas som en generell fara i bedömningsammanhang.

Alderson (1990, 2000a) och Douglas (2000) varnar också för att utvecklingsarbetet av i synnerhet CAT, datorbaserade adaptiva prov (se 3.1) tenderar att präglas av effektivisering, och generellt inte eftersträvar nytänkande. Alderson ifrågasätter i viss mån också validiteten för CAT och menar att den diskuterats alltför sparsamt. Hughes (2003) och Chapelle & Douglas (2006) ser emellertid i stort sett enbart vinster med CBT i form av mer konsistent och precis bedömning och de, i sin tur, pekar framför allt på fördelar med att CAT avbryter vid hög reliabilitet. Liknande problematik bekräftades i viss mån vid de studiebesök som tidigare nämnts, då det framkom att CAT ofta ses som en möjlighet till effektivisering vid storskalig bedömning, men att såväl utveckling som användning av CAT har visat sig komplicerat, till exempel i Norge och Danmark (Moe, 2009; Skolverket/Nyström, 2010; Wandall, 2009).

När det gäller explicit didaktiska aspekter av bedömning ser emellertid flera forskare ytterligare möjligheter med CBT. Den positiva utveckling som framför allt diskuteras gäller innovativa uppgiftsformat (Chapelle, 2010; Sireci & Zenisky, 2010), omedelbar feedback, hjälpfunktioner och insyn i testprocessen, vilket skulle innebära pedagogiska och didaktiska fördelar i form av mer individualiserade prov och övningsmaterial (Alderson, 1990, 2000a; Chapelle, 2010). Detta skulle också kunna få positiva effekter på hur prov betraktas och används, deras roll i styrdokument och undervisning, men också på tillämpad lingvistik, *applied linguistics*, i ett vidare perspektiv. Alderson manar dock 1990 till viss eftertanke när det gäller möjligheten till dessa komplexa resultat: "In fact the possibilities almost certainly outstrip the ability of teachers and applied linguists to understand or interpret or simply to digest the information. [...] the challenge is to determine what information one needs to gather about learners' performance, not simply what one can gather" (s.25).

6.4.2 Aspekter relaterade till autenticitet och interaktivitet

Autenticitet och interaktivitet betraktas som viktiga komponenter i kommunikativ bedömning och en förutsättning för meningsfullhet och generaliserbarhet, vilket har betydelse för bedömningens validitet (jämför ovan under 6.4). Douglas (2000, 2010) och Chapelle & Douglas (2006) återkommer ofta till datorbaseringsens kapacitet att mediera "äkta" interaktivitet på målspråket samt att simulera kontexter som är mer verklighetslika än vad som kan åstadkommas via papper-och-pennabedömning (PBT). De varnar dock samtidigt för övertro på autenticitet i samband med teknik och påminner om att mycket som har med språklig kompetens att göra inte alls innefattar datoranvändning. Douglas varnar för teknikifiering utan eftertanke och det han kallar *the pitfalls of technology*: "Language testing that is driven by technology, rather than technology being employed in the service of language testing is likely to lead us down a road best not travelled" (Douglas, 2000, s. 275).

Å ena sidan beskriver Douglas hur interaktiva multimedia kan skapa verklighetstrogna virtuella, simulerade situationer vid till exempel yrkesrelaterad prövning av LSP (*Language for Specific Purposes*) och å andra sidan varnar han för att betrakta sådana simulationer som helt autentiska (*the real domain*). Han betonar att de egentligen inte kan visa annat än just en prestation i en simulerad situation och manar därför till försiktighet när det gäller tolkningen

av resultat. I en senare bok om språkbedömning återkommer Douglas (2010) till frågan och betonar då vikten av att skilja på vad som är *genuint* och vad som är *autentiskt* och konstaterar att "genuineness travels well, but authenticity is more difficult to achieve" (s. 25).

Chapelle & Douglas (2006) skiljer också mellan olika aspekter av autenticitet, till exempel när det gäller virtuella format, och resonerar kring *situational authenticity*, som avser till exempel bedömningsituationens kontext, genre, innehåll och deltagare och *interactional authenticity*, som avser interaktivitet mellan testtagarens språkkunskaper och den kommunikativa uppgiften (s. 9).

6.4.3 Aspekter relaterade till begreppsvaliditet - construct och construct validity

Bachman & Palmer (2010) anger, som tidigare nämnts, att ett viktigt krav på valid bedömning är att den kan anses ge meningsfullt, tillräckligt och relevant underlag för tolkningar samt ha hög grad av generaliserbarhet. Begreppet validitet är komplext och knutet till hela bedömningsprocessen. Flera delfaktorer utgör tillsammans bedömningens validitet, vilket definierades av Messick, 1996: "Validity is not a property of the test or assessment as such, but rather on the meaning of the test scores" (s. 245). Som en konsekvens av detta drar Messick slutsatsen att: "Hence, validity becomes a unified concept and the unifying force is the meaningfulness or trustworthy interpretability of the test scores and their action implications, namely, construct validity" (s. 253).

Begreppsvaliditet (*construct validity*) karakteriseras således av Messick som en 'enande kraft' i sammanhållen validitet och han förklarar just begreppsvaliditet som: "The measure estimates how much of something an individual displays or possesses. The basic question (of construct validation) is, What is the nature of that something?" (s. 257).

Chapelle & Douglas (2006) definierar ett prov som "a procedure designed to elicit certain behavior from which one can make inferences about [...] certain examinee characteristics in particular non-test contexts" (s. 21). Sådana inferenser om generaliserbarhet måste således, enligt Messicks syn på validitet, vila på en analys av bedömningens begreppsvaliditet och med hänsyn tagen till de hot mot denna som Messick identifierat (se 3.1), nämligen *construct-irrelevant variance* (att bedömningen innehåller ovidkommande inslag i förhållande till det som avses prövas) och *construct under-representation* (att bedömningen innehåller för lite av det som avses prövas) — jämför ovan Bachman & Palmer (2010) om meningsfullhet, tillräcklighet och relevans.

En fråga som betonas i bedömningslitteraturen är i vilken mån det som faktiskt blir föremål för bedömning, *the construct*, förändras vid datorbaserad, och huruvida detta i så fall är önskvärt eller ett hot mot bedömningens validitet. Att *rubric*, dvs. tillvägagångssättet, metoden eller formatet, som används har betydelse för bedömningens validitet och påverkar resultatet benämns metodeffekt (*method-effect*) eller systemeffekt. Bachman betonade 1990 att valet av metod är mycket betydelsefullt:

[...] 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 languages tests thus varies as a function both of an individual's language ability and of the characteristics of the test method (s. 113).

Vad är det som provas vid CBT?

— What is the nature of that something? (Messick, 1996, s. 257)

Litteraturens syn på begreppsvaliditet, *construct validity*, avseende CBT relaterar i viss mån till de förhållningssätt som Chapelle & Douglas (2006) illustrerar, och som tidigare redovisats, nämligen "Tunnelseende", "Jämförelse" och "Innovation". Men framför allt följer diskussionen en viss kronologi som förmodligen speglar den utveckling som skett bara under 2000-talet. Nedan görs en uppdelning i tre relativt olika sätt att se på *construct* i CBT, som återfinns i litteraturen.

Oförändrat construct

Ett första antagande om CBT i relation till begreppsvaliditet är alltså att den prövar och avser pröva exakt samma *construct* som motsvarande pappersprov (PBT), dvs. att datorbaserad egentligen inte har, eller ska ha, någon betydelse för bedömningens validitet.

Intrycket från litteraturen blir dock att detta problematiserats alltmer i takt med utbredningen av CBT, och förmodligen i takt med en vidare användning än de effektiviseringar som tidigt avsågs. Om CBT avser pröva ett oförändrat *construct*, trots *rubric* i form av ett annat medium och en annan metod, måste eventuella metodeffekter av datorbaserad betraktas som ovidkommande inslag, *construct irrelevant variance*, och kontrolleras för att inte påverka validiteten negativt. För att undvika oönskade metodeffekter menar Douglas (2000) att det är en utmaning att analysera och balansera användningen av teknologi: "Part of the task of describing the TLU [Target Language Use] situation will necessarily include an analysis of the use of technology, and the concomitant use of technology in the test situation should ideally reflect that analysis" (s. 276).

I *Assessing reading* problematiserar Alderson (2000b) huruvida till exempel faktorer som textmängdens begränsning på skärm, svårighet att scrolla jämfört med att bläddra och ansträngning att läsa på skärm kan påverka möjligheten att generalisera läsförmåga utifrån provresultatet. Douglas (2000) drar liknande slutsatser och kritiserar till exempel att testtagaren i TOEFL, 1998 vid läsning måste behärska förmågan att scrolla med datormus, eftersom detta innebär prövning av ytterligare en färdighet och därmed förändrar *construct*. Såväl Alderson som Douglas resonerar å andra sidan kring att IKT förmodligen blivit en så naturlig del av läsning att det blivit en delfärdighet – *an element of literacy* (se nedan).

Trots mängden CBT redan i slutet på nittiotalet, menar Alderson (2000a) att det egentligen inte förelegat någon egentlig innovation eller förbättring när det gäller metoder eller att mäta *construct*. Detta är enligt författaren en anledning till att projektet DIALANG antagit ett försiktigt förhållningssätt till metodeffekter och främst arbetat med innovativ design och självbedömning. Han betonar att det faktum att DIALANG är *low-stakes* och diagnostiskt,

tillåter innovation och experiment och efterfrågar komparativ forskning kring teknologins effekter i vid bemärkelse.

Förändrat construct för CBT

Med sin fråga *New test methods – new constructs?* önskar Chapelle & Douglas (2006, s. 10) ta ett steg framåt och efterlyser forskning som undersöker CBTs möjligheter att pröva och synliggöra nya (språkliga) förmågor. De diskuterar i termer av olika *construct* – *the pure construct* eller *the construct of interest* respektive *the test construct*, och betonar vikten av att klargöra vad som faktiskt provas (*clarifying test construct*). Om testtagaren vid hörförståelse till exempel kan gå tillbaka och lyssna flera gånger eller vid skriftlig produktion kan använda online hjälp eller språkkontroller, då mäter provet implicit även dessa strategiska förmågor, vilket enligt författarna innebär att de måste betraktas vara del av *test construct*.

Vidare problematiserar Chapelle & Douglas det traditionella antagandet att *construct of interest*, t.ex. avseende läsförmåga, för att vara generaliserbart bör förstås som oberoende, dvs. förmåga att läsa vad som helst var som helst. De ifrågasätter mot bakgrund av detta att prövning av datorbaserad läsning skulle utöka *construct* och hävdar att *construct of interest* i så fall är för begränsat. En förändring i prestation som beror på god datorförmåga bör inte, enligt författarna, betraktas som irrelevant för allmän språklig förmåga, som den bör definieras på 2000-talet. De frågar sig 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 metodeffekter och vara ett exempel på *construct under-representation* enligt Messicks definition (1996):

In an environment where language users spend much of their time reading and writing in front of the computer, the constructs of academic reading, listening, and online composing might best be reflected in computer-assisted test tasks. So integral is the computer to the writing process that the idea of assessing writing ability with a paper-and-pencil writing task would be recognized by most academics as introducing bias into measurement (Chapelle & Douglas, 2006 s. 94).

Enligt Chapelle & Douglas (2006) och senare Chapelle (2010) speglar synen på validitet också möjligheterna med CBT. De menar att ett paradigm som förutsätter att språktest ska mäta språkliga kompetenser oberoende av kontext för språkanvändningen förmodligen är orimligt. En förutsättning för att forskningen kring validitet skall kunna bli konstruktiv och se både potential och nackdelar med metodeffekter rörande CBT, är enligt forskarna att begreppet validitet omvärderas för CBT och att andra dimensioner än korrelation mellan CBT och traditionella pappersprov (PBT) belyses: "[...] if a computer-based test results in a score meaning which is different from that of an otherwise similar paper-and-pencil test, it is a threat only to the extent that score users intend the scores to be equivalent" (s. 42).

Chapelle & Douglas ser emellertid, oavsett jämförelser med PBT, flera möjliga hot mot validiteten när det gäller CBT i sig. De menar att validering av CBT måste baseras på överväganden i vilken utsträckning datorkompetens, eller datoranvändningen i sig, påverkar utfallet. Författarna urskiljer några exempel på hot mot validiteten hos CBT:

- Metodeffekter (eventuell *construct-irrelevant variance*) när det gäller annorlunda provprestationer, som t.ex. oro eller ovana vid datorer
- Begränsningar (ev. *construct under-representation*) pga. uppgiftsurval eller uppgiftstyper som gör att bedömningen inte blir tillräckligt representativ eller varierad
- Otillräcklig eller irrelevant bedömning (ev. *construct under-representation*) vid automatisk rättning, pga. att CBT ännu inte kan hantera öppna svarsformat

Nytt construct för kommunikativ språklig kompetens

Chapelle & Douglas (2006) manar forskningen om CBT inom tillämpad lingvistik att beakta de språkliga kompetenser som krävs i teknologiskt medierad interaktion och kommunikation, och därför omvärdera *construct*:

Applied linguists would therefore speak of language ability as the ability to choose and deploy appropriate linguistic resources for particular types of situations. [...] These choices ultimately depend on the language user's technological and strategic competence, which together with linguistic competence may be the type of construct of relevance to language use through technology. In other words, 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 enligt författarna innebära en delning av begreppet språklig kompetens i olika kontextberoende *construct*, vilka i många situationer är förbundna med datorer. Exempel på sådana delkompetenser skulle vara *electronic literacy* och *multimodal literacy*.

En vidare syn på språklig kompetens återfinns också i de skrivningar om läs- och skrivkunnighet (*literacy*) som avser PISA (OECD – PISA, 2009, Reading Framework). I definitionen av läsning inbegrips läsning av såväl handskreven, tryckt som elektronisk text. Det senare innefattar också olika strategier kring att skrolla och hantera till exempel hyperlänkar, med motiveringen att detta är en viktig del av läsförmåga inom flera språkliga domäner. *Reading Literacy* definieras i PISA enligt följande: "an individual's capacity to understand, use and reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society" (OECD, 2009, s.14).

I USA har ETS (*Educational Testing Service*) i samarbete med flera amerikanska universitet utvecklat bedömningsmaterial för att pröva *ICT Literacy (Information and Communication Technology)* bland annat vid antagning till lärosäten och vid anställningar. I inledningen till materialet förklaras syftet:

Despite broad consensus on the need [...] to ensure that students enter college and the economy prepared for the complexities of an informationdriven society. We will surely one day succeed in closing the "digital divide" that separates the technology haves from the have-nots. But we will not succeed by merely providing access to equipment; we must also provide the intellectual tools to use technology effectively, in school and in the workplace. That is exactly what ETS is doing with a new ICT Literacy large-scale assessment.

I en artikel i tidskriften *SYSTEMICS, CYBERNETICS AND INFORMATICS* redogör Katz & Macklin (2007) bland annat för utvecklingsstudier när det gäller *ICT Literacy*. De ger också en

sammanställning av vilka komponenter som enligt ETS *ICT Literacy Assessment* utmärker språklig förmåga och kritiskt tänkande i teknologisk miljö (s. 51):

The assessment measures ICT literacy through seven performance areas, which represent important problem-solving and critical thinking aspects of ICT literacy skills.

Proficiency	Definition
Define	Using digital tools to identify and represent an information need
Access	Collecting and/or retrieving information in digital environments
Manage	Using digital tools to apply an existing organizational or classification scheme for information
Integrate	Interpreting and representing information, such as by using digital tools to synthesize, summarize, compare, and contrast information from multiple sources
Evaluate	Judging the degree to which digital information satisfies the needs of an information problem, including determining authority, bias, and timeliness of materials
Create	Adapting, applying, designing, or constructing information in digital environments
Communicate	Disseminating information relevant to a particular audience in an effective digital format

6.5 Påverkans effekter vid datorbaserad bedömning i språk

Frågor om bedömningens konsekvenser och påverkans effekter på olika nivåer avser användarrelaterade, etiska aspekter. Bachman & Palmer (2010) ställer som tidigare nämnts (i 6.4) som första krav för att kunna rättfärdiga en bedömnings användning att dess konsekvenser är av godo. De redovisar vidare, och betonar ansvarstagande för ett flertal möjliga påverkans effekter, främst avseende testtagare men också när det gäller beslut som rör såväl lärare, undervisning, organisation som administration. Shohamy (2001) understryker att det också existerar en makronivå och manar till vaksamhet:

Pretending that tests are neutral allow those in power to misuse them. Testers must realize that much of the strength of tests lies not only in their technical quality but in their use in social and political dimensions. Studies of the use of tests, as part of the validation on an ongoing basis, are essential for the integrity of the profession (s. 162).

6.5.1 Washback

Forskningen är generellt överens om att övergången från ett system till ett annat har olika implikationer på dess användning och medför påverkan på sin omgivning, vilket på olika sätt är relaterade till provningens validitet. Messick (1996) definierar fenomenet *washback* enligt följande:

Washback, a concept prominent in applied linguistics, refers to the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning (s. 241).

Eftersom *washback* och *bias*, med flera effekter, är sociala konsekvenser av ett provs användning, bör de enligt Messick ses som förknippade med bedömningens begreppsvaliditet:

The consequential aspect appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness and distributive justice, as well as to washback (s. 249).

Messick noterar att en del forskare menar att ett provs validitet påverkas av huruvida det genererar positiv eller negativ *washback*. Han framhåller dock vikten av analys och en viss försiktighet eftersom han menar att *washback* kan bero på en mängd faktorer inom ett utbildningssystem. Enligt Messick påverkar *washback* ett provs validitet endast om det kan beläggas att den är en direkt effekt av provet som sådant och inte av andra omständigheter. För att främja god *washback* menar Messick att man bör arbeta för att minimera eventuella hot mot bedömningens validitet, i form av *construct under-representation* eller *construct-irrelevant difficulty*.

Bailey ger i två artiklar (Bailey, 1996, 1999), en utförlig redogörelse för olika forskares syn på och arbete med begreppet *washback*, eller *backwash*, framför allt under 80- och 90-talen. Bailey refererar till exempel till Swain som i flera publikationer under 1980-talet framförde att kommunikativa språktest ska baseras på fyra principer: 1 *Start from somewhere*, 2 *Concentrate on content*, 3 *Bias for best* och 4 *Work for washback*. En utgångspunkt för Baileys genomgång är att *washback* kan vara både en önskad och oönskad effekt av prov och bedömning. Utifrån litteraturen konstaterar hon att *washback* generellt definieras som testnings påverkan på undervisning och lärande, att denna påverkan allmänt anses existera och vara viktig men att relativt lite empirisk forskning utförts för att dokumentera dess egentliga natur eller dess bakomliggande mekanismer (1996, s. 259).

Bailey gör vidare en uppdelning mellan *Washback to the learner* och *Washback to the programme*, framför allt baserad på forskning av Hughes (opublicerat material) och Alderson & Wall (1993). De senare föreslår femton distinktioner, vilka i en konsekvenskedja följer på varandra, som underlag för en analytisk bild av begreppet *washback*. Bailey (1999) kombinerar vidare Hughes och Alderson & Walls teorier i en utförlig schematisk bild över hur ett prov påverkar och interagerar med olika nivåer av användare (s. 264).

Bachman & Palmer (1996) betonar, liksom Shohamy (2001) och Hughes (2003), att ett provs påverkan kan urskiljas på två nivåer, dels en mikronivå som rör individuella elever och lärare och dels en makronivå som rör samhället och dess utbildningssystem. Författarna påpekar vidare att *washback* är komplext och måste sättas i ljuset av övergripande mål och värden: "In summary, the impact of test use needs to be considered within the values and goals of society and the educational system in which it takes place, and according to the potential outcomes of such use" (1996, s. 35).

Hughes (2003) hänvisar när det gäller *backwash* (en beteckning han förordar framför *washback*) till innehållsvaliditet (*content validity*), som en undergrupp av begreppsvaliditet (*construct validity*). Han varnar för att om en tongivande bedömning till exempel har skev eller snäv representation av det relevanta innehållet (*content*) tenderar detsamma att bli negligerat i undervisningen. Negativ *backwash* orsakas också ofta, enligt Hughes, av att bedömningen inte uppfattas pröva "rätt sak" (*face validity*). Hughes ser emellertid positiv *backwash* av bedömning som viktig och önskvärd och rekommenderar ett antal åtgärder vid konstruktion och utveckling av provmaterial (s. 53-56):

1. Test the abilities whose development you want to encourage.
2. Sample widely and unpredictably.
3. Use direct testing.
4. Make testing criterion-referenced.
5. Base achievement on objectives.
6. Ensure [that the] test is known and understood by students and teachers.
7. Where necessary, provide assistance to teachers.
8. Count the cost.

6.5.2 Erfarenheter av och åsikter om implementering av CBT

I vilken mån ett prov är användarvänligt, tillgängligt och möjligt att tillämpa i praktiken benämns dess praktikalitet. Bachman & Palmer (1996, 2010) definierar ett provs eller en testmetods praktikalitet som skillnaden mellan tillgängliga resurser och dess krävda resurser, och syftar på både mänskliga som materiella resurser och resurs i form av tid. De skriver:

The last test quality that needs to be considered is practicality, which is different in nature from the other five qualities. [dvs. Reliability, Construct validity, Authenticity, Interactiveness, Impact] While those qualities pertain to the uses that are made of the test scores, practicality pertains primarily to the ways in which the test will be implemented, and, to a large degree, whether it will be developed and used at all (1996, s. 35).

Chapelle & Douglas (2006), Chapelle (2010) och Douglas (2010) påpekar att man bör vara uppmärksam på att CBT kan innebära vissa säkerhetsrisker och att implementeringsfasen medför ett ökat behov av support. De understryker vikten av att implementera stegvis och undersöka aspekter som kostnader, påverkans effekter på inläring och även familiaritet med och oro inför datoranvändning, men menar också att inga av dessa möjliga negativa aspekter tyder på att CBT/CALT är mer tveksamma än andra prov. Chapelle (2010) understryker däremot möjlig positiv *washback* i form av ökad användning av datorteknologi som gör det möjligt för elever att få kontakt med målspråket i och utanför klassrummet. Författarna menar att implementering av CBT och CALT kan leda till ökad *computer literacy* men även *target language literacy*, och att båda är nödvändiga för framgång.

Även Meijer (2009) menar att implementering av CBT är ett komplext beslut och varnar för att betrakta CBT enbart som en övergång från ett instrument till ett annat. Enligt Meijer är bedömning en del av pedagogiska och socio-politiska system, som genom att mäta vad som pågår i dessa system oundvikligen förändrar desamma (*the observer effect*). Han betonar också att bedömningsinstrument upplevs som markörer dels för vad som är viktigt att lära och dels för vad som utgör kompetens inom ett område, så kallad *feedback effect* (s. 104).

Erfarenheter vid ETS (Educational Testing Service).

van Lent (2009), verksam vid ETS (*Educational Testing Service*) Global, redogör för hur en viss försiktighet präglar ETS och liknande organisationer när det gäller övergången från traditionella pappersprov (PBT) till CBT vid storskalig prövning av *high-stakes* karaktär. Han efterfrågar en *benefit-risk approach* som kan belysa viktiga hänsynstaganden för utbildningsmyndigheter som måste ta ansvar för sådana beslut. van Lent menar också att ett nytt medium kan få konsekvenser för en eventuell förändring av *construct* och att det inte är försvarbart att riskera likvärdighet och validitet för ett medium som öppnar för möjligt *construct irrelevant variance*. Han menar dock att det inte är frågan om *om*, utan *hur* en övergång ska kunna ske med bevarad och helst förbättrad validitet, likvärdighet, användbarhet och trovärdighet — när pedagogisk forskning sackar efter, vilket gör att utvecklingen av CBT för närvarande styrs av vad tekniken tillåter:

The current state of CBT can be characterized as: some CBTs offer little or no added value; some “innovative” items are likely to contribute more ‘artifactual’ than valid measurement; limited site capacity often forces continuous administration, which can introduce serious security concerns; test administration algorithms are getting smarter but remain limited. Key issues that have to be taken care of are linked to design, accessibility and security. Appropriate underpinning with relevant research is an absolute necessity (s. 84).

Bridgeman (2009), ETS USA, redogör vid samma konferens som van Lent ovan, för en bred användning av CBT i USA och listar i sitt *paper* ett tiotal storskaliga exempel på CBT, men framhäver att de största antagningsproven till amerikanska college, så kallade *admissions tests*, ges som PBT då de görs av miljontals studenter inom loppet ett par månader. Han menar att CBT har åtskilliga fördelar jämfört med PBT men manar, liksom van Lent, till en viss försiktighet då ny teknik också oundvikligen för med sig nya utmaningar som kan vara svåra att förutse.

Några av de fördelar han anger är lägre kostnader och säkrare distribution, mindre tidsåtgång, omedelbar rättning och möjligheter också till andra data som ger insyn i processen. Dock kan sårbarhet vid till exempel strömavbrott eller avbrott i uppkopplingar skapa stora svårigheter. Han manar också till vaksamhet utöver den rena datortillgången, eller provtagarnas datorvana, och hänvisar till en egen studie kring konsekvenser av användning av olika sorters datorer 2003. Denna visade att elever som gjorde läsprov på stor skärm med god upplösning fick bättre resultat än de med liten skärm med sämre upplösning, samt att de som skrev på *laptop* hade sämre resultat än de elever som skrev på vanlig dator. Han ser det också som en möjlig fördel att testtagare använder samma inbyggda hjälpmedel som t.ex. ordbok, men varnar för att alla inte kan vara lika vana att använda dem, vilket påverkar likvärdigheten.

Implementering av nationella CBT i Norge

Carlsen (2008) och Moe (2009) redogör för och diskuterar implementeringen av det norska nationella provsystemet i bland annat engelska. Universitetet i Bergen ansvarade mellan 1998 och 2003 för utvecklingen av den norska versionen av DIALANG, vilket gjorde att de hade viss erfarenhet av CBT när datorbaserade nationella prov skulle införas 2003. Eftersom det bedömdes att man varken hade serverkapacitet, möjlighet att administrera eller tillräcklig utrustning på skolorna att genomföra prov med alla elever i en årskurs samtidigt, tillhandahölls tre likvärdiga prov under en fjortondagarsperiod. Adaptiva prov skulle kräva en alltför stor

uppgiftsbank och svårigheter att kontrollera provens innehållskvalitet, varför det, efter viss försöksverksamhet med delvis adaptiva prov, beslutades göra linjära prov (Moe, 2009). Skolorna informerades om de tekniska förutsättningarna och exempeluppgifter tillhandahölls på nätet ungefär sex månader innan det första provet skulle genomföras. Några skolor hade vid implementeringen inte uppgraderat sina uppkopplingar eller sin datorkapacitet, men något år senare, 2005, genomförde cirka 200 000 elever i fyra årskurser datorbaserade prov. Resultat och enkäter från 2008 visar bland annat att de flesta eleverna 'tycker om' proven, att avsevärt fler pojkar än flickor tycker att de är 'lätta', men att detta inte avspeglas i resultaten, och att en del elever tycker att det är jobbigt att läsa på skärmen (Moe, 2009).

Senare har de norska proven mottagits väl, men opinionen var under de första åren starkt emot nationell prövning som sådan (Carlsen, 2008; Moe, 2009). Det faktum att inte bara prov har påverkans effekter på olika nivåer, utan att också prövningen i sig påverkas av dess användare och det samhälle den förekommer i, diskuteras med utgångspunkt i en norsk kontext av Carlsen (2008) i artikeln *The role of testing in an egalitarian society*. Carlsen konstaterar att utveckling av nationell prövning och betygsättning möter speciella utmaningar i ett samhälle som präglas av Jantelag och en stark demokratisk jämlikhetstradition.

Det norska nationella provsystemet hade initialt dels en pedagogisk och dels en rapportrande funktion och initierades bland annat på grund av nedslående resultat på PISA 2000 (då Norges resultat var medelgoda totalt sett, men sämre än Sverige och 'vinnaren' Finland). Dessa syften kom enligt Carlsen och Moe i viss mån att motverka varandra; arbete med att skapa positiva *washback*-effekter och ge konstruktiv *feedback* till lärare och elever kom i konflikt med att den rapportrande funktionen prioriterades och resultaten skulle Internetpubliceras på skolnivå, vilket förutsatte *single scoring* istället för profiler och detaljerad feedback. Den allmänna opinionen i Norge reagerade i början starkt och till exempel bojkottades de första proven av elever och föräldrar, då publiceringen av resultaten befarades öka segregationen i samhället och bland annat leda till fler friskoleetableringar (Carlsen, 2008; Moe, 2009). Det är, menar Carlsen, olyckligt att en bred allmän debatt för och emot bedömning i sig kan tendera att överskugga och i viss mån hindra en fruktbar diskussion om provkvalitet, testetik och rimlig och rättvis användning av resultat. Hon drar också slutsatsen att motstånd mot prövning i ett demokratiskt samhälle är något av en paradox, eftersom: "In a society where goods, positions and privileges are distributed by qualifications and not by birthright, testing is an indispensable tool" (s. 5).

6.6 Likvärdighetsaspekter – fairness, bias och familiaritet

6.6.1 Likvärdighet och bias

De svenska nationella proven vilar på uttalad demokratisk grund och den första punkten som på systemnivå beskriver provens uppdrag och syfte anger att de ska stödja en likvärdig och rättvis bedömning och betygsättning. I de gemensamma principer som omfattar de nationella provmaterialen i engelska och främmande språk definieras detta ytterligare, och också på individnivå. Där betonas att elever i proven ska ges möjlighet att visa vad de kan (och inte

främst vad de inte kan) bland annat genom att materialens olika delar erbjuder variation, både vad gäller innehåll och svarsform (NAFS-projektet — Gemensamma principer).

Bachman & Palmer (2010) skriver angående likvärdighet (*fairness*) att bedömningsforskningen generellt är överens om att likvärdighet och rättvisa är en grundläggande angelägenhet, men att begreppet förefaller svårt att definiera, då likvärdighetsaspekter är knutna till såväl process, som produkt och konsekvenser. Författarna framhåller *fairness* som centralt, i synnerhet vid bedömning av *high-stakes*-karaktär, vilket också avspeglas i de grundläggande krav de föreslår för all bedömning (se 4.4), och de förklarar att:

If the assessment procedure does not provide assessment records that are **consistent** across different groups of individuals, it is not fair. If the interpretations that we make of assessment records for individuals who are members of different groups are differentially **meaningful** or are not **impartial**, then they are not fair. If the interpretations are not **generalizable** beyond the assessment itself, or are not **relevant** and **sufficient**, then the decisions we make may not be fair. If the decisions we make ignore community **values** and legal requirements and are not **equitable** for individuals in different groups, they will not be fair. If the consequences of using the assessment and of the decisions made are not **beneficial**, then these will not be fair (Bachman & Palmer, 2010, s.131).

Deras slutsats blir att likvärdighet beror av två aspekter av bedömningen, nämligen jämlik och rättvis behandling på individnivå i processen, samt frånvaro av *bias* avseende process, utfall och tolkning av resultat. *Bias* definieras vidare som: "**Bias** can be defined as a difference in meanings of assessment records for individuals from different identifiable groups (e.g. by gender, ethnicity, religion, native language) that is not related to the ability that is assessed" (s. 129). Som troliga källor till *bias* anges vidare omständigheter relaterade till bedömningsmaterialets format, innehåll (*topical content*) och till eventuell inkonsekvens vid bedömning av resultatet.

6.6.2 Likvärdighet i relation till validitet

Medan etik inom språkbedömning generellt förknippas med grundläggande praxis och förhållningssätt och således betraktas som ett övergripande koncept som omfattar validitet och likvärdighet (Xi, 2010), diskuteras inom bedömningsforskningen emellertid förhållandet mellan bedömningens likvärdighet och dess validitet. I en artikelserie om begreppet *fairness*, kategoriserar Xi (2010) olika synsätt på i vilken utsträckning likvärdighet är knutet till validitet. Likvärdighet betraktas enligt dessa antingen som en oberoende kvalitet hos bedömningen, som en övergripande kvalitet som inbegriper validitet eller som en delaspekt av bedömningens validitet. Enligt Xi skulle en ytterligare distinktion underlätta analysen av likvärdighetsaspekter:

This approach treats fairness as an aspect of validity and conceptualizes it as comparable validity for all **relevant** groups. Anything that weakens fairness compromises the validity of a test. This conceptualization expands the scope and enriches the interpretations of fairness by drawing on well- defined validity theories while enhancing the meaning of validity by integrating fairness in a principled way (Xi, 2010, s.147).

Xi pekar också på flera avgörande faktorer som orsakar partiskhet i bedömning, som bl.a. olämpliga uppgifts- eller svarsformat, olämpligt innehåll, innehåll som underrepresenterar *construct*, samt *rubrics* som inte representerar de färdigheter som krävs i domänen - eller som representerar irrelevanta färdigheter.

I samma artikelserie replikeras Xi av bland andra Davies (2010) och Kane (2010). Davies gör gällande att "the pursuit of fairness in language testing is vain: first because it is unattainable and second because it is unnecessary" (s.171). Han anser att en vidgad syn på validitet i sig inbegriper opartiskhet och att bedömningens validitet alltid måste relateras till dess avsedda målgrupp:

Like justice, validity guarantees that an ability is being appropriately tested for a relevant population: this population will be made up of various groups but there is sufficient homogeneity across groups for them to be treated as comparable. A test that is valid for group A (say adults) cannot be valid for group B (say children) because they belong to different populations. It is not a question whether such a test is fair or unfair for group B: the test is just invalid for group B. The search for test fairness is chimaerial (Davies, 2010, s. 175).

Enligt Kane, som är mer positiv till Xis hypotes ovan, beror förhållandet mellan validitet och likvärdighet av hur de definieras, och han menar att en bredare definition av det ena av nödvändighet innefattar en smalare definition av det andra. Han anser att en bedömning som är partisk och systematiskt missgynnar vissa grupper av testtagare inte heller är valid avseende dessa grupper och att en invalid bedömning som genererar felaktiga slutsatser heller inte är likvärdig eller opartisk. Liknande resonemang förs också av Douglas (2010) som konstaterar att "there is no such thing as a valid test, only tests which have shown to be valid for certain purposes" (s. 35).

6.6.3 Jämförande studier – familiaritet

Jämförande studier i Norge har dokumenterat goda erfarenheter av digitala nationella prov (Moe, Carlsen & Hasselgren, 2006) och i Korea har Choi, Kim & Boo (2003) jämfört CBT med PBT i språk och funnit att överensstämmelsen när det gäller reliabilitet och validitet är god, med särskilt fokus på aspekter relaterade till *content* och *construct*. Vid jämförande analyser av CBT och PBT som hör- och läsförståelseprov, samt grammatik- och vokabulärprov fann Choi, Kim & Boo högst korrelationer vid grammatik och lägst vid läsning, vilket de sätter i samband med att flertalet testtagare påpekat ögontrötthet och koncentrationssvårigheter vid läsning på skärm.

När det gäller frågor om likvärdighet och eventuell *bias* avseende CBT återkommer flera forskare till själva datoranvändningen och faktorer avseende *rubric* som följer på datorbaserad. Chapelle & Douglas (2006) problematiserar att låg datorfamiliaritet kan leda till annorlunda provprestationer jämfört med pappersprov (PBT) för vissa grupper, men konstaterar att sådana effekter eventuellt kan minskas genom att testtagare erbjuds exempelprov före bedömningstillfället. Acceptans av CBT avseende TOEFL har undersökts av Stricker, Wilder & Rock (2004) och deras studie har sedan följts upp av Stricker & Attali (2010), efter det att TOEFL Internetbaserades 2009 (TOEFL iBT™). Den uppföljande studien uttrycker ambition att göra mer finmaskiga undersökningar, men båda studierna finner generellt god acceptans för CBT och inga nämnvärda samband mellan familiaritet och prestation. Detta tolkas som att testtagare, i

USA och andra länder, redan är väl förtrogna med datorbaserade prov trots att teknologin ännu är i sin linda (2004, s. 49).

Kikis- Papadakis & Kollias (2009) uttrycker emellertid stark oro över 2000-talets enorma ökning av IKT-prövningar och CBT (*e-assessments*) när det gäller *placement-tests* och *admission-tests* av *high-stakes*-karaktär. De menar att det som av många ses som en effektivisering och ökad tillgänglighet, samtidigt systematiskt diskriminerar och exkluderar de socio-ekonomiska grupper som inte har likvärdiga möjligheter när det gäller familiaritet varken med datormediet som sådant eller med IKT. Liknande rapporteras också av Sireci & Zenisky (2010) som angående nya format anser att "In particular there is very little research on the social considerations of innovative item formats. This is unfortunate because computer familiarity is likely to be associated with socioeconomic status and other demographic variables" (s. 345).

Kikis-Papadakis & Kollias hävdar också att skolkulturen i Europa överlag är konservativ när det gäller såväl datorutrustning som användning av och undervisning om IKT, och även när det gäller tilltro till värdet av IKT, vilket framför allt missgynnar de elevgrupper som är beroende av skolan i detta avseende. Emellertid ser Sireci & Zenisky också i princip det omvända, nämligen risker för *bias* när det gäller traditionella pappersprov (PBT) på grund av *construct-irrelevant variance (CIV)* som då drabbar vana datoranvändare:

Increasingly, many skills used in real life are done in a computer environment. Without movement toward greater use of computer-based testing, the potential exists for CIV to creep into a paper-based testing environment because of unfamiliarity with paper test methods (s. 345).

6.6.4 Jämförande studier – genusaspekter

En annan distinktion som studeras när det gäller skillnader i provresultat mellan datorbaserade prov och pappersprov avser eventuella könsskillnader. Halldórsson, McCelvie & Björnsson (2009) jämförde utifrån ett genusperspektiv isländska elevers resultat på PISA 2006 som pappersprov och CBAS 2006 (*Computer-Based Assessment of Scientific Literacy*), mot bakgrund av att Island var det land som vid PISA 2003 hade de största könsskillnaderna i resultat, till flickornas fördel. I samtliga tre deltagande länder (Island, Danmark och Korea) hade däremot pojkar bättre resultat än flickor på CBAS 2006. Studien indikerar att isländska pojkar visar högre motivation, familiaritet och självförtroende när det gäller uppgifter som involverar IKT, men att dessa faktorer inte visar sig ha signifikant förhållande till resultat. Enligt författarna förklaras resultaten delvis av att mindre mängd löpande lästext i CBAS jämfört med PISA PBT är till fördel för pojkars resultat. Enligt forskarna stöds detta av erfarenheter som visat att könsskillnader minskar, även vid pappersprov, om textmängden i uppgifter reduceras till under 100 ord, samt av motivationsfaktorer som indikerar att pojkar visar tendens att oftare 'ge upp' vid upplevda svårigheter som t.ex. större textmängd.

Liknande resultat rapporteras av Martin & Binkley (2009) som anser att fokus på klassisk läsförståelse med löpande text visat sig innebära *bias* till flickors fördel, som t.ex. i PISAs läsförståelsedel, eftersom flickor generellt har större familiaritet med den lästypen. De menar att CBT, med innovativa format och ny *rubric*, kan möjliggöra större variation och utjämna könsskillnader som beror på att flickor och pojkar generellt tycks föredra olika typer av stimuli och format. En utökad syn på läskompetens (*reading literacy*) som också omfattande IKT-

kompetens skulle, enligt författarna, förmodligen göra att pojkar fick större möjlighet att visa sin förmåga.

Jämförande studier som involverat prov i flera ämnen har också gjorts när det gäller de nationella ämnesproven i engelska, svenska och matematik för årskurs nio i Sverige, i relation till *rubric* men inte datoranvändning. Åberg-Bengtsson & Erickson (2006) har redovisat starka korrelationer på individnivå mellan prestationer när det gäller liknande format oavsett lingvistiskt eller matematiskt ämnesfokus. De framhåller därför vikten av varierade uppgiftsformat inom prov, då vissa format tydligt och systematiskt gynnar vissa typer av elevers profiler och lärstilar. Författarna framhåller också att hänsyn bör tas till att en uppgifts svårighetsgrad i sig har olika bevekelsegrunder – den kan på ett teoretiskt plan bedömas 'svår' i förhållande till *construct*, empiriskt visa sig 'svår' eller upplevas av testtagare som 'svår'.

6.6.5 PISA 2009

I Skolverkets rapport *Eleverna och nätet* (Skolverket, 2011b) redovisas och diskuteras resultat från PISA 2009. För första gången erbjöds då, förutom det traditionella läsprovet, också ett frivilligt prov i digital läsförståelse och båda dessa läsprov genomfördes av elever från 16 OECD-länder och tre andra länder, och däribland ca 2000 svenska elever i årskurs nio. De båda proven var inte identiska och en viktig skillnad var att navigation ingick som en färdighet i det digitala läsprovet, vilket innebär att med hjälp av IT söka information för att konstruera text (se 6.4.3 om förändrat *construct*).

Korrelationen mellan traditionell läsning och digital läsning betraktas som hög. Totalt bland de 16 deltagande OECD-länderna var sambandet mellan resultaten på de båda läsproven 0.83 och för de svenska eleverna var korrelationen 0.87. Enligt analysen samvarierar också generellt resultatet på det traditionella läsprovet med resultaten på övriga prov – i digital läsning, matematik och naturvetenskap. Förmodligen, menar författarna, ligger högre resultat på det traditionella läsprovet bakom att vissa länder har högre resultat än Sverige på övriga prov, och också på det i digital läsning. Svenska elever har ett högre medelvärde på det digitala läsprovet än på det traditionella, och de presterar bättre än genomsnittet inom OECD på detta prov. Bland OECD-länderna har fyra länder (Sydkorea, Nya Zeeland, Australien och Japan) medelvärden som är signifikant högre än det svenska medelvärdet på det digitala läsprovet, tre länder (Island, Irland och Belgien) medelvärden som inte skiljer sig ifrån det svenska medelvärdet och 10 länder medelvärden som är signifikant lägre än det svenska medelvärdet (s. 44). I studien belyses provprestationer i förhållande till bakgrundsvariabler som datorvana, utländsk bakgrund, socio-ekonomisk bakgrund och genus. Skillnaderna är mindre avseende grupp på det digitala läsprovet än på det traditionella för Sveriges del, men diskrepanserna är också i flera avseenden större än i flera andra länder.

94 % av de svenska eleverna anger dels att de surfar på nätet hemma mer än en gång i veckan, dels att läsa eller skriva digitalt känns mer bekant än på papper. Resultaten visar att elever som har tillgång till dator hemma presterar bättre på det digitala lästestet, men också att de måttliga datoranvändarna, med måttligt självförtroende, presterade bättre än de frekventa med högre självförtroende. Det framkommer emellertid inte något direkt samband mellan datoranvändning i skola eller hem och resultaten i digital läsning, vilket samvarierar i alla deltagande länder. Författarna problematiserar avseende detta att elever använder datorer till

olika saker och att skolan sällan har undervisning i egentlig dator- eller IKT-kompetens, utan oftast låter elever arbeta ensamma vid datorer. "Datoranvändning" är enligt författarna sannolikt ett för oprecist begrepp att använda vid liknande jämförelser.

Generellt visar resultaten att diskrepansen i resultat är större i Sverige än i OECD som helhet avseende elever med inhemsk respektive utländsk bakgrund. Elever med utländsk bakgrund, oavsett om de är födda i Sverige eller inte, har lägre resultat på båda läsproven, men lyckas signifikant bättre på det digitala läsprovet än på det traditionella. Också avseende socio-ekonomisk bakgrund tycks för Sveriges del resultaten visa att dess påverkan är mindre på det digitala lästestet än på det traditionella, men socio-ekonomisk bakgrund verkar ha större betydelse i Sverige än i övriga nordiska länder. Jämfört med alla övriga deltagande OECD-länder ligger Sverige på ett medelvärde när det gäller denna samvarians. Avseende genus har i alla länder flickor signifikant bättre resultat på båda proven, men skillnaden mellan pojkars och flickors resultat är mindre på det digitala provet. Författarna menar att förklaringar till pojkarnas fördel avseende detta kan vara att texterna är kortare och att det är färre löpande texter i det digitala provet, samt att pojkar känner sig mer motiverade av mediet (jfr ovan Halldórsson, McCelvie & Björnsson, 2009; Martin & Binkley, 2009).

6.7 Avslutande kommentar: Ett nytt steg för de nationella provmaterialen i språk.

Efter avsnitt 5 om de svenska nationella provmaterialen i engelska användes ovanstående rubrik som en fråga. Det förefaller nu rimligt att använda den igen, men som ett påstående. De svenska provmaterialen i språk har, som tidigare återgivits, försiktigt påbörjat en datorbaserad och Skolverket har på huvudmannanivå inlett diskussioner om nästa steg. Ovanstående litteraturgenomgång och spegling av internationella erfarenheter, samt det faktum att internationella prövningar som PISA nu prövar digital läsning, bl.a. i ljuset av Europarådets nyckelkompetenser (se 6.2), visar att datorbaserade och digitaliserade bedömningsformer som också delvis prövar nya kompetenser har stor spridning. Språkforskningen varnar emellertid för att teknologin måste inordnas i dess tjänst och inte tillåtas rusa iväg eller övervärderas (se t.ex. 6.1 och 6.4.2) och höjer ett varnande finger angående en mängd validitetsaspekter som måste beaktas av huvudmän och de lärosäten som utvecklar provmaterial. Litteraturen visar att ett prov dock kan vara av hur god kvalitet som helst men om det inte fungerar ute i verksamheten, tas väl emot och får goda konsekvenser och *washback*-effekter (se 6.5.1), faller det inte väl ut och brister också i validitet (se t.ex. 4.2 och 6.4.3), eller (se 6.6.2) som Douglas (2010) konstaterar: "there is no such thing as a valid test, only tests which have shown to be valid for certain purposes" (s. 35).

Som tidigare nämnts (se 5) finns det inom NAFS-projektet en lång tradition och erfarenhet av breda utprövningar av provmaterial för att undersöka bl.a. användarrelaterade aspekter som mottagande, reaktioner och praktikalitet. Detta tillvägagångssätt och samma avsikt ligger till grund för nästa avsnitt, den empiriska studien — Hur fungerar datorbaserade prov på ett antal svenska grundskolor och vilka erfarenheter, reaktioner och resultat ger en sådan studie? Detta är, i något annan ordalydelse, denna undersöknings andra forskningsfråga och denna studie innebär i sig ett nytt steg för de nationella provmaterialen i språk, då en datorbaserad utprövning inte tidigare gjorts.

7 Empirisk studie – metod och design

Inom ramen för arbetet med denna uppsats genomfördes en empirisk studie, som bestod i att elever och lärare vid sex skolor gjorde och reflekterade över datorbaserade norska nationella prov i engelsk läsning. Nedan redovisas den empiriska studiens ansats och utformning. Vidare behandlas etiska frågor, urvalsgrupp, genomförande samt underlag och resultat. Bearbetningen av materialet redovisas, och undersökningens trovärdighet och generaliserbarhet problematiseras.

7.1 Metod

7.1.1 Ansats

För att kunna skapa en bred och god bild av attityder, uppfattningar och reaktioner avseende de datorbaserade proven användes olika metoder för insamling av underlag. Den mesta informationen samlades in i form av enkätsvar, men också i samband med deltagande observation samt vid informella samtal och kontakter via e-post.

Dataunderlaget i studien har analyserats induktivt. Såväl kvantitativa som kvalitativa data har lagrats på ett sådant sätt att de möjliggör analyser på olika nivåer för att finna såväl mönster som variationer. Sammanställning och analys av kvantitativa data med hjälp av statistikprogram som SPSS har drag av positivism (Alvesson & Sköldberg 2008; Stukát, 2005). På ett övergripande plan har studien en kvalitativ ansats som i viss mån liknar en induktiv etnografisk fallstudie, då det egentligen undersökta (datorbaserade provmaterial) inte går att skilja från sin unika (skol)kontext (Hammersley & Atkinson, 1983; Stukát, 2005; Wallace, 1998).

Den empiriska studien har som fokus att kasta (nytt) ljus på fenomen, upptäcka eventuella problem och framför allt undersöka validitetsaspekter med fokus på begrepps- och konsekvensvaliditet, *construct-* och *consequential validity* (Messick 1989) samt, enligt praxis för god bedömning (EALTA, 2006), belysa praktikalitet, likvärdighet och användbarhet avseende datorbaserade prov. Denna framåtsyftande avsikt att bidra till att utveckla relevanta, användbara, pålitliga och likvärdiga provmaterial kan sägas bära förtecken av ett pragmatiskt synsätt på bedömning och prövning (Alvesson & Sköldberg, 2008). God bedömning för lärande förutsätter också tydlig användarfokus i form av intresse för attityd- och motivationsaspekter när det gäller framför allt elevens och lärares upplevelse av bedömningssituationen och provmaterialet, vilket inspirerats av motivationsforskning (Dörnyei, 2001; Giota, 2002).

7.1.2 Undersökningens struktur

Den empiriska studien planerades och genomfördes enligt följande arbetsgång:

- Initiala muntliga och skriftliga kontakter med skolor
- Planeringsbesök med samtal bland annat om tekniska förutsättningar
- Enkät till elever före genomförandet av proven
- Genomförande av datorbaserade provuppgifter
- Deltagande observation under ett genomförande per skola
- Enkäter till elever, lärare och skolledare samt samtal efter genomförandet

- Rapport av provresultat samt avstämning med medverkande lärare

7.1.3 Forskningsetiska överväganden

I enlighet med Vetenskapsrådets forskningsetiska principer (Vetenskapsrådet) inleddes studien med omfattande såväl skriftlig som muntlig information, initialt till rektorer och lärare och sedan till elever och föräldrar, om undersökningens bakgrund, syfte och genomförande, samt kommande användning av resultaten i rena forskningsändamål. Samtliga involverade ställde sig positiva till att delta. Då studien bland annat riktade sig till omyndiga elever rekommenderades skolorna att ta in skriftliga medgivanden från vårdnadshavare. De elever som inte lämnat sådant medgivande deltog inte i studien. Samtliga enkäter avidentifierades inför inmatning i statistikprogrammet SPSS. Endast uppgifter om respondentgrupp, årskurs, skola och kön bevarades för att möjliggöra analys ur flera aspekter. Provresultaten registrerades på individnivå och meddelades lärarna efter studiens avslutande men har sedan avidentifierats enligt ovan inför analysen.

7.1.4 Urvalsgrupp

När det gäller urval av skolor fanns en strävan att inom det begränsade antalet få viss spridning avseende till exempel storlek och geografiskt läge. Eftersom flera besök och ett relativt nära samarbete var en förutsättning för att få en god bild av genomförandet, riktades urvalet till sex kommunala skolor i Västra Götaland, tre i vardera två olika kommuner — en storstadskommun (skola A, B och C i resultatredovisningen nedan) och en landsortskommun (skola D, E och F). Skolorna har mellan 100 och 500 elever; två av skolorna (A och E) har åldersblandade klasser, en skola (A) har grupper som undervisas enligt Montessoripedagogik och en skola (F) har enligt sin hemsida en tydlig IT-satsning. En skola (B) deltar med både yngre och äldre elever.

Avsikten var att genomföra proven med elever i såväl yngre som äldre åldrar i grundskolan. Totalt engagerades 276 elever från de olika skolorna, enligt nedan. Efter ett visst bortfall besvarade totalt 248 elever någon av enkäterna och 209 elever gjorde båda enkäterna. Det var i stort sett lika många pojkar som flickor, samt ungefär lika många elever fördelat över äldre (åk 8) och yngre åldrar (åk 5/6) som deltog. Däremot var det nästan dubbelt så många elever från de tre storstadsskolorna; 165 elever jämfört med 83 från de tre landsbygdsskolorna.

Skola	åk5	åk6	åk8
A	22	36	
B		48	24
C			60
D		22	
E	13	9	
F			42
	35	115	126

Figur 1. Översikt över det planerade antalet deltagande elever per årskurs och skola.

De tretton lärare, tre män och tio kvinnor, som involverades i studien är alla lärare i engelska eller klasslärare som undervisar i engelska. Av dessa tretton lärare kom elva att genomföra prov

med elevgrupper. De sex skolorna representeras också av åtta skolledare. En skola (F) har två rektorer och en skola (B) har rektorsuppgiften för yngre respektive äldre åldrar uppdelad på två tjänster.

7.2 Material

7.2.1 Uppgiftsdomän

När det gäller datorbaserade prov har erfarenheter rapporterade i litteraturen och även förmedlade vid studiebesök i våra grannländer (Skolverket, 2010/Nyström) tydligt visat att de produktiva språkfärdigheterna kräver manuell bedömning. Vidare förutsätter receptiva uppgifter av hörförståelsetyp mer avancerad utrustning än läsförståelse. Det beslutades därför att använda uppgifter med fokus reception – läsning, med slutna svarsformat som rättas löpande på maskinell väg. Som tidigare nämnts i 6.4 är detta är den uppgiftstyp som hittills visat sig fungera bäst i datorbaserad form (Moe, 2009; Skolverket, 2010/Nyström).

7.2.2 Provmaterial

Kontakter inleddes under hösten 2008 kring möjligheten att i den engelska studien använda datorbaserade uppgifter som utvecklats i något av våra grannländer. Förhoppningen var att eventuellt kunna genomföra en studie i två delar – en som baserades på CBT (*Computer Based Test*) och en på CAT (*Computer Adaptive Test*) med utgångspunkt i provmaterial från Norge respektive Danmark. Relativt långt fram under förberedelserna pågick ett samarbete som skulle innebära användande av danska uppgifter till ett adaptivt prov, men detta fullföljdes aldrig av administrativa, logistiska och innehållsliga skäl.

En överenskommelse med Universitetet i Bergen möjliggjorde att använda 2008 års norska nationella prov i engelsk läsning. Digitala prov i engelsk läsförståelse infördes i Norge 2004 efter uppdrag av Stortinget och avser pröva måluppfyllelse av kunskapsmålen för läsning i engelska för fjärde och sjunde trinn (årskurs). Dessa prov bygger på beskrivningarna för läsförståelse i CEFR, den europeiska referensramen för språk (Moe 2008).

7.2.3 Provuppgifter

De nationella proven för engelsk läsning i Norge genomförs under en tvåveckorsperiod alldeles i början av 5:e respektive 8:e trinn/åk. Proven är linjära (se 3.1) och varje år ges tre likvärdiga versioner som slumpas ut, för att de ska kunna genomföras vid något olika tidpunkter på skolorna. Till den svenska studien användes version tre av vardera provet för åk 5 och 8:

Årskurs 5		Årskurs 8	
Itemformat	Prøve 3	Itemformat	Prøve 3
Click item	3	Click item	1
Click& drag	4	Click& drag	2
Colour	3	Colour	3
Click picture	6	Click picture	1
Click text	7	Click text	2
Gap filling	6	Gap filling	4
Multiple choice	3	Multiple choice	10
Who could say	6	Who could say	13
		Double click word	4

Figur 2: Uppgiftstyper i version 3, Nasjonale prøveversionar engelsk 2008 (Moe, 2008).

I båda proven förekommer rikligt med bilder och olika typer av uppgiftsformat (se Figur 2), men endast slutna svarsformat, dvs. olika typer av flervalsuppgifter som rättas löpande. Ämnesmålen för läsförståelse i engelska i de norska respektive svenska styrdokumenterna är i stort mycket lika varandra, men de norska proven har också ett uttalat syfte att testa elevens förmåga att använda datormediet, IKT, dvs. att de förhåller sig till ett delvis annat *construct* än de svenska materialen (jfr. 5.1). Provet för åk 5 består av 38 uppgifter som prövar elevens förmåga att finna information, förstå vardagliga engelska ord och uttryck samt förstå huvudinnehållet i texter. De 40 uppgifterna för åk 8 prövar förmåga att finna information, förstå huvudinnehållet i samt reflektera över innehållet i texter. Några uppgifter handlar således om att lokalisera detaljer och andra om att dra slutsatser utifrån given information. Uppgifterna i de båda proven är när det gäller format och tillvägagångssätt nästan identiska, men de yngre eleverna behöver aldrig skrolla i längre texter. Eleven svarar till exempel genom att flytta bildobjekt, klicka på ord eller bilder, färglägga föremål eller välja bland givna namn, ord eller fraser i flervalsalternativ (Moe, Carlsen & Hasselgren, 2006).

Provuppgifterna rättas maskinellt, och normalt sett får de norska eleverna sitt resultat omedelbart efter avslutat prov. För studiens del beslutades dock att meddela skolorna resultaten först efter studiens slut, då eventuellt framför allt elevernas utvärdering av utprövningen annars skulle riskera att färgas av deras individuella prestationer.

7.2.4 Enkäter och fältanteckningar

Samtliga enkäter utformades när det gäller frågeformat och även en del frågeställningar enligt modeller som under många år använts inom NAFS-projektet för Test Taker Feedback (TTF) och Teacher Feedback (TF), (se 5.2). De innehåller slutna format med fem-punkts så kallade Likertskalor (Andersson, 1994), i vilka respondenten ombeds ta ställning till ett påstående genom att markera *Ja, absolut* – *Nej, absolut inte* med tre steg däremellan, samt möjlighet att lämna öppna kommentarer. Eleverna gjorde en enkät före (Elevenkät I; Bilaga 1) och en enkät efter provtillfället (Elevenkät II; Bilaga 2). Lärarna och skollädaarna besvarade digitala enkäter efter periodens slut, vilka utväxlades per e-post. Eftersom planeringsbesöken med lärare och skollädaare erbjöd möjlighet att få en uppfattning om attityder, datorvana och liknande, bedömdes det för dem vara tillräckligt med en enkät efter avslutandet av studien.

Vid ett provtillfälle per skola genomfördes en deltagande observation vid vilken fältanteckningar fördes. Med ett undantag, vid A-skolan, besöktes det första provtillfället.

7.2.5 Bearbetning

Provresultaten registrerades maskinellt och översändes i Excel-format, varefter de lades in i statistikprogrammet SPSS. Även enkätsvaren registrerades i SPSS. Vid de deltagande observationerna fördes fältanteckningar vilka, liksom också den mailväxling med digitala enkäter som förekom, sparades och sammanställdes i Word-format. För lärarenkätens del lades resultaten även in i SPSS. Analysen av resultaten medger såväl jämförelser som selektiva iakttagelser på individ-, grupp-, skol- eller kommunnivå, dock givetvis med stor försiktighet, eftersom urvalen är selekterade och mycket begränsade.

Genom avtalet med Universitetet i Bergen förmedlades samarbete med det norska företaget ENOVATE AS som utvecklar nätbaserad programvara för utbildningssammanhang. Företaget skötte tekniken avseende de norska nationella proven och bistod också denna studie med distribution och löpande support.

7.2.6 Reliabilitet, validitet och generaliserbarhet

Att använda ett redan befintligt prov från ett annat land innebar en kvalitetssäkring, då materialet utvecklats, utprovats och analyserats med höga krav på reliabilitet och validitet (Moe, 2008). Universitetet i Bergen har även god erfarenhet av CBT då de, som tidigare nämnts, ansvarat för den norska versionen av DIALANG (Moe, Carlsen & Hasselgren, 2006). Väsentligt är dock att vara medveten om att det här aktuella provet, jämfört med de svenska provmaterialen, är utvecklat i en annan kursplanekontext (Utdanningsdirektoratet. *Læreplan i engelsk: LK06*) och också använder delvis andra uppgiftstyper. Eftersom studiens syfte emellertid inte är att pröva elevernas färdigheter i engelsk läsning mot de svenska kursplanerna utan att undersöka praktikalitet och attityder till datorbaserad språkbedömning som format, bedömdes detta prov fungera väl. Det norska provet kan, liksom de svenska motsvarigheterna (se 3.2 och 4.3 ovan), karakteriseras som ett summativt, kriteriere- och domänrelaterat, indirekt kunskapsprov (Council of Europe/Skolverket, 2009a).

Frågeställningarna i enkäterna i den empiriska studien återkommer i såväl identiska som något varierade formuleringar riktat till de olika respondentgrupperna, enligt så kallad triangulering (Stukát, 2005; Alvesson & Sköldberg, 2008). I de två elevenkäterna återkommer också likartade och identiska frågeställningar före och efter genomfört prov. En ankring finns även till flera års elevenkäter i anslutning till de nationella Ämnesproven i engelska för åk5, då en fråga upprepas därifrån. Varje enkät ger också utrymme för öppna kommentarer och förklaringar. Enkäterna kompletterades vidare med deltagande observation och samtal vid besök på skolorna.

En varningsklocka att beakta är att "som man frågar får man svar" (Andersson 1994; Stukát 2005) – och att man får olika svar beroende på vem som frågar vem, hur och när etc. Dörnyei (2001) problematiserar bland annat detta när det gäller motivations- och attitydforskning. Inga egentliga intervjuer kompletterar här heller elevenkäterna, vilket Dörnyei (2001) framhåller som en brist. Dessutom kan så kallad *compliance-effekt* göra att respondenter (framför allt i

ett riktat urval) svarar med tillmötesgående eller det de tror förväntas av dem. Själva deltagandet kan även i sig göra respondenter positiva, s.k. *Hawthorne effect* (Dörnyei, 2001). Min egen roll (och förförståelse) som engelsklärare, provutvecklare och representant för Skolverket har varit en förutsättning men också inneburit ett deltagande i studien som iakttagare och samordnare, vilket bör beaktas i ljuset av så kallad *ethnographic reflexivity* (Hammersley & Atkinson, 1983).

I enlighet med en så kallad *Case study approach* (Hammersley & Atkinson, 1983; Stukát, 2005; Wallace, 1998) gör den aktuella studien inte anspråk på att vara generaliserbar, utan på att representera de deltagande respondenterna. Dock kan resultaten förmodligen bära en viss sannolikhet, *probability* (Wallace, 1998) och därmed överförbarhet, *transferability* (Hammersley & Atkinson, 1983). Resultaten kan kasta (delvis nytt) ljus på ett ämne eller problem, vilket i sin tur kan leda till idéer och hypoteser som genom vidare empirisk forskning eventuellt kan generaliseras, enligt så kallad *illuminative approach* (Wallace, 1998).

7.3 Genomförande

En preliminär planering av pilotstudien redovisades i en skrivelse till Skolverket enligt nedan (Lindqvist, 2009) och i stort följdes sedan denna, med enstaka justeringar beroende på skolornas förutsättningar.

När det gäller den här pilotstudien i engelska har vi valt att ställa en förfrågan om deltagande till åtta grundskolor [...] enligt följande:

- Inledande kontakt med intresserade pedagoger under september – november 2008
- Bekräftelse med aktuella skolledare, främst i form av ett beskrivande brev
- Inledande planeringsbesök under december 2008 – januari 2009 med samtal bland annat om de tekniska förutsättningarna
- Enkäter före genomförandet riktade till elever, lärare och skolledare
- Genomförande av datorbaserade uppgifter i två klasser per skola under februari – mars 2009
- Deltagande observation under genomförandet
- Enkäter och ev. samtal efter genomförandet

Enligt uppdraget inleddes under hösten 2008 kontakter med intresserade lärare och/eller skolledare, och samtliga tillfrågade skolor visade sig intresserade av att delta. Kontakten togs först muntligt eller per e-post och följdes av ett beskrivande brev (bilaga 5).

Under januari och februari 2009 genomfördes planeringsmöten på samtliga skolor med vidtalade lärare och vid tre skolor även med skolledare. Då diskuterades det praktiska genomförandet, skolans förutsättningar vad gäller teknisk utrustning samt till exempel nödvändiga schemamässiga anpassningar. Även pedagogiska frågor och specifika frågor kring den egna skolans förutsättningar, lärarens datorvana eller den aktuella elevgruppen berördes.

Vid planeringsmötet fick lärarna en presentation av studien, med bifogad blankett för godkännande, att distribuera till vårdnadshavare. De fick också skriftlig information att delge eleverna och den inledande attitydenkäten – Elevenkät I – att genomföra innan eleverna tagit del av exempeluppgifter eller gjort provet. Kontakten med och godkännande av vårdnadshavare lämnades därmed över till skolan att ansvara för och följa upp.

Genomförandet planerades i samtliga grupper till veckorna före påsklovet, vecka 11-14 2009, och ett besök för deltagande observation vid ett provtillfälle per skola planerades in, i de flesta fall till det första provtillfället. Elevenkät II genomfördes direkt efter provet för de flesta elever och kunde distribueras vid observationstillfället, då också den inledande enkäten samlades in.

Utöver informationen vid planeringsmötena fick varje lärare, när det gäller själva provmaterialet, en utskrift av *Informasjon om nasjonale prøver i lesing i engelsk – Del 1 och 2*, för respektive årskurs (Utdanningsdirektoratet, 2008). Materialet innehöll bland annat målbeskrivningar, förklaringar till exempeluppgifter, samt en översikt över provets komposition på uppgiftsnivå. Lärarna uppmanades att förklara upplägget för eleverna och låta dem göra exempeluppgifterna på Utdanningsdirektoratets hemsida.

Inför och under genomförandet hölls sedan relativt tät e-post- och telefonkontakt med information under vägs. Varje skola skickade in digitala klasslistor, vilka lades över i Excelformat och vidarebefordrades till ENOVATE, som lade upp användarnamn och lösen för varje elev. Dessa mailades tillbaka till skolorna i PDF-format. Inför själva provlektionen fick varje lärare en detaljerad arbetsgång per e-post samt egen tillgång till själva proven via en demo-användare.

Provresultaten registrerades på individnivå i och med att eleverna avslutade sina prov och loggade ut. Dessa sammanställdes, distribuerades i Excel-format och delgavs lärarna via e-post under slutet av april. De sammanställningar på individnivå som upprättats har byggts på skolornas klasslistor (oftast med enbart förnamn och aldrig med födelse- eller personnummer). Vid bearbetning har underlaget aidentifierats med id-nr.

För lärarenkäter och skolledarenkäter användes digitalt format och dessa utväxlades per mail under vecka 14 och 15. Varje lärare ombads till slut via e-post ge respons kring elevernas prestationer på proven i början av maj.

8 Empirisk studie - resultat

Resultaten redovisas nedan i separata avsnitt. Först återges intryck från de deltagande observationerna i punktform. Därefter redovisas de fyra genomförda enkäterna med kontinuerliga summeringar; först återges elevernas feedback från de båda elevenkäterna var för sig, med en avslutande begränsad jämförelse mellan svar från de båda enkäterna, och därpå följer lärar- respektive skolledarenkäterna. Sist ges en redogörelse för elevernas resultat på de genomförda proven. Varje avsnitt avslutas med en sammanfattande kommentar.

8.1 Deltagande observation

Vid ett provtillfälle per skola genomfördes en deltagande observation. De iakttagelser som gjordes redovisades skolvis i de två åldersgrupperna till Skolverket i form av relativt utförliga, informella fältanteckningar, oundvikligen med visst fokus på det som inte fungerade helt friktionsfritt (Lindqvist, 2009). Här återges en sammanfattning i punktform av dessa anteckningar:

Vid samtliga sex skolor var eleverna mycket intresserade och engagerade. Helhetsintrycket blev att såväl lärare som elever välkomnade studien och att de generellt är både positiva och optimistiska när det gäller datoranvändning.

Samtliga besökta skolor hade datasalar eller kombinerade studie- och datasalar med väggfasta datorer, vid vilka eleverna sitter relativt tätt bredvid varandra. Generellt föreföll den tekniska standarden relativt låg och vid varje skola var någon dator ur funktion. Vid fyra skolor förutsatte datorparken att man delade en klass och därmed behövde engagera flera lärare.

Det förekom problem vid tre skolor med några elevers egna nätverksinloggningar, så att lokal tekniksupport fick kontaktas.

En skola hade två relativt moderna datasalar. Vid provtillfället tillät emellertid inte tillgänglig bredbandskapacitet att två grupper genomförde provet samtidigt och det fick avbrytas för att göras med en grupp i taget vid ett senare tillfälle.

Vid fyra skolor hade av olika skäl inte alla elever gjort demo-versionen med exempeluppgifter, vilket förmodligen var skälet till att de vid provtillfället kände viss oro inför en del uppgiftsformat. Lärares förkunskaper om provet föreföll synnerligen viktig.

Flera elever ville förstora texten på skärmen och en del elever anmärkte på funktioner i provet som till exempel hindrade att de gick tillbaka och ångrade svar, eller inte medgav att de kunde 'bläddra' framåt i provet.

Ett par lärare påpekade att deltagande elever med olika funktionsnedsättningar arbetade bättre och lugnare med dessa prov än med traditionella pappersprov.

8.2 Enkäter till eleverna

De två elevenkäterna har som syfte att fånga upp elevernas förväntningar och tankar innan de sett proven, att jämföra med deras erfarenheter och slutsatser efter provtillfället. Eleverna har på enkäterna uppgett årskurs, skola, kön samt initialer, som sedan översatts till id-nummer,

vilket möjliggör jämförelser på dessa nivåer. Några jämförelser på individnivå mellan de två enkäterna har dock inte, av tids- och utrymmesskäl, gjorts i samband med sammanställningen av denna studie.

Efter ett visst bortfall var det totalt 248 elever som besvarade någon av enkäterna och av dem gjorde 209 elever båda enkäterna. 237 elever gjorde Enkät I och 220 Elevenkät II. Det var i stort sett lika många pojkar som flickor, samt ungefär lika många elever fördelat över äldre (åk 8) och yngre åldrar (åk 5/6) som gjorde någon av enkäterna. Däremot var det nästan dubbelt så många elever från de tre storstadsskolorna; 165 elever jämfört med 83 från de tre landsbygdsskolorna.

Elevernas enkäter redovisas nedan först var för sig, följt av någon jämförelse dem emellan då samma eller liknande frågeställningar förekommer på båda enkäterna. I några fall, när det vid närmare analys visat sig finnas markanta/intressanta skillnader, redovisas jämförelser i diagramform avseende kön, kommuntillhörighet eller åldersgrupp. I huvudsak utgår dock redovisningen från hur många elever i procent som svarat positivt på att det aktuella påståendet överensstämmer med deras egen självbild. För att få fram denna procentsats har de två yttersta ja-alternativen på den femgradiga så kallade Likert-skalan slagits ihop.

8.2.1 Elevenkät I

Elevenkät I (Bilaga 1) har besvarats av 237 elever och genomfördes alltså innan elevernas uppfattningar färgats av det datorbaserade provet. Den tar upp 14 korta påståenden kring deras inställning till och kunskaper om datorer respektive engelska, samt deras tankar inför att kombinera dessa två till att ha datorbaserade prov i engelska. Dessutom innehöll enkäten en öppen fråga (5) kring deras datorvanor, samt som avslutning utrymme för öppna kommentarer och tankar.

8.2.1.1 Slutna svarsformat

Datorer

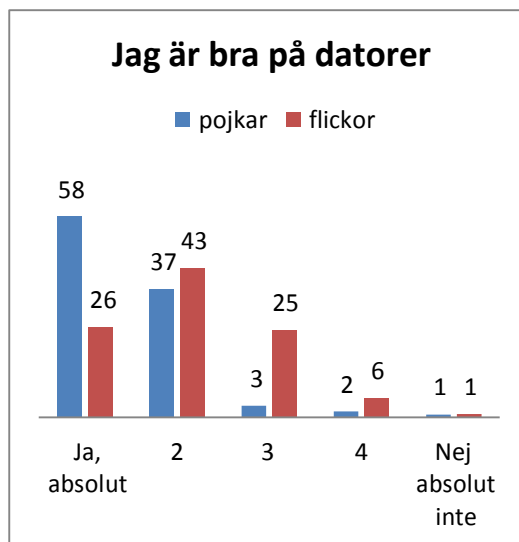
Enkätens fem första ställningstaganden berör elevernas inställning till och kunskaper om datorer, samt deras datorvanor i skolan och hemmet. En stor andel av eleverna menar att påståendena stämmer mycket väl in på dem och markerar något av de två positiva alternativen på Likert-skalan, dock med en markant lägre överensstämmelse när det gäller datoranvändning i skolan:

1. Jag gillar datorer	93%
2. Jag är bra på (att använda) datorer	82%
3. Jag använder ofta dator i skolan	35%
4. Jag använder ofta dator hemma	90%
5. Jag sitter ungefär (X) timma/r per dag vid datorn och då håller jag oftast på med (X)	

På den öppna frågan svarade eleverna att de sitter mellan 0 och 8,5 timmar per dag vid datorn och att de oftast är inne på till exempel MSN, Billdagboken eller spelar olika onlinespel. I

genomsnitt uppskattar dessa elever att de tillbringar drygt två timmar vid datorn hemma varje dag. Fem elever av 237 har skrivit 0 timmar/dag.

Vid en uppdelning av respondenterna med avseende på kön, framträder här relativt stora skillnader. Som framgår av figur 3 har pojkarna större självförtroende när det gäller att använda datorer (påstående 2, ovan):



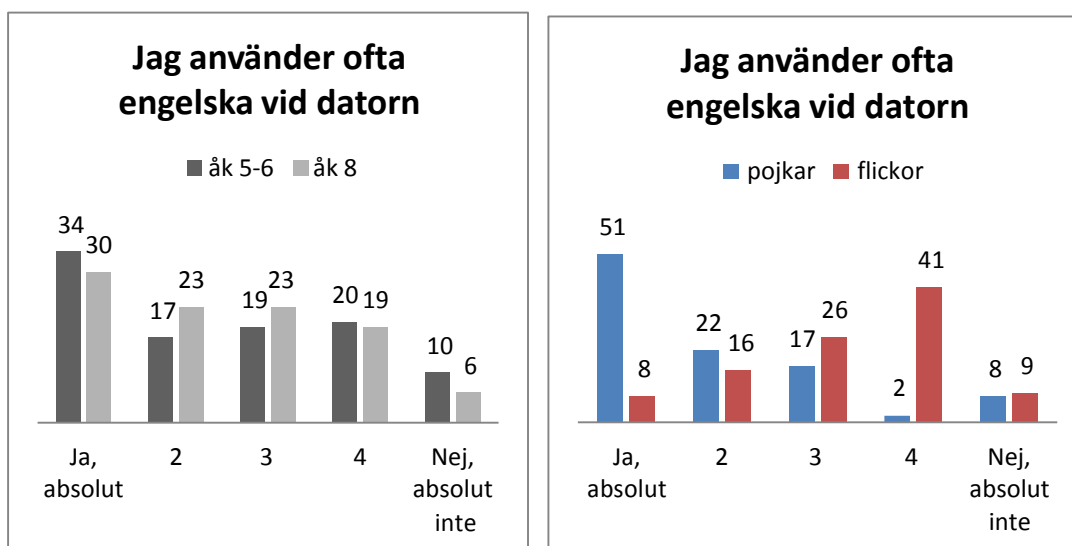
Figur 3: Elevernas självvärdering på fråga 2 "Jag är bra på (att använda) datorer" (Svar uppdelat på kön och i procent av totalt 237 elever)

Ämnet engelska

Vid en analys av enkätsvaren visar det sig att de elever som deltog i undersökningen tycker om ämnet engelska, att de har ett gott självförtroende i ämnet samt att ungefär hälften av dem har erfarenhet av att använda engelska vid datorn, även när det gäller pedagogiska sammanhang:

6. Jag gillar engelska	77%
7. Jag är bra i engelska	71%
8. Jag använder ofta engelska vid datorn	51%
9. Jag har gjort språkövningar på datorn förut	48%

En ytterligare granskning av svaren på fråga 8, huruvida eleverna använder engelska vid datorn, visar inte några stora skillnader mellan åldersgrupperna men däremot markanta könsskillnader (se figur 4 och 5 nedan). Ett mycket snarlikt påstående förekom i elevenkäten till de nationella proven i engelska för årskurs 5 (*I use English when I do things on a computer*) och visar där liknande mönster enligt Skolverkets ämnesrapport: "Mötet med engelska i samband med datoranvändning ökar för varje år. I år uppger 65% av eleverna att de använder engelska i sådana sammanhang. Bland dessa elever finns dubbelt så många pojkar som flickor" (Skolverket, 2008, s.22).



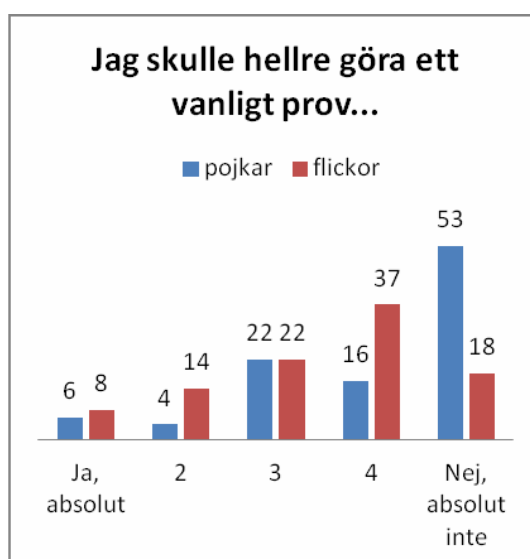
Figur 4 och 5: Elevernas svar på fråga 8 "Jag använder ofta engelska vid datorn" (Svar uppdelat på åldersgrupp respektive kön, i procent av totalt 237 elever)

Engelskprov på dator

Redan innan eleverna sett och provat på de datorbaserade proven är de allra flesta mycket positiva till tanken och känner tillförsikt. Några av dem skulle dock föredra ett traditionellt pappersprov och ungefär lika många uttrycker genom sina svar att de känner en viss osäkerhet:

- | | |
|--|-----|
| 10. Jag tycker det verkar bra att ha engelskprov på dator | 86% |
| 11. Jag skulle hellre göra ett vanligt prov med papper och penna | 16% |
| 12. Jag behöver hjälp för att kunna göra prov på dator | 15% |
| 13. Jag känner mig beredd på att göra det här provet | 83% |

Figur 6 visar att skillnaden åter är ganska stor mellan flickors och pojkars inställning till att hellre göra vanligt prov med papper och penna:



Figur 6: Elevernas svar på fråga 11 "Jag skulle hellre göra ett vanligt prov med papper och penna" (Svar uppdelat på kön, i procent av totalt 237 elever)

Två påstående relaterar till elevernas känsla inför sina prestationer på prov i engelska. Fler än hälften brukar vara nöjda med sina prestationer på engelskprov. Dock verkar svaren tyda på att en del elever ser en möjlighet att prestera ännu något bättre på datorbaserade prov:

14. Jag brukar vara nöjd med hur det går för mig på prov i engelska	65%
15. Jag tror att jag kommer att bli nöjd med hur det går för mig på det här provet	74%

8.2.1.2 Öppna kommentarer

Av de 237 eleverna som fyllt i den första elevenkäten har 99 stycken lämnat öppna kommentarer om att ha datorbaserade prov i skolan, som komplement till påståendena. Kvantitativt sett har de till dessa kommentarer sammanlagt använt ca 1380 ord (dvs. i genomsnitt ca 14 ord per person). Vid en innehållslig analys framträder några tydliga undergrupper till de två större kategorierna, positiva respektive mera tveksamma kommentarer, vilket redovisas nedan med inslag av autentiska citat. Några elever har lämnat mer neutrala upplysningar som redovisas kortfattat i en egen kategori. Naturligtvis förekommer också viss överlappning och tolkningsproblematik när det gäller en sådan generalisering.

Positiva kommentarer

Det var kul – lättare – bäst...

Hälften av de öppna kommentarerna, 48 stycken av 99, uttrycker en odelat positiv inställning till att ha datorbaserade prov i engelska i skolan. De allra flesta skriver kort och gott att det verkar bra eller mycket bra, och många uttrycker också att det ska bli roligt eller roligare att göra ett sådant prov. En del elever motiverar sig något och gör då en implicit eller uttalad jämförelse med pappersprov, och några elever menar att det är lättare eller går snabbare eller smidigare att göra prov på datorn. Ett antal elever relaterar till den egna åldersgruppen och menar att det passar dem att göra datorbaserade prov. Ytterligare andra verkar eventuellt i första hand tycka det ska bli roligt att prova på att göra datorbaserade prov.

- *Jag tycker att det är bra med prov på datan för att man kanske har kul samtidigt som man har prov.*
- *Prov på datorn äger*
- *Jag tror att jag kommer se fram mot engelskproven nu mera*

Att skriva...

Vid analysen av elevernas reflektioner framträder aspekten skriva som viktig att fokusera för ett tiotal elever. Hälften av dem menar att det är odelat bättre att få skriva på dator, medan några andra problematiserar kring att de/man eventuellt behöver skriva för hand – för att träna till exempel sin handstil eller för att det kan vara viktigt för ämnets skull. Enstaka elever tycker att det vore bättre att få skriva för hand, om man inte är så snabb vid tangentbordet.

- *Det kanske skulle vara bättre i och med att man slipper vässa pennor, suddas m m*

- *Det är nog bra bara för vissa kanske dåliga på att stava och får till orden bättre då. Det tycker jag är en stor möjlighet.*
- *Det är bra för dom som inte gillar att skriva men jag tycker ändå att det är viktigt att man kan skriva bra engelska på papper.*
- *Man får inte kramp i händerna*

Tveksamma eller negativa synpunkter

Att använda dator...

Förutom de elever ovan som skriver om datorn som skrivverktyg, är det ytterligare några som tar upp datoranvändning, och som är mer tveksamma till andra aspekter. Några elever tar upp skillnad i förutsättningar beroende på datorvana och några elever efterlyser IT-utbildning för elever, men eventuellt också för lärare. Det förekommer också ett par andra reflektioner, till exempel när det gäller skärm-/migränkänslighet.

- *Det kan vara svårt för dem som inte använder datorer ofta och som inte riktigt förstår. Men annars verkar det bra.*
- *Det värkar bra män det hade varit bra att ha lektioner i datakunighet*
- *Man kan bli beroende*
- *Det kan vara jobbigt att se på en skärm som "brusar" för migrän känsliga om det skulle vara så.*

Att fuska...

En grupp elever väljer att skriva något om att de ser nackdelar kring möjligheterna att fuska vid datorbaserade prov. De verkar ha vissa farhågor om att det kan bli svårt att genomföra proven under sådana förhållanden som de förväntar sig vid prov. Flera elever utvecklar sina kommentarer och en elev tar upp möjligheter att till exempel gå ut på Internet för att få hjälp på ett prov, medan de andra skriver om att det är lätt att se varandras prov när man sitter i skolornas datasalar.

- *Vid en dator kan du oftast komma ut på internet eller nå filer vilket kan leda till fusk.*
- *Vad som inte är så bra är att om alla gör provet i en datasal är det lätt att tjuvkika på grannens prov.*

(Vissa) förbehåll ...

Några elever har förbehåll som handlar om att datorbaserade prov kanske kan vara bra under förutsättning att instruktioner är tydliga och att proven är lättlästa och inte för svåra. En enda elev i hela underlaget är odelat negativ i sin kommentar och två elever verkar helst ha det som det är.

- *Blä*
- *Det kan vara bra men det är också dåligt för vi har det ju bra så det är kanske lite komplicerat.*
- *Om de ska finnas prov på datorer tycker jag de ska vara lättläst*

Upplysningar...

Sju elever har valt att lämna upplysningar de vill förmedla om datorer, sig själva eller sin skola som de tycker berör ämnet, som till exempel:

- *Det är väl helt okej, jag har gjort egna slags tester/prov på datorn förut med programmeringsspråket PHP. Och jag vet att det funkar bra.*
- *I skolan tycker jag att alla ska ha en dator som man får av skolan. Som i it gymnasium.*
- *Jag gillar eng för det låter kul när man pratar eng. Jag gillar inte så mycke dator för jag vill ha andra fritid*

8.2.2 Elevenkät II

Den andra elevenkäten, Elevenkät II (bilaga 2), besvarades i de flesta elevgrupperna av eleverna direkt efter det att de genomfört det datorbaserade provet. Den besvarades av 220 elever, alltså något färre än antalet (237 elever) som gjorde den första enkäten på grund av ett visst bortfall (209 elever gjorde båda enkäterna). Enkäten tar upp 16 påståenden, varav några har direkt anknytning till eller är likadana som påståenden på den första enkäten. Generellt syftar denna elevenkät till att fånga upp elevernas utvärdering - deras nyvunna erfarenhet av, upplevelse av och attityder till att göra ett datorbaserat prov, samt möjliggöra vissa jämförelser med deras tankar innan de gjort utprövningen. Slutligen fanns också möjlighet för eleverna att lämna öppna kommentarer eller reflektioner.

8.2.2.1 Slutna svarsformat

Tillgänglighet och likvärdighet

De fyra första påståendena på Elevenkät II vill fånga upp den omedelbara upplevelsen av och attityden till att göra proven. De inbjuder eleverna att ta ställning till om de tyckte om att göra provet samt om de kände sig osäkra inför tillvägagångssätt och uppgiftstyper, men berör också deras uppfattning om likvärdighet när det gäller datorbaserad prov.

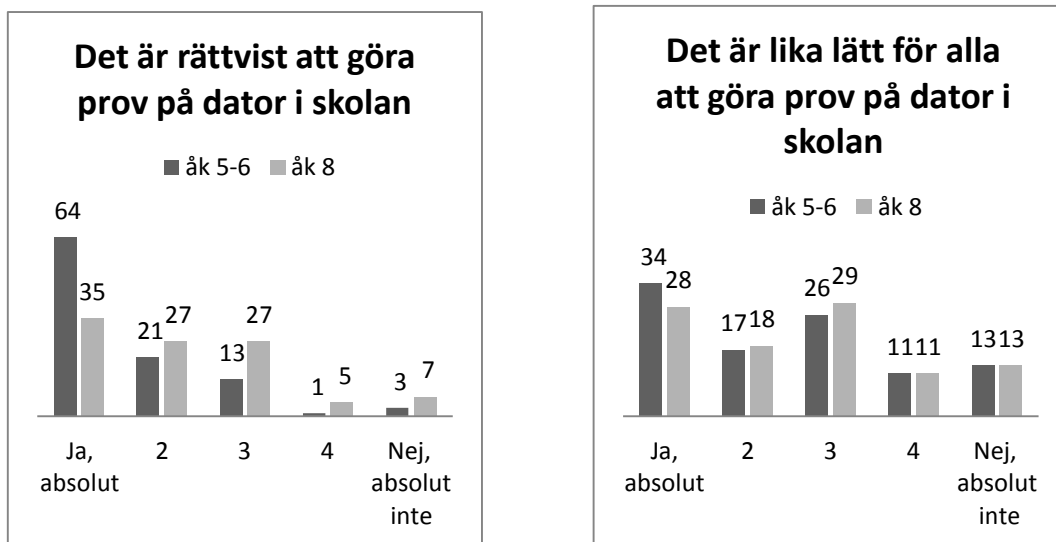
De allra flesta tyckte om att göra provet, att det kändes lättillgängligt och ganska få elever upplevde att de för egen del hade behövt mer hjälp. Dock menar hälften av dem att det inte är lika lätt för alla att göra datorbaserade prov. Påståendet kring rättvisa kan naturligtvis tolkas på en mängd sätt, men indikerar hur elever problematiserar kring begreppet och kan då komplettera frågor kring tillgänglighet. Eleverna verkar å ena sidan inte tycka att datorbaserade prov innebär någon stor risk för ojämlikhet eller särbehandling ur rättvisesynpunkt, men å andra sidan kan i detta sammanhang andelen som markerar Nej-alternativ eller skalans mittalternativ, 27 respektive 51 %, anses vara anmärkningsvärd.

1. Jag gillade att göra engelskprov på dator	71%
2. Det var lätt att förstå hur man skulle göra	75%
3. Det var till hjälp att göra exempeluppgifterna innan ¹	56%
4. Jag skulle ha behövt mer hjälp för att kunna göra prov på dator	10%

¹ De exempeluppgifter som fanns att tillgå före provet låg på norska Utdanningsdirektoratets hemsida, men fungerade inte för alla grupper och 78 elever har inte kunnat besvara nr.3.

- | | |
|--|-----|
| 12. Det är rättvist att göra prov på dator i skolan | 73% |
| 13. Det är lika lätt för alla att göra prov på dator | 49% |

Som framgår av figur 7 och 8 ser svaren ut på följande sätt vid en spridning av elevsvaren på påståenden 12 och 13 över hela den femgradiga skalan:



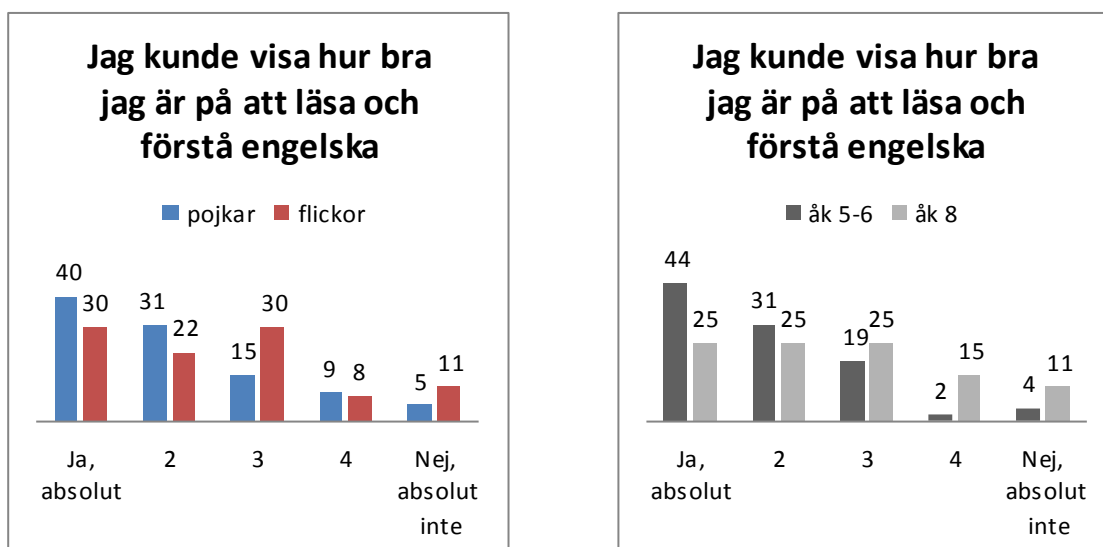
Figur 7 och 8: Elevernas svar på fråga 12 "Det är rättvist att göra prov på dator i skolan" respektive 13 "Det är lika lätt för alla att göra prov på dator" (Svar i procent av totalt 220 elever, uppdelat på åldersgrupp)

Provsituation och lärande

Nästa grupp påståenden handlar om att belysa elevernas upplevda tillfredsställelse under och med själva provsituationen, samt hur de ser på provets didaktiska kvaliteter och sin egen prestation. Eleverna verkar generellt tycka att det kändes bra, att de inte var nervösa eller hade svårt att koncentrera sig, men ungefär 10 % av dem besvarar dessa utsagor negativt. När det gäller huruvida det kändes som ett bra lärtillfälle och ett rättvisande provtillfälle svarar i det ena fallet nästan hälften av dem och i det andra fallet drygt hälften att de upplevde provet så.

- | | |
|--|-----|
| 5. Jag kunde koncentrera mig när jag gjorde provet | 75% |
| 6. Jag kände mig lugn när jag gjorde provet | 81% |
| 7. Jag kände mig nöjd när jag gjorde provet | 73% |
| 8. Jag lärde mig något när jag gjorde provet | 48% |
| 9. Jag kunde visa hur bra jag är på att läsa och förstå engelska | 63% |
| 10. Jag tror att jag kommer att bli nöjd med mitt resultat på det här provet | 62% |

Vid närmare analys av påstående 9 ovan fördelat över hela Likert-skalan och kön respektive åldersgrupp, som illustreras i figur 9 och 10, framkommer att äldre elever och flickor är mer tveksamma än yngre elever och pojkar till om provet gör rättvisa åt deras kunskaper:



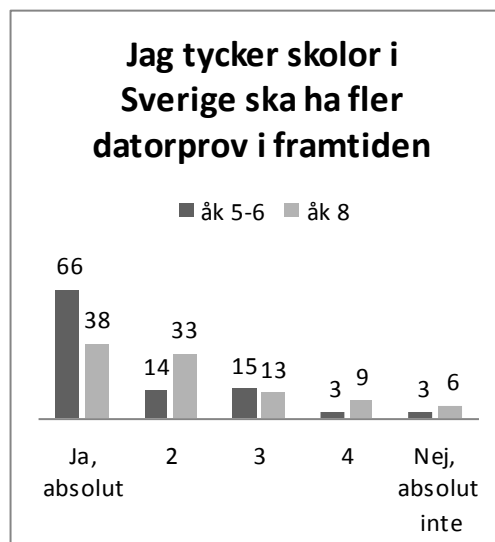
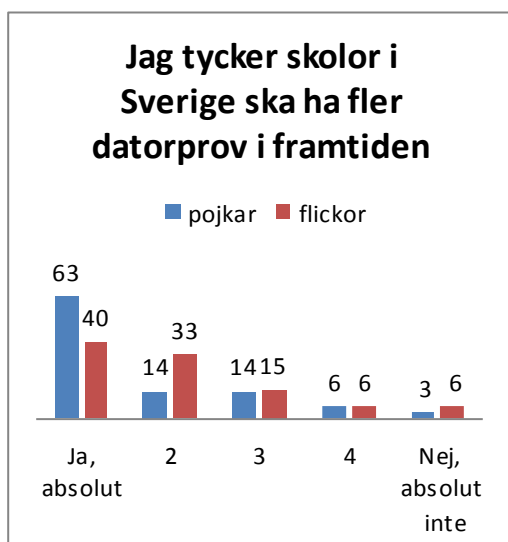
Figur 9 och 10: Elevernas svar på fråga 9: "Jag kunde visa hur bra jag är på att läsa och förstå engelska" (Svar uppdelade på kön respektive åldersgrupp i procent av totalt 220 elever).

Elevernas slutsatser och framtiden

Fyra frågor handlar om elevernas helhetsintryck och sammanfattande åsikter kring de datorbaserade provens vara eller icke vara, på deras hemmaplan och i deras framtid. Tre fjärdedelar av eleverna är positiva eller mycket positiva till datorbaserade prov i engelska och det samma gäller genomförandet på den egna skolan. 46 elever av 220 föredrar att göra vanliga pappersprov. Vid en närmare granskning av den sista frågan framkommer att pojkar är mer positiva än flickor, och att åk 8 är lite mer tveksamma än åk 5/6 inför tanken att införa fler datorbaserade prov i framtiden (se diagram nedan).

11. Det fungerade bra på min skola att göra prov på dator	74%
14. Jag tycker det är bra att ha engelskprov på dator	78%
15. Jag gör hellre vanliga prov med papper och penna	22%
16. Jag tycker skolor i Sverige ska ha fler datorprov i framtiden	76%

Spridningen över hela den femgradiga skalan för påstående 16 avseende flickor och pojkar samt åldersgrupp visar i figur 11 och 12 några skillnader:



Figur 11 och 12: Elevernas svar på fråga 16. "Jag tycker skolor i Sverige ska ha fler datorprov i framtiden" (Svar uppdelade på kön respektive åldersgrupp i procent av totalt 220 elever)

8.2.2.2 Öppna kommentarer

Efter de 16 påståendena fick eleverna uppmaningen: "Förklara eller kommentera gärna dina svar – eller skriv annat om datorbaserade engelskprov". Det är fler elever som lämnat öppna synpunkter i den andra enkäten, än i den första. Av totalt 220 elever har 124 stycken skrivit sammanlagt 2160 ord (i genomsnitt drygt 17 ord per elev), vilket är något mer än i Enkät I. Vid en kategorisering av deras kommentarer är det möjligt att urskilja några olika innehållsliga grupper. De flesta eleverna verkar fundera kring datorbaserade prov och användningen av sådana generellt, medan andras kommentarer handlar om det just genomförda provet. Elevernas reflektioner redovisas nedan i undergrupper för att bli överskådliga och med inslag av autentiska citat. Naturligtvis kan viss överlappning och tolkningssvårighet förekomma.

Generella reflektioner kring datorbaserade prov

Motivation

Drygt en tredjedel av elevernas kommentarer berättar om hur de upplever en ökad motivation när de föreställer sig möjligheten att göra datorbaserade prov. Till övervägande del skriver de att de tycker det är roligt, bra och att de vill ha fler liknande prov, men det finns också en relativt stor grupp som utvecklar sina omdömen ytterligare och menar till exempel att det blir intressantare, lättare att koncentrera sig, inte så nervöst och skönt att slippa skriva. Endast två elever av 124 i hela urvalet öppna kommentarer är enbart negativa, utan att på ett konstruktivt sätt motivera eller precisera sig.

- Jag tycker att det var as bra! Och jag ser fram till det i framtiden
- Jag tyckte att det var jätteroligt att göra engelska prov på datorn. Det blir enklare för det är kul. Det kan bli tröttsamt att sitta med penna och stirra på ett papper. Jag tycker defenitivt att vi ska ha prov på datorn.
- Det är lättare att förstå det kommer bilder. Man behöver inte få ont i handen när man skriver. Det är bättre tyckte jag.

- *Det vore något nytt och bra att ha engelskprov på dator, man blir mer intresserade på språket*
- *Kunde inte koncentrera mig, det tog tid och det var tråkigt.*

Teknik och miljö

Drygt 20 elever tar upp problem som relaterar till datoranvändning i allmänhet och miljön i skolornas datasalar i synnerhet. De flesta tar upp att de sitter tätt bredvid varandra och därför lätt blir störda eller ser varandras skärmar och kan fuska. Ungefär tio elever skriver specifikt om att det är tröttande att läsa på datorskärmen och ytterligare andra om obekväma stolar eller att program kan haka upp sig. De allra flesta elever väver dock in positiva kommentarer. En elev uttrycker också att datorbaserade prov är mer miljövänliga i vidare mening.

- *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.*
- *Om man hade bättre stolar att koncentrera sig utan att få ont*
- *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.*
- *Det är miljövänligt och inget slöseri med papper!*

Förutsättningar

En andel kommentarer och reflektioner, ca 20 %, belyser elevernas funderingar kring sina egna och andras olika förutsättningar att genomföra datorbaserade prov, i både positiva och negativa hänseenden. Några verkar inspirerade av enkätfrågorna och anknyter till formuleringar i dem, men flera tar också upp nya infallsvinklar. De resonerar kring datorvana, ämneskunskaper och 'läs- och skrivsvårigheter' och menar att det finns olika beredskap och förmåga att göra proven. Några tar upp provens domäntäckning och variation, som till exempel att inte få visa andra kunskaper inom ämnet än det provet testas och önskemål om att variera med skriftliga pappersprov. Någon elev berör bedömning och tycker att det verkar lättare på papper. En elev tar upp att det är bättre om man kan förbereda sig inför provet.

- *Frågan: "Det är lika lätt för alla att göra prov på dator" det beror på vad man har för data kunskaper. När man gör det här provet kan man ju inte visa hur bra man är på att läsa och uttala allt på engelska*
- *Jag tycker att det var kul men i vissa ämnen kan man väl ha skriftliga prov*
- *Jag fattade typ ingenting, man har ingenting att träna till så nej jag gilla det inte*
- *Det blir bra. Jag är dyslektiker och för mej så kändes det bra.*
- *Det kan både vara bättre och sämre att göra det på dator, för att man kanske kan utvärdera mer när man har skriftligt.*

Synpunkter på de aktuella proven

Svårighetsgrad

Ungefär 30% av de 124 öppna kommentarerna verkar främst handla om proven som sådana. De flesta av eleverna har synpunkter på svårighetsgraden. Kommentarer varierar mellan att det var över eller under deras nivå, med ungefär jämn fördelning. Några elever går in mer i detalj och tar till exempel upp svåra ord eller textmängd, men majoriteten tillägger att de är positiva till proven.

- *Jag lärde mig inte nåt, jag kunde allt innan. Men kommer det ett prov på 6:ans svårighetsgrad lär jag mig säkert nåt. Vill göra fler prov på datorn.*
- *Jag hade inte lärt mej dom orden som stod där ibland så de var lite svårt. T.ex enjoj, vad betyder de?*
- *Det va coolt men krånglit*
- *Det var ganska kul men väldigt lätt. Kanske lite för lätt, men jag vet inte*
- *Det var väl bra. Bara lite jobbigt med så långa och många texter.*

Grafik och teknik

Ett tiotal elever kritiserar funktioner i provet och har också konstruktiva förslag på förbättringar. Flera av dem har råkat ut för att inte kunna gå tillbaka i provet och rätta till eller kontrollera sina svar, vilket bör ha fungerat inom ramen för en uppgift, men inte mellan uppgifter. Några efterlyser information om hur långt man kommit i provet. Detta ges också i de norska originalproven men funktionen fungerade inte vid alla provtillfällen i denna studie. En elev har synpunkter på bland annat bilder och innehåll.

- *Man borde se hur mycket man har kvar tex 14 av 20 sidor kvar*
- *Jag tycker att det är värdelöst att man inte kan gå tillbaka och ändra svar som man skrev t ex i början. Det kan man med papper och penna!! Annars var det bra.*
- *Billigt gjort. Jag tycker att det är tråkiga bilder och de måste vara bättre tecknade bilder. De måste vara roligare teman helt enkelt. Och så tycker jag att skolans internet måste vara bättre innan man kan fortsätta att göra såna här prov.*

8.2.3 Elevenkät I och II

Eftersom flera påståenden är helt eller i stort sett identiska på de båda enkäterna möjliggörs jämförelser mellan hur eleverna uppfattade något innan de sett eller gjort proven, samt hur de utvärderar (i stort sett) samma frågeställning efteråt. Här redovisas i korthet tre sådana möjliga jämförelser.

Jag tror att jag kommer att bli nöjd...

En skillnad mellan de båda enkäterna framkommer när eleverna uttalar sig om huruvida de tror att de kommer att vara tillfreds med sin prestation. Påståendena är inte formulerade identiskt, då ordet *resultat* i den andra enkäten antagligen direkt associerar till poäng på provet, medan *hur det går för mig* förmodligen kan tolkas gälla själva genomförandet av provet. Med detta i beaktande är det dock tydligt att andelen elever som instämmer, och tror att de kommer att bli nöjda, sjunker i Elevenkät II (se nedan). Möjligen påverkas resultatet av att den andra enkäten gjordes i direkt anslutning till provet. Någon närmare analys av vilka elever som blivit mer osäkra görs dock inte här.

Enkät I

15. Jag tror att jag kommer att bli nöjd med hur det går för mig på det här provet (74%)

Enkät II

10. Jag tror jag kommer att bli nöjd med mitt resultat på det här provet (62%)

Dator eller papper och penna...

Två par utsagor som har direkt anknytning till studiens syfte att ta reda på attityder till datorbaserade prov jämfört med traditionella prov är fråga Enkät I:10 och II:14 respektive I:11 och II:15 enligt nedan:

Enkät I

10. Jag tycker det verkar bra att ha engelskprov på dator (86 %)

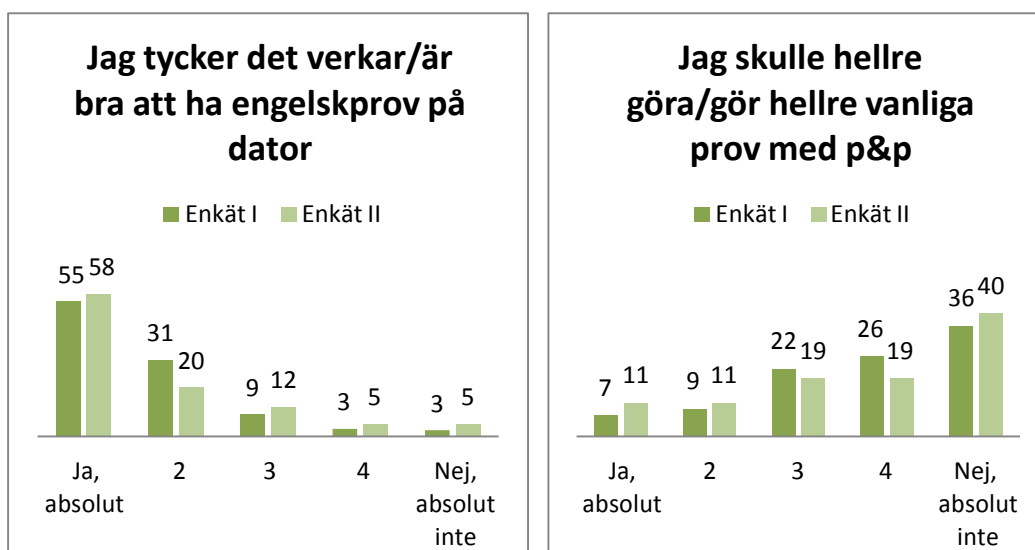
11. Jag skulle hellre göra ett vanligt prov med papper och penna (16 %)

Enkät II

14. Jag tycker det är bra att ha engelskprov på dator (78 %)

15. Jag gör hellre vanliga prov med papper och penna (22 %)

En närmare granskning (se nedan) visar dock att det, bakom procentsatserna ovan, döljer sig att något fler elever efter utprövningen blir (ännu mer) positiva till datorbaserade prov i jämförelse med vanliga papper-och-penna-prov, men också att något fler än i Elevenkät I svarar negativt. Liksom när det gäller den tidigare frågan har eleverna blivit lite mer försiktiga. Men, de har också blivit något mer säkra på sin uppfattning efter det att de fått en viss erfarenhet och genomfört de datorbaserade proven, då svaren tenderar att bli fler i vardera ytterkanterna av skalan i Elevenkät II, enligt figur 13 och 14:



Figur 13 och 14: Jämförelse mellan Elevenkät I och II: Elevernas svar på fråga I:10 och II:14 respektive I:11 och II:15 (Svar i procent elever, per skalsteg och enkät).

8.2.4 Sammanfattande kommentar

Eleverna har lagt ner stort arbete i sina enkäter på att ge konstruktiv feedback, samt på att i de öppna kommentarerna klä sina reflektioner i ord och tillföra ytterligare värdefulla aspekter som inte tas upp i de givna påståendena. Att sammanfatta i korthet är en grannliga uppgift, men anmärkningsvärt är att även om de har vissa farhågor och invändningar, avslutar de allra flesta sina kommentarer med ett positivt omdöme om utökad datorbasering av provmaterial i skolan. De tycker generellt att det fungerade bra att göra proven, att det var roligt, ungdomsanpassat, motivationshöjande och lättare (framför allt att inte skriva),

men de funderar också insiktsfullt kring såväl lärande- och likvärdighetsaspekter som tekniska och lokalmässiga förutsättningar på skolorna. Deras synpunkter på proven som sådana handlar ofta om svårighetsgrad och textmängd. Många elever tyckte proven var lätta, men flera menade också att det var mycket text och att det kan vara jobbigt att läsa på skärmen. Dessa och flera av deras synpunkter kring proven sammanfaller med norska resultat från utvärderingar med elever (Moe, 2009). Elevernas synpunkter och farhågor kan också speglas i aktuell motivationsforskning (avsnitt 5), liksom i de frågor som tas upp i relation till aktuell bedömningsforskning och erfarenheter från andra länder kring datorbaserad (avsnitt 6), vilket också berörs i avsnitt 9.

8.3 Enkäter till lärarna

De tretton lärare som involverades är alla lärare i engelska eller klasslärare som undervisar i engelska. Av dessa tretton lärare besvarade tolv den enkät som skickades ut elektroniskt efter genomförandet. En lärare valde att inte besvara den eftersom hon endast var med under planeringsstadiet. Antalet deltagande lärare är för litet för att kunna utgöra grund för generaliseringar, och de är också medvetet valda och har förklarat sig villiga att medverka. Trots detta kan dessa lärare genom sina ställningstaganden belysa situationen på sina respektive skolor, samt förmodligen indikera vissa tendenser.

Sex lärare var involverade i genomförandet av prov med yngre elever, åk 5 och åk 6, och sex lärare med äldre elever i åk 8. Sju av dem representerar landsbygdskommunen och fem storstadskommunen. Samtliga lärare kontaktades personligen och var positiva till att delta i studien med sina elevgrupper.

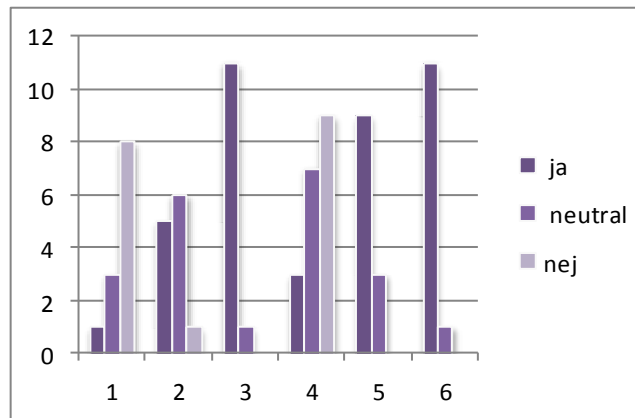
Den digitala enkät (Bilaga 3) lärarna besvarade efter studiens slut, när samtliga elever genomfört proven, bestod av två delar. Den innehöll en slutna del, i vilken de på en femgradig Likert-skala (*Ja, absolut – Nej, absolut inte*) tog ställning till sex påståenden kring datoranvändning och datorbaserade prov. Enkäten innehöll också en del som gav utrymme för öppna, positiva respektive negativa kommentarer, samt sådan information om den egna elevgruppen som lärarna bedömde skulle kunna vara av intresse för tolkningen av resultatet. Den senare redovisas nedan, i avsnittet om elevernas resultat på det genomförda provet.

8.3.1 Slutna svarsformat

Lärarna ombads ta ställning till nedanstående sex påståenden:

1. Jag använder ofta datorer i min engelskundervisning
2. Jag kommer att använda fler datorbaserade övningar och prov i min undervisning
3. De datorbaserade proven var lätta att förstå sig på och använda
4. På min skola finns tillräcklig beredskap att genomföra datorbaserade prov
5. Jag tycker det finns pedagogiska fördelar med datorbaserade prov
6. Jag skulle ställa mig positiv till datorbaserade nationella prov

Figur 15 visar hur de tolv lärarna svarat. För tydlighetens skull har de två svarsalternativen i vardera ytterkanten av Likert-skalan slagits ihop till ett ja- respektive ett nej-alternativ.



Figur 15: Antal lärares markeringar per sex påståenden i lärarenkäten, enligt ovan

Som framgår av lärarnas individuella värdering av påståendena är det en majoritet som inte använder datorer i sin engelskundervisning, medan fem lärare menar att de kommer att göra det oftare efter studien. Med ett undantag menar lärarna att de datorbaserade proven var lätta att förstå sig på och använda. Något mer än hälften av lärarna anser inte att deras skola har tillräcklig beredskap att genomföra datorbaserade prov, men de flesta är efter studien positiva till, samt ser pedagogiska fördelar med, datorbaserade prov i engelska.

Jämförelser inom lärargruppen

För att försiktigt undersöka eventuella skillnader inom lärargruppen delades respondenterna upp i lärare för yngre respektive äldre åldrar, samt lärare i de två kommunerna, storstad respektive landsbygd.

Lärare i yngre respektive äldre åldrar

När lärarnas ställningstaganden fördelas med avseende på vilka åldrar de undervisar framkommer skillnader framför allt när det gäller tre påståenden:

2. Jag kommer att använda fler datorbaserade övningar och prov i min undervisning
4. På min skola finns tillräcklig beredskap att genomföra datorbaserade prov
6. Jag skulle ställa mig positiv till datorbaserade nationella prov

Det är lika många lärare för yngre som för äldre åldrar som markerar att de inte brukar använda datorer i sin undervisning, men det är fler lärare i yngre åldrar som är beredda att göra det oftare efter studien. Lärarna för de yngre eleverna är mer tveksamma till om deras skolor har tillräcklig beredskap att genomföra datorbaserade prov. På frågan om de ställer sig positiva till datorbaserade prov svarar fem lärare för de yngre åldrarna det yttersta alternativet *Ja, absolut*, medan de sex lärarna i äldre åldrar möjligen inte är fullt så positiva, då de fördelar sina ställningstaganden jämnt mellan de två möjliga positiva alternativen.

Lärare i storstad respektive landsbygd

En geografisk uppdelning inom lärargruppen visar på vissa skillnader när det gäller fyra påståenden:

1. Jag använder ofta datorer i min engelskundervisning
4. På min skola finns tillräcklig beredskap att genomföra datorbaserade prov

5. Jag tycker det finns pedagogiska fördelar med datorbaserade prov
6. Jag skulle ställa mig positiv till datorbaserade nationella prov

Något fler av lärarna i landsbygdsskolorna använder ibland datorer i sin undervisning, men de är också mer oroliga än sina storstadskollegor för sina skolors beredskap att genomföra datorbaserade prov. Lärargruppen i storstadsskolorna är något mer positiva till datorbaserade prov och anser i något högre grad att de innebär pedagogiska fördelar, då fem av dem markerar det yttersta alternativet *Ja, absolut* vid båda dessa påståenden. Samtliga fem lärare i landsbygd är också positiva, men fördelar sina markeringar mellan de två positiva svarsalternativen.

8.3.2 Öppna kommentarer

Lärarenkäten erbjöd två skrifvärt för öppna kommentarer, ett för positiva och ett för negativa. I en inledande text ombads lärarna lämna synpunkter utifrån studiens syfte och mål, men också kring annat som de ville lyfta fram. De ombads också delge ytterligare information som de bedömde vara till stöd för vår tolkning av resultaten.

Positiva kommentarer - förväntningar

Tio av tolv lärare har skrivit i fältet för positiva kommentarer. Vid en kategorisering av deras reflektioner framträder fyra innehållsliga grupper. I något fall går de naturligtvis i varandra eller det kan vara en tolkningsfråga vad läraren egentligen syftar på. De fyra kategorierna redovisas nedan med inslag av citat ur lärarenkäterna.

Eleverna är positiva

De flesta kommentarerna handlar om att eleverna är positiva till de datorbaserade proven, eller möjligen till själva studien kring datorbaserade prov. Åtta lärare påpekar detta, även om två av dem uttrycker att eleverna alltid är positiva till datorer. Två av dem anger en förmodad förklaring till elevernas inställning:

- *Eleverna positivt inställda. Jag tror eleverna uppfattar datorprov som mer 'rättvisa' kanske*
- *Det var uppskattat av såväl eleverna som mig. Roligt! Jag tror att man kan få med sig fler elever som i vanliga fall inte gillar att skriva för hand.*

Det är smidigt för läraren

Lika många lärare, åtta stycken, berör sin egen situation och pekar på arbets- och tidsbesparing. En lärare skriver helt enkelt att det är "administrativt smidigt" och menar möjligen ett vidare perspektiv än bara lärarens. Fyra lärare nämner emellertid specifikt lättnaden i rättningsarbete som en fördel, men även andra aspekter förekommer. Två lärare menar att datorbasering erbjuder goda möjligheter i form av snabb feedback för eleverna och objektiv bedömning.

- *Att man kunde genomföra provet smidigt och lätt vid olika tidpunkter utan att behöva härja med pappersprov. Bra att man inte behöver hålla ordning på buntar med prov. Super att man slipper rättningsarbetet givetvis.*

Datorbaserade prov – och framtiden

Det finns nio kommentarer som direkt eller indirekt relaterar till det genomförda provet som sådant. En lärare betonar att det var bra att få exempeluppgifter att träna på innan och två lärare talar om uppgifterna som elevnära respektive roliga. Några kommentarer har mera karaktären av reflektioner kring framtiden. Två av dem handlar om att de genomförda proven inte täcker allt som bör testas vid ett nationellt prov och två om möjligheten att ha adaptiva prov.

- *Mycket bra DELVIS vid nationella prov – den skriftliga biten måste göras för hand. Jag tror att det kommer vara ett bra komplement till andra test.*

Negativa kommentarer — farhågor

De synpunkter lärarna valt att skriva i fältet för negativa kommentarer är nästan dubbelt så omfattande som de positiva (449 respektive 235 ord, dvs. i medeltal drygt 37 resp. drygt 19 ord per lärare). Även här har tio av tolv lärare lämnat reflektioner. Vid en kategorisering av dessa framträder ytterligare fyra innehållsliga grupper och även här kan de naturligtvis i någon mån överlappa varandra. Kategorierna redovisas på samma sätt som ovan.

Datortekniska problem på skolan

När det gäller lärargruppens farhågor är de i stort överens om att den egna skolans datorpark ofta är otillräcklig och opålitlig. Åtta av de tio lärarna som lämnat kommentarer skriver om detta. Tre lärare påpekar att en del av datorerna (ibland) är ur funktion eller att de helt enkelt är för få för att kunna genomföra med en undervisningsgrupp samtidigt. De menar också att det förekommer att datorerna låser sig, att Internetuppkopplingar och programåtkomst går mycket långsamt eller att elevernas egna inloggningar i det lokala nätverket krånglar. Några av lärarna belyser också att de råkar ut för liknande bekymmer under genomförandet av proven.

- *Vissa datorer fungerade mycket långsamt. En fjärdedel av våra datorer gick inte att använda.*
- *Två elever kunde inte genomföra provet då datorn låste sig mitt i programmet. En elev kunde inte logga in på sitt lösen. Dessa tre elever blev besvikna.*
- *Vi har enbart 8 datorer vilket försvårar genomförandet, (men fungerar i grupper)*
- *Inloggning och internet krånglar ofta.*

Miljö och provsäkerhet

Några lärarkommentarer handlar om svårigheten att skapa en tillfredsställande provsituation främst med tanke på möbleringen i datasalarna. En lärare följer en liknande tanke, men syftar på den känslighet som möjligen kan finnas med Internetbaserade material.

- *Jag har svårt att se hur man i våra nuvarande datosaler ska kunna genomföra prov i helklass. Det är säkert så att eleverna skärper till sig när det "är på riktigt", Nu hängde de över varandras datorer.*
- *Eleverna satt tätt, vilket underlättar fusk*
- *Kanske blir det lättare för "hackers" att komma åt proven i förväg, jag vet inte.*

Pedagogiska och ämnesrelaterade reflektioner

Hälften av lärarna tar upp pedagogiska och ämnesrelaterade, samt bedömningsrelaterade

farhågor kring de datorbaserade proven. Ett par lärare menar att elever hade svårt att läsa,

eller blev trötta av att läsa på skärmen och flera påpekar att studien endast erbjöd prov i receptiv läsförmåga och att detta bara utgör en del av funktionell språkfärdighet. Ytterligare några lyfter upp svarsformaten, och saknar elevproducerade svar.

- *Flera elever i klassen tyckte att det var svårt att läsa så mycket text på datorn. Det tog längre tid än de hade väntat sig. De såg trötta ut när de lämnade datasalen.*
- *Man kan nog använda datorbaserade prov, men jag anser att de måste kompletteras. Språkförmåga är svår att mäta med enbart flervalsfrågor.*

Provet som sådant

En lärare har flera synpunkter på det genomförda provet som sådant, både avseende funktion och innehåll. Kommentarererna handlar om att lösenord vid inloggningen till provet inte fungerade för någon elev, att det inte gick att ångra givna svar i provet samt om att de vid något tillfälle upplevde att instruktioner var otydliga eller att frågorna innehöll svåra ord i relation till nivån. En lärare påpekar också att de Internetbaserade övningsexempel, som var tänkta som hjälp före provet, inte var tillgängliga vid vissa tillfällen.

- *På en viss fråga gick det inte att ångra ens svar. Man skulle färglägga soffan gul. Eleven råkade färglägga väggarna av misstag men det gick inte att ångra.*
- *Jag såg inte några minus, mer än att själva övningsexemplen inte fungerade när vi skulle öva.*

8.3.3 Sammanfattande kommentar

Att åtta lärare av tolv markerar *Nej, absolut inte* vid påståendet om huruvida de ofta använder datorer i sin undervisning kan uppfattas som en förvånande hög siffra. Förmodligen har den samband med de tekniska hinder lärarna belyser, samt inte minst med att flera skolor inte har datorer så det räcker till en hel undervisningsgrupp, vilket innebär att läraren behöver dela grupper och gå emellan. Några lärare ställer sig positiva till att hädanefter oftare använda datorer. Kanske har de påverkats av elevernas entusiasm, men de förväntningar de här tolv lärarna har när det gäller datorprovets smidighet verkar grusas av en oro över respektive skolors datorkapacitet. Att så många lärare ändå ställer sig tveklöst positiva till datorbaserade prov kan kanske ses som en förhoppning om framtida förbättringar på IT-området. De antagna pedagogiska fördelarna med datorbaserade prov skulle vara intressant att följa upp och få mer specificerade, men kanske kan dessa tolkas som ett utslag av en positiv inställning till studien som sådan och en visst tillmötesgående, så kallad *compliance-effekt* dvs. att lärarna svarar det de tror förväntas.

8.4 Enkäter till skolledarna

Rektorerna vid de sex deltagande skolorna fick inledningsvis ett informerande brev om studien (Bilaga 5) och de ombads också vara med vid planeringssamtalen. Vid tre skolor deltog rektor i dessa och vid de övriga ombads lärarna informera sina skolledare. Efter det att studien avslutats skickades en digital enkät till skolornas åtta rektorer (Bilaga 4). Fem skolledare som arbetar vid fyra av skolorna har besvarat enkäten.

Skolledarenkätens upplägg var identiskt med lärarenkäten, med dels påståenden att ta ställning till och dels utrymme att lämna öppna reflektioner. Enkäten som riktade sig till

rektorerna gav dock endast tre påståenden att ta ställning till. Dessa förekommer även på

lärarenkäten för att möjliggöra eventuella jämförelser. Resultatet redovisas nedan på liknande sätt som för lärarna, men endast som en sammanhållen grupp då antalet är så litet.

8.4.1 Slutna svarsformat

Samtliga fem skolledare markerar de två positiva alternativen på svarsskalan som anger att de anser att deras skola har tillräcklig beredskap att genomföra datorbaserade prov, att de tycker det finns pedagogiska fördelar med datorbaserade prov, samt att de skulle ställa sig positiva till datorbaserade nationella prov.

8.4.2 Öppna kommentarer

Skolledarenkäten erbjöd två skrivfält för öppna kommentarer, ett för positiva och ett för negativa. Liksom när det gäller lärarna ombads de lämna synpunkter utifrån studiens syfte och mål, men också kring annat som de ville lyfta fram, samt eventuellt lämna information om den egna skolan som de bedömde vara relevant.

Positiva kommentarer - förväntningar

Fyra av de fem rektorerna lämnar öppna kommentarer. De menar att datorbaserade prov skulle innebära fördelar för såväl elever som lärarpersonal och ett par påpekar att datorer idag är ett modernt och naturligt verktyg. En rektor lyfter också fram *”fördelar med statistik osv”* och refererar då förmodligen till huvudmannanivån. De fördelar de ser för elever är främst ökad motivation och omedelbar feedback i form av resultat på provet. För lärarnas del betonar rektorerna den arbetsbesparing när det gäller rättningsarbete de datorbaserade proven skulle innebära.

- *Datorn är idag ett naturligt hjälpmedel för ungdomar. De inspireras och tycker det är roligt att jobba med mediet. Kanske är det så att man lyckas bättre när man tycker något är kul att göra.*
- *Eleverna kan vara mer interaktiva med bilder och text. Arbetsbesparande för personalen.*
- *Datorbaserade prov ser jag som bra komplement i de olika verktyg som pedagogerna har i deras arbete att bedöma elevernas resultat gentemot målen. Det sparar tid och arbetsbelastning för rättning, och eleverna får feedback snabbare.*

Negativa kommentarer - farhågor

Samma fyra rektorer har också reflekterat kring eventuella nackdelar med datorbasering av nationella prov. Tre av dem har farhågor när det gäller tekniken och menar till exempel att en satsning på datoriserade prov ställer stora krav på att den egna datorparken är tillräcklig och fungerar tillfredsställande. Ett par kommentarer handlar om provens begränsade omfattning när det gäller att spegla elevens förmåga och utgöra underlag för bedömning. En rektor menar också att det kan förekomma olika attityd till datorer bland lärarpersonalen.

- *Tillgängligheten och allt krångel när det gäller tekniken. Fungerar datorerna idag eller just när jag ska använda dem? Vi har datasalar så att samla en klass med elever som gör provet samtidigt är inga problem men alla elever i en årskurs kan inte göra det vid samma tillfälle. Det måste alltså vara prov som inte går att samtala om.*
- *Lärarens inställning till datorer. Är det ett medie för möjligheter eller är det något skräm-*

mande?

- *Svårt för att få ett heltäckande, behöver kompletteras med ytterligare redovisningsformer*

8.4.3 Sammanfattande kommentar

De här fem skolledarna är, liksom lärarna, generellt positiva till datorbaserade prov och ser flera fördelar, även pedagogiskt. Även om de markerat endast ja-alternativen i det slutna formatet, verkar de dock i de öppna kommentarerna hysa vissa tvivel kring den egna skolans förutsättningar att klara ett reellt genomförande och har också några pedagogiska invändningar. Detta är dock en för liten studie för att några generella slutsatser ska kunna dras.

8.5 Elevernas resultat på proven

Av de till studien engagerade 276 eleverna genomförde, på grund av ett visst bortfall, totalt 219 elever proven; sammanlagt 113 elever i åk 5 och 6 gjorde provet för yngre åldrar och 106 elever i åk 8 det för äldre åldrar. Resultaten registrerades på servern samtidigt som den enskilde eleven loggade ut från provet och översändes till oss som en sammanställning på individnivå i Excel-format när provtiden för samtliga skolor var slut.

I enkäten till lärarna ombads de, innan de fått resultaten, ge sådan information som kunde hjälpa oss tolka resultaten. Då framgår att sammantaget en handfull elever undantagits på grund av att de går i särskilda undervisningsgrupper. För övrigt deltar elever med såväl vissa funktionsnedsättningar som elever med åtgärdsprogram i engelska på grund av att de ännu ej når målen för ämnet, samt elever som har undervisning i nybörjarengelska. En lärare i åk6 var tvungen att undanta en relativt stor grupp elever på grund av att de inte lämnade in medgivanden från sina vårdnadshavare. Två lärare skriver att några elever inte tog provtillfället på allvar. De flesta lärarna belyser också här, liksom i sina enkäter i övrigt, att eleverna var glada, uppspelta men också koncentrerade och lite spända när de gjorde proven. Någon lärare påpekar speciellt att prov på dator gynnade någon elev med funktionsnedsättning som annars kan behöva viss anpassning vid prov.

Eftersom föreliggande studie främst har som fokus att belysa attityder till och förutsättningar att genomföra datorbaserade prov i sig, inte att pröva elevernas kunskaper i engelska, lades inte större vikt vid provresultaten som sådana. Elevernas resultat på proven bedömdes som viktiga och samlades in dels för att eleverna har rätt till ett resultat på sitt prov, dels för att, vilket är det viktiga för studiens del, få en uppfattning om huruvida datorbaseringen kan antas ha påverkat deras prestation.

Underlaget från genomförandet av de båda proven matades in i SPSS och medger analys av resultaten såväl på grupp, individ som uppgiftsnivå, men redovisas här endast per årskurs. Resultaten för populationen i studien visas nedan som lösningsfrekvens och i jämförelse med utfallet av testen när de gick som nationella prov i Norge 2008 (Lie, 2008):

Elevgrupp	Prov, version 3	Range min-max denna studien	Lösning-frekvens denna studien	Lösning-frekvens Norge 2008
29 elever i årskurs 5	åk 5, max 38 poäng	17-37 poäng	0.75	0.67
84 elever i årskurs 6	åk 5, max 38 poäng	22-38 poäng	0.87	(-)
106 elever i årskurs 8	åk 8, max 40 poäng	1-40 poäng	0.68	0.58

Figur 16: Sammanställning av elevernas resultat på proven.

Gruppen elever som deltog fick alltså jämförelsevis hög lösning-frekvens på de datorbaserade proven för såväl yngre som äldre åldrar och därmed goda resultat.

Efter det att lärarna delgivit sina respektive elevers resultat ombads de kommentera dessa utifrån sin professionella bedömning av elevernas förmåga när det gäller engelsk läsförståelse — dvs. *test construct* (se 6.4.3), samt eventuellt jämföra med motsvarande delar i det nationella provet i engelska. Detta bedömdes som viktigt då avvikande resultat på proven eventuellt skulle kunna sättas i samband med datorbaseringen som sådan — dvs. *rubric* (jämför min första forskningsfråga 2.2). Åtta lärare gjorde detta, och de menar generellt att elevernas prestationer stämmer väl överens med deras bedömning. Ett par lärare för yngre åldrar menar att överensstämmelsen med resultaten på läsdelen i Äp5 är god eller mycket god. Flera av lärarna menar att enstaka elevers resultat överraskade dem något, avseende såväl högre som lägre poäng. En lärare i årskurs 8 frågar sig om provet lämnade tillräckligt utrymme för de duktigaste eleverna att visa sin förmåga, vilket förmodligen kan relateras till de aktuella provens relativt begränsade utformning (se 7.2).

Kommentar

Provresultaten för eleverna som deltog är mycket goda. Emellertid bör beaktas att en direkt jämförelse med de norska resultaten egentligen inte är adekvat. Å ena sidan är alla de svenska eleverna i studien äldre än de norska som går i motsvarande årskurs - men å andra sidan börjar många/de flesta svenska elever med engelska i skolan något år senare. Dessutom, eftersom föreliggande studie främst har som fokus att belysa attityder till och förutsättningar att genomföra datorbaserade prov i sig, lades inte större vikt vid att skolorna skulle genomföra under provlika former. För studiens del kan elevernas prestationer dock sägas bekräfta att svenska elever i regel har goda kunskaper i engelsk läsförståelse enligt dessa prov och att datormediet i sig — dvs. *rubric* — i detta fall inte tycks ha hindrat dem.

9 Sammanfattande diskussion

Nedan förs en sammanfattande diskussion rörande först uppsatsens design, metod och material och därefter kring ämnets större kontextuella sammanhang och problemområde. Sedan följer två avsnitt med utgångspunkt i de två forskningsfrågor som jag med denna uppsats avser försöka besvara. Vidare dras, mot bakgrund därav, några slutsatser om implikationer avseende vidare utveckling av de svenska nationella proven i engelska. Resonemangen under de olika rubrikerna nedan kan i någon mån gå in i varandra. Med generaliserande begrepp som "litteraturen", "forskningen" och "studien" avses nedan den litteratur, de internationella erfarenheter och den empiriska studie som redovisats ovan.

9.1 Design, metod och material

En övergripande utgångspunkt för denna undersökning är att datorer redan har en, om än omdiskuterad, stor roll i internationell språkbedömning och kommer, i större eller mindre omfattning, att göra sitt intåg även i Sverige. Uppsatsens design speglar en arbetsgång i inledningsskedet av ett sådant förändringsarbete, och avser belysa: 1) Var vi befinner oss, 2) Vart vi (sannolikt) är på väg och varför, samt vad vi behöver tänka på, samt 3) Hur det skulle kunna fungera och tas emot i praktiken i vår kontext.

Undersökningen består alltså av tre innehållsliga delar som inbördes beror av varandra:

Avsnitt 3 – 5 utgör en bakgrund och redovisar den teoretiska grunden för var svensk språksyn, styrdokument och bedömning av språklig kompetens står 'idag'

Avsnitt 6 belyser aktuell forskning rörande datorbaserad av språkbedömning och aktuella erfarenheter av dess effekter, vilket förmodligen är 'nästa steg'

Avsnitt 7 redovisar resultatet av en empirisk studie i svenska grundskolor med avsikt att undersöka datorformatets praktikalitet samt användarnas förväntningar och reaktioner, i ljuset av deras erfarenheter av dagens språkbedömning – för att se hur 'nästa steg' kan tas emot och fungera i några svenska grundskolor

Dessa tre delar måste på olika plan beaktas och samvariera för att en datorbaserad avspråkbedömning skall kunna inledas. De utgör också förutsättningar för att kunna belysa och uppfylla de utvärderingskrav på bedömningar som ställs av Bachman & Palmer (1996, 2010) samt enligt Messicks (1989) syn på ett enhetligt validitetsbegrepp (se t.ex. 4.4, 6.4 och 6.6.2).

Litteraturstudien rörande datorbaserad avspråkbedömning i språk utgår från ledande forskares arbete inom området, som t.ex. Dan Douglas och Carol Chapelle, samt forskningsorganisationer som till exempel ETS (Educational Testing Service) och JRC (Joint Research Centre). För att kunna ge en bred bild men också sammanställa materialet och analysera resultatet, med syfte att svara på min första forskningsfråga, användes Bachman & Palmers (1996) vedertagna kriterier för utvärdering av en bedömnings lämplighet och användbarhet, nämligen: reliabilitet, begreppsvaliditet (*Construct validity*), autenticitet, interaktivitet samt påverkan (*Impact*) och praktikalitet. Min bedömning är att dessa faktorer fungerade väl som ett verktyg för att åskådliggöra de problemställningar som diskuteras i ämnet. Till synes kan eventuellt redovisningen spänna över ett bredare fält än forskningsfrågan indikerar men utifrån en enhetlig validitetssyn (se 4.4) och en vidgad syn på språkbedömning (se 4.2), vilket jag återkommer till nedan, ser jag inte att en snävare

redovisning hade kunnat motiveras.

Den empiriska studien genomfördes och analyserades i mycket enligt den arbetsgång för utprövningar som är vedertagen inom NAFS-projektet (se avsnitt 5). Materialet samlades in enligt triangulering med flera beröringspunkter i form av enkäter till flera olika grupper respondenter, vid flera tillfällen och vid deltagande observationer, liksom vid samtal. Egentliga intervjuer, t.ex. med några lärare eller elever, hade naturligtvis kunnat bidra med djupare insikt på individplanet men valdes bort dels på grund av omfattningen på materialet, dels på grund av att ämnets natur i den här aktuella uppsatsen snarare fokuserar på grupp- än individnivå. I ett senare skede av ett implementeringsarbete med datorbaserade nationella prov skulle intervjuer däremot kunna tillföra mycket, som jag ser det, till exempel angående elever med olika typer av funktionsnedsättningar och behov av anpassningar av provmaterial, för att alla elever ska kunna beredas goda och likvärdiga möjligheter att visa vad de kan på nationella prov.

Resultatet lagrades kvantitativt samt analyserades och kategoriserades kvalitativt. Avseende undersökningens pålitlighet och generaliserbarhet togs flera omständigheter i beaktande: dels respondenternas beroende av sin skolkontext, dels deras eventuella positiva inställning till studien som sådan och vilja att tillmötesgå och dels min egen roll (se 7.1 och 7.2). Antalet respondenter och skolor är emellertid sammantaget relativt stort och elevernas goda resultat på proven förstärker att genomförandet fungerade väl när det gäller deras förmåga att använda datorer i provsammanhang, och därmed att datorbaseringen i sig inte hindrade dem. Min bedömning är att respondenternas attityder och förväntningar som de kommer fram i underlaget och hur proven faktiskt fungerade ute på dessa skolor, vilket är min andra forskningsfråga, är tillräckligt representativa, med tanke på det breda underlaget, för att vara värdefulla. En faktor som kan påverka generaliserbarheten i synnerhet i detta sammanhang är dock tiden. Den här aktuella studien gjordes vårterminen 2009, samma år som PISA CBT, och sedan dess tycks fler och fler skolor ha utrustat sina elever med egna datorer eller läsplattor. Detta är onekligen ett steg på väg, men storskaliga prov som ska göras av ca 100 000 elever, helst samtidigt, kräver dock mer likvärdig teknik och mer eftertanke än så (se 6.5.2 t.ex. Bachman & Palmer, 2010; Douglas, 2010; Meijer, 2009). Rent konkret krävs också god server- och bredbandskapacitet, samt nätverksanslutning och inte minst support, men den tekniska beredskapen i form av fler datorer, har sannolikt ökat i många svenska skolor.

När det gäller provmaterialet var detta ett verktyg för att undersöka datorbaserad av engelskprov, inte elevernas engelskkunskaper. De norska proven prövar förståelse av fraser och korta texter, vilket ryms inom svenska målbeskrivningar, men också viss datoranvändning (se 7.2.3). De svenska eleverna var därtill ett år äldre än provens norska målgrupp (se 7.2.2). Snarare var det en fördel att proven upplevdes som lätta ur ämnessynpunkt av de flesta elever. Svårighetsgraden skulle inte utgöra ett hinder då de genomfördes, vilket elevernas goda provresultat visar. Det var också en fördel att prov och instruktioner, liksom demoversioner med exempeluppgifter, i Norge är helt enspråkiga på engelska, liksom i Sverige, vilket inte är fallet i alla länder. Dessa saker diskuterades med lärare och elever vid planeringssamtalen och vid de deltagande observationerna, samt framgick av den skriftliga informationen till lärarna. Mitt samlade intryck är att dessa prov och de instruktioner som gavs fungerade mycket väl för ändamålet.

9.2 En förändrad syn på språklig kompetens – en didaktisk utmaning

Arbetet med denna uppsats har fått mig att inse att, inte enbart bedömning av språklig kompetens, men framför allt hela synen på språklig kompetens och själva språkbegreppet står inför ett paradigmskifte. Att datorer används alltmer inom bedömning av språk, Europarådet anger digital kompetens som en nyckelkompetens för framtiden (6.1) och att PISA 2009 prövar digital läsning och traditionell läsning är tecken på detta som bland andra (se 6.2) diskuteras i denna uppsats. Flera forskare menar att datoranvändning bör ses som en del av själva det kommunikativa språkbegreppet i det att man talar om till exempel digital läs- och skrivkunnsighet (*ICT-literacy*) som en ny språklig kompetens, i sig bestående av flera färdigheter och delkompetenser (se 6.4.3).

Min undersökning handlar om bedömning, närmare bestämt datorbaserad bedömning i språk. En övergripande insikt och slutsats är emellertid att ovanstående och allt vad det för med sig förmodligen är toppen av ett isberg. Bedömningsforskningen betonar alltmer bedömningens roll som en oskiljaktig del av lärande och undervisning (se 4.2) och menar att dessa tre förutsätter varandra. Det som bör gälla för god bedömning bör också gälla god undervisning för att stödja och utveckla lärande. Att ta kunna ansvar för att, oavsett om det gäller process eller produkt, bedöma "rätt" saker på "rätt sätt" så att det får goda konsekvenser (se t.ex. 4.4) bör också gälla för god undervisning och de utvärderingsfaktorer, som här använts för att belysa forskningens syn på datorbaserad bedömning i språk, är också relevanta för didaktiska sammanhang med mer specifikt fokus på undervisning och lärande. Flera forskare, (se t.ex. 6.1) konstaterar att den utmaning språkforskningen, och inte bara specifikt bedömningsforskningen som här fokuserats, står inför inte ligger på det teknologiska planet utan på det konceptuella, men också att ett av de viktigaste övergripande budskapen handlar om att tekniken måste inordnas i språkforskningens tjänst och inte tillåtas löpa amok, styrs från huvudmannanivå eller av andra intressenter.

Flera forskare ser stor pedagogisk och didaktisk potential i datorbaserade undervisningsmaterial (se 6.4). Men de hävdar också att den exploderande användningen av såväl datorbaserade bedömnings- som undervisningsmaterial kan få negativa påverkanseffekter på undervisning om man inte ger språkanvändning via dator större specifikt undervisningsdidaktiskt avtryck och forskningsutrymme. I den svenska rapporten om PISA 2009 (se 6.6.5), liksom i flera jämförande studier (6.6.3) konstateras att "datoranvändning" och "datorfamiliaritet" är trubbiga och förmodligen inte längre relevanta begrepp. Nästa steg bör därför handla om att identifiera, fånga och ytterligare beskriva dessa nya språkliga färdigheter, strategier och kompetenser, (jfr. 6.4.3, s. 33) för att i praktiken kunna förmedla dem via utbildning. Detta förefaller inte minst viktigt då man i nämnda studier (se 6.6.2-4) pekar på ökande risker för segregering avseende främst vissa socio-ekonomiska grupper som annars riskerar att ställas utanför informationssamhället. Chapelle & Douglas (2006) och Chapelle (2010) betonar, som tidigare citerats i 6.1 och 6.4.3, nödvändigheten av deras eget område, nämligen språkforskning som enligt en innovativ agenda tar sig an komplexiteten i detta och överskrider gränserna mellan bedömning, språk och teknologi för att hitta innovativa och fruktbara vägar att beforska, förstå och eventuellt omvärdera eller "uppdatera" begreppet kommunikativ språklig kompetens, med syfte att förmedla kunskapen till andra. Jag tror att detta är en stor och svår utmaning för den didaktiska språkforskningen. Möjligen kan denna uppsats, även om dess egentliga fokus är

ett annat, ses som ett bidrag genom att kanske inspirera någon till vidare forskning.

9.3 Förhållandet mellan construct/ 'Vad' och rubric/ 'Hur', samt effekter därav, vid bedömning av språklig kompetens via dator

Som behandlats i avsnitt 3, Begreppslig bakgrund, definierar aktuell bedömningsforskning i språk kommunikativ språklig kompetens utifrån en handlingsorienterad grundsyn (*action oriented approach*) och ser språkanvändaren som en social aktör. Begreppet kan således inte ses som statiskt utan beroende av sin samtid. Detta innebär att bedömning av densamma för att kunna anses rättfärdig enligt de principer som anges av Bachman & Palmer (2010), kontinuerligt bör utvecklas och moderniseras. En vidgad syn på validitet utifrån Messicks definition (1996), betonas och diskuteras genomgående i bedömningslitteraturen. Denna förutsätter en analys av bedömningens begreppsvaliditet, dvs. förhållandet mellan vad som faktiskt bedöms, *the construct*, vid den aktuella bedömningen, liksom i vid mening vad som avses bedömas enligt dess syfte, avsedda användning/konsekvens, aktuell språksyn och styrdokument. Mitt samlade intryck av litteraturgenomgången är att bedömningsforskningen befinner sig vid något av ett vägskalet när det gäller sitt förhållande till samtidens tekniska/teknologiska (r)evolution, i relation till god bedömning och språkbegreppet, vilket också anförs av flera forskare.

Min slutsats är, utifrån min analys av aktuell forskningslitteratur, att användningen av dator i bedömnings-sammanhang, med fokus på metodeffekter i förhållandet mellan *rubric* och *construct*, kan identifieras och kategoriseras i tre steg enligt de aspekter på begreppsvaliditet som redovisats:

1. Datorn används som 'maskin' och enhetligt format. *Rubric* kan anses ha begränsad effekt vid en relevant definition av den aktuella bedömningens *construct*.
2. Datorn används som 'verktyg' och integrerad metod. *Rubric* bör betraktas som en delkomponent i form av datorkompetens (*computer literacy*) i bedömningens *construct*.
3. Datorn används för bedömning av informations- och kommunikationsteknologisk kompetens (*ICT-literacy*), utifrån ett *rubric* oskiljaktligt från ett moderniserat *construct* för kommunikativ språklig kompetens.

Det första steget att datorbasera bedömning i språk är således att ersätta pappret med skärmen och pennan med mus och tangentbord, samt göra ev. in- och uppspelning av ljud via datorn. Datoranvändningen blir en del av bedömningens format, och därmed av *rubric*, men metodeffekter som kan påverka bedömningens validitet förefaller relatera mer till uppgiftsformat än datoranvändningen i sig. Datoranvändning i denna bemärkelse förefaller i dagsläget vara bekant för flertalet testtagare. Demoversioner har vidare visat sig kunna skapa god familiaritet med tillvägagångssättet.

Det andra steget att datorbasera innebär integrerad användning av sådan mjukvara som anses betjäna eller berika den aktuella bedömningen. Innovativa uppgiftstyper, virtuella och datorspelsliknande format, textbehandling och en mängd hjälpfunktioner skulle utgöra väsentliga delar av *rubric*. Färdigheten att använda dessa prövas därmed som en integrerad delkomponent av provets *construct*, om än inte i alla avseenden relevant för det egentliga språkbegreppet. Insiktsfulla överväganden kring *test construct* respektive *pure construct/construct of interest* förefaller väsentliga, vilket också betonas av Chapelle &

Douglas (2006).

Dessa två första varianter har stora likheter i det att de möjliggör relativt stor kontroll över, och ansvarstagande för, bedömningens innehåll för provutvecklare och huvudmän. Chapelle & Douglas (2006) understryker att CBT inte riskerar sämre validitet än annan bedömning, under förutsättning att den definieras och valideras som just datorbaserad bedömning av språklig kompetens via dator. De framhåller också att datorbundet *rubric* ger större konsistens och därmed högre reliabilitet och likvärdighet, medan andra forskare, som t.ex. Bridgeman (2009), delvis emotsäger detta. Jag tycker mig, i min empiri avseende läsning se att ett mera 'låst' rubric kan hindra elevens kontroll över processen och egna strategier att lösa uppgifter, vilket skulle kunna exemplifiera *construct-irrelevant variance*. Eventuellt innebär en ytterligare styrd process att bedömningen, förutom *proficiency*, också i högre grad än tidigare berör *performance* och såväl process som produkt. Jag har inte funnit detta problematiserat i litteraturen, förutom som fördelar vid CBT och simulerade miljöer vid just *performance testing* (se 6.4.2), vilket diskuteras av t.ex. Douglas (2000, 2010).

All datorbaserad inbegriper nya intressenter som utvecklar program och mjukvara och flera forskare varnar för att låta tekniken berusa istället för att berika. Med tekniken i bedömningens tjänst tycks högre grad av interaktivitet och autenticitet dock kunna betjäna god begreppsvaliditet. I provsammanhang som innefattar utökad digitalisering och användning av mjukvara, framstår överväganden avseende *bias* och *construct-irrelevant variance* eller *-difficulty* än viktigare och flera forskare anser att innovation ännu lämpar sig bäst i formativa material för undervisningssammanhang, dvs. CALL, *Computer Assisted Language Learning*. Flera forskare bekräftar elevens möjligheter till mer autentiska lärandesituationer i CALL (t.ex. Chapelle, 2010), och god formativ bedömning som i DIALANG (Huhta m.fl., 2002). Skolans och undervisningens pedagogiska betydelse för elevers datorkompetens, så kallad *computer literacy* till skillnad från 'datoranvändning', understryks inte minst ur ett likvärdighetsperspektiv av rapporteringen från PISA CBT (Skolverket, 2011b). Emellertid är t.ex. ordbehandling förmodligen så vedertaget att papper och penna som verktyg för skriftlig produktion kan orsaka *construct-irrelevant difficulty* för många testtagare. Formatet i sig tycks också på ett positivt sätt vara motiverande för många elever, vilket jämförande studier i litteraturgenomgången och även min egen empiri bekräftar.

I relation till det kommunikativa språkbegreppet verkar bedömningsforskningen överens om att dessa typer av CBT rätt använda kan utgöra goda komplement till, men inte ersätta, andra bedömningsformer. Läsning på skärm gör till exempel att löpande längre text måste undvikas, vilket medför *construct under-representation* och maskinell rättning utesluter (ännu) öppna svarsformat och all produktion, vilka också måste anses väsentliga komponenter i relevant bedömning av kommunikativ språklig kompetens. Autenticitet och interaktivitet kan simuleras till exempel i form av styrda stimuli vid t.ex. muntliga prov, så kallad *semi-direct* prövning (Hughes, 2003), men utgör knappast alternativ till direkt interaktion. Avseende maskinell rättning och flervalsformat framförs vinster i effektivitet men också i objektivitet och konsistens, medan andra ser dessa kvaliteter främst som egenskaper hos bedömningen i sig, vilket jag ser som en viktig invändning. Kritiker menar också, framför allt när det gäller CAT (*Computer Adaptive Testing*), att dikotom bedömning av isolerade moment eller uppgifter på bekostnad av integrerande format tangerar tillbakagång till en syntetisk språkfärdighetssyn och frestelse att göra det lätt mätbara till det viktiga, vilket framstår som oförenligt med god bedömningskultur. Denna typ av prov verkar emellertid attrahera på makronivå för dess effektivt rapporterade funktion. Liknande politiska aspekter på och washback-effekter på t.ex. utbildningssystem av framför allt

storskalig nationell prövning diskuteras bl.a. av Shohamy (2001) och Bailey (1996, 1999) och är, som jag ser det, väsentliga att beakta framför allt vid överväganden om att övergå från en tradition till en delvis annan (se 6.5). Ett nära exempel på detta är mottagandet och diskussionerna i Norge vid implementeringen av det datorbaserade prov som användes i den här aktuella empiriska studien (se 6.5.2).

Det tredje steget när det gäller CBT i språk är, som jag ser det och som också litteraturgenomgången visar, något helt annat. Denna bedömning utnyttjar och förutsätter, förutom den datoranvändning/kompetens som beskrivs ovan, användning och behärskning av det datorn kan ge tillgång till, genom Internet och sociala medier. Flera forskare menar att det kommunikativa språkbegreppet på 2000-talet bör innefatta även språklig informations- och kommunikationsteknologisk (IKT) kompetens, *ICT-literacy*, vilket naturligtvis i grunden förutsätter god datorkompetens. Inte minst efterfrågas detta av näringslivet och digital kompetens är en av åtta rekommenderade nyckelkompetenser (Europarådet, 2006). Dessa har också influerat Skolverkets skrivningar i Gy 2011 om så kallade högskoleförberedande förmågor som studenter behöver ha med sig från gymnasieskolan, då man konstaterar att "Synen på generella kompetenser har ändrats" (Skolverket, 2011c/Gy 2011, s. 13). "Digital kompetens" återfinns i Gy 2011 bland sexton s.k. nyckelord, liksom att "Läsa längre texter på engelska" och "Hantera stora textmängder" (s. 27). Se vidare om den svenska kontexten nedan.

Litteraturen manar till viss försiktighet inför storskalig bedömning av språklig IKT-kompetens, *ICT-literacy*, och betonar etiska riktlinjer (se 4.5 och 6.5). Förespråkare eftersöker, enligt mitt intryck, främst *washback*-effekter i utbildningssammanhang (t.ex. Ripley, 2009). PISA CBT 2009 erbjuder, parallellt med traditionell läsning, prov i digital läsning, navigering, textkonstruktion, källanvändning mm., vilket innebär prövning av nya förmågor inte som medel, utan mål. Att öppna porten till Internet skulle innebära stora utmaningar i storskaliga provsammanhang, inte minst ur säkerhets- och likvärdighetssynpunkt. Erfarenheter i litteraturen visar att dessa kompetenser (först) behöver få större utrymme i undervisningssammanhang och att risken för *bias* är relativt stor gentemot de grupper av elever som är mest beroende av skolan för sitt lärande.

Sammanfattningsvis drar jag slutsatsen att CBT i språk innebär en sådan förändring av *rubric* att detta, i egenskap av en betydande metodeffekt på olika nivåer, påverkar bedömningens validitet, såvida det inte också medför relevant omvärdering av *construct* för den aktuella versionen av CBT. Den bedömning som görs av språklig kompetens via dator bör således betraktas som en bedömning även av förmågan att använda dator på det sätt som provet förutsätter. En väsentlig fråga, enligt litteraturgenomgången, blir då i vilken mån den datoranvändning som förutsätts är relevant för språkbegreppet, enligt ovan. Beroende på hur *construct* definieras för så kallade *proficiency tests* kan således såväl dator- som pappersprov orsaka hot mot bedömningens validitet. En vidgad syn på validitet med ansvarstagande på alla nivåer förefaller väsentlig i dessa nya sammanhang, inte minst med tanke på *washback* och signaleffekter på lärande, undervisning och samhälle, avseende storskalig bedömning som markör för relevant kunskap, vilket också benämns *feedback*-effekter — jfr. 6.5.2: t.ex. Meijer (2009).

9.4 Datorbaserade prov i engelska på några svenska skolor — förväntningar, reaktioner, resultat

Vid sex tämligen olika skolor genomförde och övervägde 12 lärare och 219 elever i åk 5, 6 och 8 datorbaserade provuppgifter som ett alternativ eller komplement till traditionella pappersprov. Studien har genererat ett brett underlag och resultaten stämmer, enligt mitt intryck, mycket väl överens med den bild av överväganden avseende CBT som framkommer i litteraturgenomgången. Resultaten visar att involverade respondenter generellt är mycket positiva till prov i datorbaserad form och att deras invändningar utgör konstruktiv kritik att tillvarata under kommande arbete med datorbaserade material för den svenska skolan. Kärnan i både lärarnas och elevernas reaktioner verkar handla om förutsättningar, på ett konkret plan tekniska och materiella sådana för skolornas del, men också om att datormediet tillför en ytterligare och ny aspekt i bedömningsprocessen och elevers olika beredskap och möjligheter att visa vad de kan i engelska.

Sammantaget indikerar studien dock att den tekniska beredskapen som helhet, till exempel vad gäller datorpark, bredbandskapacitet och nätverksåtkomst, på flera skolor är långt ifrån tillräcklig för att genomföra storskaliga nationella prov med alla elever i en årskurs samtidigt, vilket stämmer överens med erfarenheter från Norge. Lärare och elever är också överens om att datorer relativt sällan används i engelskundervisningen, även om flera lärare ser pedagogiska möjligheter i språk. Flera av de synpunkter som framkommer i enkäterna som nackdelar respektive fördelar med själva datoranvändningen bekräftas av tex. Bridgeman (2009) och Chapelle & Douglas (2006).

Forskning avseende *washback*, dvs. påverkans effekter, har visat att prov uppfattas, och ska kunna uppfattas, som en indikator på viktiga kompetenser i ämnet. Det är därför intressant att iaktta att flera synpunkter från respondenterna relaterar till provets begreppsvaliditet, vilket också iakttagits i större bedömningsstudier (Erickson & Gustafsson, 2005). Att (bara) 63 % av eleverna svarar att provet innebar att de fick visa hur bra de är på att läsa och förstå engelska (se 8.2.2), kan indikera att de saknar moment som de anser viktiga och några gav också uttryck för att provet på olika sätt var trivialt. Några lärare och skolledare problematiserar provets domäntäckning men däremot ifrågasätter ingen respondent explicit relevansen i att använda dator vid prov i språk.

Relativt många resultat kan tolkas som kopplade till olika typer av motivationsaspekter som uttryck för ansträngning, kapacitetsupplevelse och oro (se 5.2). Elevernas acceptans och självförtroende när det gäller datorer, och engelska, förefaller generellt gott, vilket är gynnsamt för provsituationen (se 5.2). 10 % av eleverna tror före provet också att det kommer att gå bättre än vanligt för dem, även om denna siffra sjunker ungefär lika mycket i retrospektion (s. 8.2.1.1 och 8.2.3). Med tanke på att detta prov i studien presenteras som ett alternativ till de prov eleverna är vana vid och i ljuset av god bedömningspraxis, är elevernas svar på i synnerhet några frågor oroande. Efter genomfört prov (se Elevenkät II, 8.2.2.1) är det 10 % av eleverna som ger uttryck för att de behöver hjälp för att göra provet och drygt 20 % som menar att de inte kände sig riktigt tillfreds med själva provsituationen, samt 22 % som hellre gör traditionella pappersprov. Detta är sammantaget relativt många elever och manar till eftertanke och behöver undersökas ytterligare avseende eventuell datorbasering av storskaliga provmaterial.

Kommentarer som relaterar till likvärdighet och familiaritet (se 6.6) är vanliga. 10 % av eleverna anger att de relativt sällan använder dator hemma, vilket stämmer väl med resultat från PISA-undersökningen (Skolverket, 2011b) och när det gäller skolan är motsvarande siffra så hög som 65%. Det sammantagna intrycket från hela den empiriska studien blir dock att de elever (och lärare) som gjort exempeldemon hade god beredskap i form av familiaritet. Det är intressant att notera att 73 % av eleverna markerar att det är rättvist att göra prov på dator i skolan, medan endast 49 % anser att det är lika lätt för alla. I de öppna kommentarerna (se 8.2.2.1 och 8.2.2.2) återkommer eleverna till rättvisefrågor främst avseende datorvana och skriftlig färdighet, men det skulle vara värdefullt att veta ännu mer om hur de resonerar, och en intressant fråga att följa upp ytterligare.

Några jämförelser gjordes mellan olika grupper respondenter och det är tydligt att pojkar och elever i yngre åldrar är mest positiva till detta prov. De tycker dels att provet gör deras kunskaper rättvisa och dels att de vill ha fler datorbaserade prov i framtiden. Avseende genus stämmer detta väl med de jämförande studier som rapporterats i 6.6.4 och möjligen kan åldersaspekten avspegla en mer avdramatiserad inställning till prov hos yngre elever. En annan intressant aspekt är att en jämförelse mellan överensstämmande frågor i de båda enkäterna (se 8.2.3) visar att antalet elever som hellre vill göra traditionella pappersprov i framtiden ökat efter provets genomförande och att acceptans för CBT minskat på motsvarande sätt. Eleverna tenderar också att bli mer säkra i sina uppfattningar om dator- respektive pappersprov i retrospektion, vilket troligen kan sättas i samband med deras erfarenheter av det aktuella provet och provsituationen. Det faktum att 24 % av eleverna inte tycker att vi i Sverige ska satsa på CBT i engelska måste betraktas som en relativt hög siffra.

Med tanke på den snabba utvecklingen inom teknikens och skolans område, och i synnerhet gällande 10- till 15-åringars datorvanor, blir en avslutande kommentar att den empiriska studien visserligen skulle möjliggöra åtskilliga fler analyser men att det vore mer intressant att göra en liknande studie i jämförande syfte, på de aktuella skolorna, med (ev. samma) lärare och nya elever – i samband med utprövningar av ev. kommande svenska nationella datorbaserade prov i språk. Till sist kan konstateras att det mest övergripande intrycket ändå är att kvaliteten i vidare utveckling av CBT för den svenska skolan har allt att vinna på samverkan mellan berörda parter på den politiska, akademiska, tekniska och lokala nivån. Detta gäller inte minst de bidrag som lämnas av dem det allra mest berör, nämligen eleverna.

9.5 Implikationer för det svenska nationella provsystemet i engelska

I november 2011 uppstod tekniska problem i Transportstyrelsens system för kunskapsprov. Proverksamheten fick avbrytas, flera tusen personer fick avstå teoriprov och Sveriges körskolor stannade, eftersom kunskapsprov föregår förarprov (Transportstyrelsen, 2011). Samtidigt utsågs Sveriges IT-kommun 2011. Västnytt rapporterade att Ale nominerats som en av tre kandiderande kommuner, bl.a. grundat på satsningar i skolan: "Ale kommun är först i Sverige med att införa "en till en-datorer" i de yngsta åldrarna (6 – 9 år). Kommunen har valt den gruppen eftersom man har sett att det är där som datoriseringen har störst effekt, särskilt på läs- och skrivinläringen" (Svtplay, 2011).

Och det är väl ungefär där vi står. Inledningsvis kan jag således konstatera att vidare utveckling av det svenska nationella provsystemet i språk avseende eventuell datorbaserad står inför avsevärda såväl konceptuella som tekniska utmaningar, vilket belysts av både litteraturgenomgången och den empiriska studien.

De tekniska förbehåll rörande bland annat säkerhet och praktikalitet som framkommit tyder på att datorbaserad av nationella prov, för att fungera väl, kräver mer resurser än de som nu finns tillgängliga på provinstitutioner och flera skolor i Sverige. Det skulle också, i svensk bedömningskontext med nationella prov som stöd för lärarens bedömning, vara oroande om datorbaserad av nationella prov genomförs av anledningar som handlar om effektiviseringar och övertro på central rättning, det senare något som ibland diskuteras i relation till Skolinspektionens "omrättning" av nationella prov. En slutsats jag vågar dra från mina resultat är att detta inte skulle vara av godo vare sig för elever eller för bedömningens validitet.

Med tanke på den tradition av god bedömningspraxis som utmärker de nationella provmaterialen i språk förefaller det mest väsentliga dock vara avväganden avseende CBTs validitet och rättfärdigande i vår kontext. De svenska nationella proven i engelska för grund- och gymnasieskolan är så kallade mål- och kriterierelaterade *proficiency tests* av *high stakes*-karaktär, med syfte att "stödja en likvärdig och rättvis bedömning och betygssättning" samt "ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå". Vidare anges: "De nationella proven kan också bidra till: att konkretisera kursplanerna och ämnesplanerna" samt "en ökad måluppfyllelse för eleverna"(Skolverket, *Nationella prov*). Viktiga aspekter är alltså vad som avses prövas dvs. avsett *construct*, likvärdighet och meningsfull *washback*, dvs. såväl summativa som formativa, didaktiska och demokratiska aspekter. Och så den här avgörande frågan om huruvida datorkompetens är relevant i sammanhanget.

Först kan dock konstateras att CBT ännu förefaller ha sådana begränsningar visavi god helhetsbedömning av kommunikativ språklig kompetens att en datorbaserad av nationella prov i språk som helhet inte är att överväga, på ganska länge. Min slutsats är, å andra sidan, att datorbaseringens möjliga positiva effekter tycks väl motsvara forskningens krav på kommunikativa språkprov. I GERS talas om generell kompetens som används vid språkliga aktiviteter samt kommunikativ språklig kompetens som "den kompetens som gör det möjligt för en person att agera särskilt med hjälp av språkliga medel" (GERS, s. 9). Liknande övergripande mål inkluderas med fokus på datorkompetens i de norska engelskproven som här använts.

Utifrån en vidgad syn på bedömning som en självklar del av lärande och undervisning går dessa tre hand i hand (t.ex. Erickson, 2009a; Taras, 2005) och stöd för datoranvändning i de svenska styrdokumenterna för engelska finns. Som beskrivits ovan knyter de nya kurs- och ämnesplanerna för språk än mer till Europarådets språkpolicy som den uttrycks i GERS. Avseende explicit datorbaserad finns, förutom det som nämnts ovan, emellertid mycket lite att hämta i GERS eftersom den tillkom i början av 1990-talet. En dokumentökning ger, för att illustrera detta, 15 träffar totalt på 'dator', 0 på 'digital' och 'teknologi' samt 2 vardera för 'informationsteknik' och 'ny teknik'. I Kursplan för engelska GR2011 anges däremot i syftestexten användning av "olika hjälpmedel för lärande, förståelse, skapande och kommunikation" och i Kommunikationens innehåll för reception "talad engelska och texter

från olika medier” för åk 1-9, respektive ”söka, värdera, texter och talat språk från internet

och andra medier” för åk 4–9 (Skolverket, 2011d). I Gy2011 berörs, liksom i GERS, generella kompetenser och därvid digital kompetens som en framför allt högskoleförberedande förmåga (Skolverket, 2011c, s. 9 och 27), en begränsning som framstår som (mycket) förvånande, eftersom digital kompetens med all sannolikhet kommer att vara nödvändig för alla i framtiden (se 6.2). I Ämnesplanens syftestext för engelska framhålls emellertid att ”interagera i tal och skrift samt producera talat språk och olika texter, på egen hand och tillsammans med andra, och med stöd av olika hjälpmedel och medier” samt ”hur man söker, värderar, väljer och tillägnar sig innehåll från olika källor för information” (Skolverket, 2011e). I ljuset av detta skulle således en försiktig datorbaserad av vissa delkomponenter i de nationella proven i engelska vara meningsfull och förmodligen gynna provens validitet.

Aspekter avseende likvärdighet och demokrati relaterar till bedömningens användning och konsekvenser, samt de effekter den får från individ- till systemnivå. Etik och ansvarstagande har, som tidigare redovisats, fått en alltmer framträdande betoning i bedömningsforskningen, inte minst i relation till datorisering. Framför allt ser jag det, också ur ett validitetsperspektiv, som oroande att internationella jämförande studier som här redovisats, visar på förhållandevis stora brister i likvärdighet gentemot de grupper av elever som förmodligen är mest beroende av skolans undervisning när det gäller datorkompetens och IKT. Liknande resultat för Sverige rapporteras dessutom från bla. PISA CBT 2009. Å andra sidan förefaller vissa skillnader av olika skäl utjämnas något vid datorisering och många elever, och ev. vissa (nya) grupper, gynnas och motiveras av CBT, vilket naturligtvis är viktigt att beakta och också framkom i den empiriska studien i det att flera elever med funktionsnedsättningar tycktes arbeta bättre vid datorn än med papper och penna (t.ex. 8.1 och 8.2.2.2). Förhoppningsvis kan en *washback*-effekt av CBT i vissa bedömningssammanhang innebära satsningar i undervisningen avseende språkliga färdigheter via dator.

Sammanfattningsvis drar jag slutsatsen att det vid datorisering av nationella prov i språk är klokt att ”skynda långsamt” – men att det vore motiverat att förbereda utveckling av något eller några datorbaserade delprov. Det skulle då företrädesvis gälla skriftlig produktion med ordbehandling och möjligen också ett, eventuellt maskinellt rättat, komplementärt delprov i reception med slutna svarsformat. Inget av dessa skulle dock avse pröva *ICT-literacy* eller datoranvändning som en del av kommunikativ språklig kompetens i en vidgad bemärkelse. Detta måste, som jag ser det, först vara (ytterligare) etablerat i såväl undervisningssammanhang som i styrdokumentet för språk, innan det kan bli aktuellt för nationell bedömning. Fortsatta studier, med olika metoder och ansatser, av språkundervisning och aktuell bedömningsforskning, samt erfarenhetsutbyten internationellt och lokalt, liksom grundligt utvecklings- och utprövningsarbete förefaller därför i alla avseenden väsentliga.

För övrigt vore det en svindlande tanke att genomföra detta uppsatsarbete utan ordbehandling, statistikprogram, digitala enkäter, mail och Internet. Och kanske kan viss datorisering av nationella prov bidra till att skolan kan ge fler elever språklig datorkompetens och IKT-kompetens.

Referenser

- Alderson, C. (1990). Learner-Centered Testing through Computers: Institutional Issues in Individual Assessment. I J. de Jong & D.K. Stevenson (Red.), *Individualizing the assessment of language abilities*. Clevedon, UK: Multilingual Matters.
- Alderson, C. (2000a). Technology in testing: The present and the future. *System*, 28(4), 593-603. Hämtad 20100829 från <http://www.sciencedirect.com/science/article/pii/S0346251X00000403>
- Alderson, C. (2000b). *Assessing Reading*. Cambridge: Cambridge University Press.
- Alvesson, M. & Sköldbberg, K. (2008). *Tolkning och reflektion. Vetenskapsfilosofi och kvalitativ metod*. Lund: Studentlitteratur.
- Andered, B. (2001). Europarådets Framework – en inspiration för de nya kursplanerna. I *Språkboken. En antologi om språkundervisning och språkinlärning* (s. 26–37). Stockholm: Skolverket.
- Andersson, B-E. (1994). *Som man frågar får man svar – En introduktion i intervju- och enkätteknik*. Mölnlycke: Prisma ePan.
- APA - American Psychological Association. (2010). *PsycINFO Database Record: International Guidelines on Computer-Based and Internet-Delivered Testing*. Hämtat 20111027 från <http://psycnet.apa.org>
- Assessment Reform Group. (2002). *Research-based principles to guide classroom practice. 10 principles*. Hämtat 20111027 från <http://www.aiaa.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- 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.
- Bailey, K.M. (1996). Working for Washback: A Review of the Washback Concept in Language Testing. *Language Testing*, 13(3), 257–279.
- Bailey, K.M. (1999). Washback in Language Testing. *TOEFL Monograph Series*. Educational Testing Service. Princeton, New Jersey. RM-99-4. Hämtat 20111107 från <http://www.ets.org/Media/Research/pdf/RM-99-04.pdf>
- Bjerkestrand, O. (2009). The European Coherent Indicators and Benchmarks and Implications for Computer-Based Assessment. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 24–28). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Bridgeman, B. (2009). Experiences from Large-Scale Computer-Based Testing in the USA. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 39–44). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Börjesson, L. (2009). *Jämförelse mellan skrivningarna i Gemensam europeisk referensram och Kursplaner 2000*. Opublicerat internt utredningsuppdrag. Stockholm: Skolverket.
- Carlsen, C. (2008). The role of testing in an egalitarian society. *CAMBRIDGE ESOL : Research Notes, Issue 34, 2008*. Hämtat 20101212 från http://www.cambridgeesol.org/rt_notes/
- 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.
- Choi, I-C., Kim, K.S. & Boo, J. (2003). Comparability of a paper based language test and a computer based language test. *Language Testing*, 20(3), 295–320.
- Chun, C. (2006). An Analysis of a Language Test for Employment: The Authenticity of the PhonePass Test. *Language Assessment Quarterly*, 2006: 3(3), 295–306.

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. Language policy division. Hämtat 20111025 från http://www.coe.int/T/DG4/Linguistic/Default_en.asp
- Council of Europe. *European Language Portfolio*. Hämtat 20111025 från <http://www.coe.int/t/dg4/education/elp/>
- Davies, A. (2010). Test fairness: a response. *Language Testing*, 27(2), 171–176.
- DIALANG. *About DIALANG*. Hämtat 20111112 från <http://www.lancs.ac.uk/researchenterprise/dialang/about>
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder Education.
- Dörnyei, Z. (2001). *Teaching and Researching Motivation*. Harlow: Longman.
- EALTA —European Association for Language Testing and Assessment, (2006). *Guidelines for Good Practice in Language Testing and Assessment*. Hämtat 20110919 från <http://www.ealta.eu.org/documents/archive/guidelines/Swedish.pdf>
- Eklöf, H. (2006). *Motivational beliefs in the TIMSS 2003 context: Theory, measurement and relation to test performance*. Umeå: Umeå University, Department of Educational Measurement.
- ENOVATE AS. Hämtat 20110929 från <http://enovate.no>
- Erickson, G. (1998). Test-taker feedback – Elevers synpunkter som del av arbetet med nationella prov i engelska. I Ljung, B. & Pettersson, A. (Red.), *Perspektiv på bedömning av kunskap* (s. 79-100). Stockholm: 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. (2009a). "Att bäras åt" — Om den goda bedömningens flerfaldighet och ömsesidighet. I U.Tornberg, et al. *Språkdiraktiska perspektiv. Om lärande och undervisning i främmande språk* (s. 159–174). Stockholm: Liber.
- Erickson, G. (2009b). *Nationella prov i engelska – en studie av bedömersamstämmighet*. Hämtat 20110930 från <http://www.nafs.gu.se/publikationer/>
- Erickson, G. (2009c). *Slutredovisning av en studie av ett nationellt prov i engelska i relation till GERS*. Opublicerat internt utredningsuppdrag. Stockholm: Skolverket.
- Erickson, G. (2010). Good Practice in Language Testing and Assessment —A Matter of Responsibility and Respect. I Kao, T. and Lin, Y. (Red.), *A New Look at Teaching and Testing: English as Subject and Vehicle* (s. 237–358). Taipei, Taiwan: Bookman Books Ltd
- Erickson, G. (2011). Putting the CEFR to Good Use – A Collaborative Challenge. I Mader & Urkun (Red.), *Selected articles by the presenters of the IATEFL Testing, Evaluation and Assessment Special Interest Group (TEA SIG) and EALTA Conference in Barcelona, Spain 29-30 October, 2010* (s. 36–43). Hämtat 20111201 från http://www.ealta.eu.org/documents/resources/IATEFL_EALTA_Proceedings_2010.pdf
- Erickson, G. & Nihlén, C. (1994). PM inför konferens kring nya prov i engelska på grundskolenivå. Skolverket, Stockholm, 21-22 februari, 1994. (Opublicerat material).
- Erickson, G. & Börjesson, L. (2001). Bedömning av språkfärdighet i nationella prov och bedömningsmaterial. I *Språkboken. En antologi om språkundervisning och språkinläring* (s. 255–269). Stockholm: Skolverket.
- 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 20110920 från <http://www.ealta.eu.org/resources.htm>
- ETS/ Educational Testing Service. *About e- rater*. Hämtat 20111012 från <http://www.ets.org/erater/about/>
- ETS/ Educational Testing Service. (2004). *ICT Literacy Assessment*. Hämtat 20111013 från http://www.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/020_2heapaper.pdf

- ETS/ Educational Testing Service. (2009). *Guidelines for the Assessment of Language Learners*. Hämtat 20111113 från http://www.ets.org/Media/About_ETS/pdf/ELL_Guidelines.pdf
- Europarådet. (2006). *Nyckelkompetenser för livslångt lärande – en europeisk referensram*. Hämtat 20111202 från http://ec.europa.eu/dgs/education_culture/publ/pdf/ll-learning/keycomp_sv.pdf
- Fulcher, G. (2000). Computers in language testing. I Brett P. & Motteram, G. (Red.), *A special interest in computers: Learning and teaching with information and communications technologies* (s. 93–107). Manchester: IATEFL publications. Hämtat 20110312 från <http://languagetesting.info/gf/glennfulcher.php#publications>
- Giota, J. (2002). Skoleffekter på elevers motivation och utveckling. *Pedagogisk forskning i Sverige*. Hämtat 20091215 från <http://www.ped.gu.se/dokument/Asa/Publikationer/Joanna/Giota2002b.pdf>
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press.
- Halldórsson, A., McCelvie, P. & Björnsson, J. (2009). Are Icelandic boys really better on computerized tests than conventional ones? — Interaction between gender, test modality and test performance. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 178–193). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Hammersley, M. & Atkinson, P. (1983). *Ethnography: Principles in practice*. London: Routledge.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge Handbooks for Language Teachers. (Second Edition). Cambridge: Cambridge University Press.
- Huhta, A. m.fl. (2002). DIALANG – A Diagnostic Language Assessment System for Adult Learners. I *Council of Europe, Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Case Studies*. Strasbourg: Council of Europe Publishing.
- ILTA – *Code of Ethics*. (2000). Hämtad 20111113 från http://www.iltaonline.com/images/pdfs/ILTA_Code.pdf ,
- Kane, M. (2010). Validity and fairness. *Language Testing* 27(2), 177–182.
- Katz, I. & Macklin, A. (2007). Information and Communication Technology (ICT) Literacy: Integration and Assessment in Higher Education. *SYSTEMICS, CYBERNETICS AND INFORMATICS* 5(4), 50–55. Hämtat 20111009 från [http://www.iiisci.org/journal/CV\\$/sci/pdfs/P890541.pdf](http://www.iiisci.org/journal/CV$/sci/pdfs/P890541.pdf)
- Kikis – Papadakis, K. & Kollias, A. (2009). Reflections on Paper-and-Pencil Tests to eAssessments: Narrow and Broadband Paths to 21st Century Challenges. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 99–103). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Kozma, R. (2009). Assessing and Teaching 21st-Century Skills Assessment Call to Action. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 13–23). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Lie, S. (2008). Kort analyse av data fra Nasjonale prøver 2008. (Material som erhållits vid personlig kontakt med Eli Moe, Universitetet i Bergen)
- Little, D. (2005). *The European Language Portfolio: where pedagogy and testing meet*. Dublin: Trinity College. Hämtat 20111025 från http://www.alte.org/further_info/cardiff/dl_1105.pdf
- Lindblad, T. (1990). Provet och bedömning i främmande språk. I *Undervisning i främmande språk – Kommentarmaterial Lgr80*. (s. 274–298). Stockholm: Skolöverstyrelsen och Utbildningsförlaget.
- Lindblad, T. (1991). Provet i engelska, franska och tyska. I Lindblad, T. (utg.), *"Allt är relativt"*. (Rapport nr 1991:03, s. 97–136). Göteborg: Göteborgs universitet, Institutionen för pedagogik.
- Lindqvist, AK. (2009). *Datorbaserade prov i engelska – en pilotstudie 2009*. Intern rapport till Skolverket i maj 2009. Hämtat 20111025 från <http://www.nafs.gu.se/publikationer/>
- Malmberg, P. (2000). De moderna språken i grundskolan och gymnasieskolan från 1960 och framåt. I *Språk. Grundskola och gymnasieskola. Kursplaner, betygskriterier och kommentarer Gy 2000:18* (s. 7–26). Stockholm: Skolverket.
- Malmberg, P. (2001). Språksynen i dagens kursplaner. I *Språkboken. En antologi om språkundervisning och språkinläring* (s. 16–25). Stockholm: Skolverket.

- Martin, R. & Binkley, M. (2009). Gender differences in cognitive tests: a consequence of gender dependent preferences of specific information presentation formats? I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 75–82). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- 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. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Messick, S. A. (1989). Validity. I Linn, R.L. (Red.), *Educational Measurement*. (Third edition), (s. 13–103). New York: American Council on Education/Macmillan
- Messick, S. A. (1996). Validity and washback in language testing. I *Language testing*, 13(3), 241–256.
- Moe, E. (2008). Teknisk rapport – pilotering 2007 / prøveversionar engelsk 2008. (Material som erhållits vid personlig kontakt med Eli Moe, november 2008.)
- Moe, E. (2009). Introducing Large-Scale Computerized Assessment – Lessons Learned and Future Challenges. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 51–56). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Moe, E., Carlsen, C. & Hasselgren, A. (2006). Digitale leseprøver i engelsk – utfordringer og muligheter. *Norsk Pedagogisk Tidsskrift*, 2006(5), 391–403.
- NAFS-projektet — Nationella Prov i Främmande Språk, Institutionen för pedagogik och specialpedagogik, Göteborgs universitet. *Projektet NAFS - Nationella prov i främmande språk*. Hämtat 20110920 från <http://www.nafs.gu.se>
- OECD. (1999). *Measuring Student Knowledge and Skills – A New Framework for Assessment*. Paris: OECD Publications.
- OECD. (2009). *PISA 2009. Reading Framework*. Hämtat 20111013 från <http://www.acer.edu.au/ozpisa/reading/>
- Oscarson, M. (1999). Estimating language ability by self-assessment: A review of some of the issues. I *Language Learning, Teaching and Assessment. Festskrift till Torsten Lindblad*. IPD-rapport 1999(2), (s. 161-189). Göteborg: Göteborgs universitet, Institutionen för Pedagogik och Didaktik.
- Ripley, M. (2009). Transformational Computer-based Testing. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 92–98). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Shohamy, E. (2001). *The Power of Tests: a critical perspective on the uses of language tests*. Harlow: Pearson Education.
- Sireci, S. & Zenisky, A. (2010). Innovative Item Formats in Computer-Based Testing: In Pursuit of Improved Construct Representation. I Long, M.H. & Doughty, C.J. (Red.), *The Handbook of Language Teaching* (s. 329–347). Chichester: Wiley-Blackwell.
- Skolverket. *Nationella prov*. Hämtat 20111019 från http://www.skolverket.se/prov_och_bedomning/2.1100
- Skolverket. *Sfi - Svenskundervisning för invandrare*. Hämtat 20110920 från http://www.skolverket.se/prov_och_bedomning/2.1100/2.1190
- Skolverket. (2004). *Engelska i åtta europeiska länder: En undersökning av ungdomars kunskaper och uppfattningar*. (Rapport nr 242: Författare Erickson, G.). Stockholm: Skolverket.
- Skolverket. (2005). *Lärare och elever om gymnasieskolans nationella prov – en enkätstudie*. Stockholm: Skolverket. Hämtat 20110930 från http://www.skolverket.se/prov_och_bedomning/2.1278/2.3288
- Skolverket. (2008). *Rapport - Ämnesproven 2008 i grundskolans årskurs 5*. Hämtad 20091113 från <http://www.skolverket.se/sb/d/306/a/14246>
- Skolverket (2009a). *Gemensam europeisk referensram för språk: Lärande, undervisning, bedömning - GERS*. Stockholm: Skolverket.
- Skolverket. (2010). *Adaptiva och andra datorbaserade prov – en kunskapsöversikt*. Peter Nyström (Red.). Stockholm: Skolverket. Hämtat 20111019 från http://www.skolverket.se/prov_och_bedomning/2.1100

- Skolverket. (2011a). *Kommentarmaterial till kursplanen i engelska*. Hämtat 20111019 från <http://www.skolverket.se/2.3894/publicerat/publikationer>
- Skolverket. (2011b). *Eleverna och nätet – PISA 2009 om 15-åringars förmåga att söka, läsa och värdera digital information*. (Rapport 361.) Hämtat 20111101 från <http://www.skolverket.se/publikationer?id=2607>
- Skolverket. (2011c). *Gymnasieskola 2011*. Stockholm: Skolverket.
- Skolverket. (2011d). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket.
- Skolverket. (2011e). *Ämnesplan – Engelska. Gy 2011*. Hämtat 20111130 från http://www.skolverket.se/forskola_och_skola/gymnasieutbildning/2.2954/amnesplaner_och_kurser_for_gymnasieskolan_2011/subject.htm?subjectCode=ENG
- Stobart, G. (2006). The Validity of Formative Assessment. I J. Gardner (Red.), *Assessment and Learning* (s. 133–146). London: SAGE Publications.
- Stricker, L., Wilder, G. & Rock, D. (2004). Attitudes about the computer-based Test of English as a Foreign Language. *Computers in human behavior*, 20/2004, 37–54.
- Stricker, L. & Attali, Y. (2010). *Test Takers' Attitudes About the TOEFL iBT™*. (TOEFL iBT Research Report. TOEFLiBT-13, January 2010). Hämtat 20111003 från <http://www.ets.org/Media/Research/pdf/RR-10-02.pdf>
- Stukát, S. (2005). *Att skriva examensarbete inom utbildningsvetenskap*. Lund: Studentlitteratur AB.
- Svtplay. *Västnytt, 15 november 2011*. Hämtat 20111123 från http://svtplay.se/v/2603618/vastnytt/ale_bland_de_basta_pa_it
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. I *British Journal of Educational Studies*, VOL.53(4), 466–478. Blackwell Publishing Ltd.
- TOEFL – Test of English as a Foreign Language. ETS Testing Service. <http://www.ets.org/toefl>
- Tornberg, U. (2000). *Om språkundervisning i mellanrummet och talet om "kommunikation" och "kultur" i kursplaner och läromedel från 1962 till 2000*. Uppsala: Acta Universitatis Upsaliensis.
- Toulmin, S. (2003). *The uses of argument*, 2nd ed. Cambridge: Cambridge University Press.
- Transportstyrelsen. *Fortsatt stopp för kunskapsprov*. Hämtat 20111123 från <http://www.transportstyrelsen.se/sv/Nyhetsarkiv/Fortsatt-stopp-for-kunskapsprov-pa-forarprovskontoren-i-Sverige/?transportarea=road>
- Utdanningsdirektoratet. *Læreplan i engelsk : LK06*. Hämtat 20091229 från <http://www.udanningsdirektoratet.no/grep/Lareplan/?laereplanid=122422>
- Utdanningsdirektoratet. *Norske nasjonale prøver*. Hämtat 20091212 från <http://www.udanningsdirektoratet.no/Artikler/Nasjonale-prover/Nasjonale-prover-i-regning-og-engelsk-for-5-og-8-trinn-2009/>
- van Ek, J. & Trim, J. (1998). *Threshold 1990*. Cambridge: Cambridge University Press / Council of Europe. Hämtat 20111002 från http://www.ealta.eu.org/documents/resources/Threshold-Level_CUP.pdf
- van Lent, G. (2009). Risks and Benefits of CBT versus PBT in High-stakes Testing. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 83–91). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Velling Pedersen, D. (2004). "Det var kul för man fick tänka efter lite grann". I *Att visa vad man kan – En samling artiklar om ämnesproven i år 5*. Stockholm: Skolverket.
- Vetenskapsrådet (u.å.). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. Hämtat 20090307 från <http://www.codex.vr.se/texts/HSFR.pdf>
- Wallace, M. (1998). *Action Research for Language Teachers*. Cambridge: Cambridge University Press.
- Wandall, J. (2009). National Tests in Denmark – CAT as a Pedagogic Tool. I F. Scheuermann & J. Björnsson (Red.), *The Transition to Computer-Based Assessment* (s. 45–50). JRC Scientific and Technical Reports. Hämtat 20091227 från <http://crell.jrc.it/CRELL%20publications.htm>
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing* 27(2), 147–170.
- Åberg-Bengtsson, L. & Erickson, G. (2006). Dimensions of national test performance: A two-level approach. *Educational Research and Evaluation*. 12(5), 469–488.
- Åhs, M. (2005). *NP – Mer av vad man egentligen kan*. (C-uppsats i ämnesdidaktik.), Göteborg: Göteborgs universitet, Institutionen för pedagogik och didaktik.



GÖTEBORGS UNIVERSITET
INSTITUTIONEN FÖR PEDAGOGIK OCH DIDAKTIK

Vad tycker du?

Elevenkät I – Datorbaserade prov i engelska

Läs påståendena om datorer, engelska och prov och kryssa i hur väl de stämmer in på dig – om du håller med eller inte.

Jag är en kille / tjej i åk _____ på _____ (skola) och mina initialer är _____

	Ja, absolut			Nej, absolut inte	
Jag gillar datorer	_____	_____	_____	_____	_____
Jag är bra på (att använda) datorer	_____	_____	_____	_____	_____
Jag använder ofta dator i skolan	_____	_____	_____	_____	_____
Jag använder ofta dator hemma	_____	_____	_____	_____	_____

Jag sitter ungefär _____ timma/r per dag vid datorn och då håller jag oftast på med _____

Jag gillar engelska	_____	_____	_____	_____	_____
Jag är bra i engelska	_____	_____	_____	_____	_____
Jag använder ofta engelska vid datorn	_____	_____	_____	_____	_____
Jag har gjort språk-övningar på datorn förut	_____	_____	_____	_____	_____

Jag tycker det verkar bra att ha
engelskprov på dator

Jag skulle hellre göra ett vanligt
prov med papper och penna

Jag behöver hjälp för att
kunna göra prov på dator

Jag känner mig beredd på
att göra det här provet

Jag brukar vara nöjd med
hur det går för mig på prov
i engelska

Jag tror att jag kommer att
bli nöjd med hur det går för
mig på det här provet

*Skriv gärna andra kommentarer och tankar om att ha datorbaserade engelskprov i skolan:
(t.ex. om vad som kan vara bra eller dåligt – för dig själv, på din skola eller i allmänhet)*



Tack för hjälpen!



GÖTEBORGS UNIVERSITET
INSTITUTIONEN FÖR PEDAGOGIK OCH DIDAKTIK

Vad tycker du om datorbaserade prov i engelska?

Elevenkät II

Läs påståendena och kryssa i hur väl de stämmer in på dig – om du håller med eller inte.

Jag är en *kille / tjej* i åk _____ på _____ (skola) och mina initialer är _____

	Ja, absolut		Nej, absolut inte		
Jag gillade att göra engelskprov på dator	_____	_____	_____	_____	_____
Det var lätt att förstå hur man skulle göra	_____	_____	_____	_____	_____
Det var till hjälp att göra exempeluppgifterna innan	_____	_____	_____	_____	_____
Jag skulle ha behövt mer hjälp för att kunna göra prov på dator	_____	_____	_____	_____	_____
Jag kunde koncentrera mig när jag gjorde provet	_____	_____	_____	_____	_____
Jag kände mig lugn när jag gjorde provet	_____	_____	_____	_____	_____
Jag kände mig nöjd när jag gjorde provet	_____	_____	_____	_____	_____
Jag lärde mig något när jag gjorde provet	_____	_____	_____	_____	_____
Jag kunde visa hur bra jag är på att läsa och förstå engelska	_____	_____	_____	_____	_____
Jag tror att jag kommer att bli nöjd med mitt resultat på det här provet	_____	_____	_____	_____	_____

	Ja, absolut			Nej, absolut inte	
Det fungerade bra på min skola att göra prov på dator	_____	_____	_____	_____	_____
Det är rättvist att göra prov på dator i skolan	_____	_____	_____	_____	_____
Det är lika lätt för alla att göra prov på dator	_____	_____	_____	_____	_____
Jag tycker det är bra att ha engelskprov på dator	_____	_____	_____	_____	_____
Jag gör hellre vanliga prov med papper och penna	_____	_____	_____	_____	_____
Jag tycker skolor i Sverige ska ha fler datorprov i framtiden	_____	_____	_____	_____	_____

*Förklara eller kommentera gärna dina svar – eller skriv annat om datorbaserade engelskprov:
Vi vill verkligen veta vad du tycker och tänker om detta!*





GÖTEBORGS UNIVERSITET INSTITUTIONEN FÖR PEDAGOGIK OCH DIDAKTIK

Datorbaserade prov i engelska - lärarenkät

Det är mycket värdefullt för oss att få veta hur Du ställer dig till och funderar kring datorbaserade prov i engelska i grundskolan. Vi hoppas att Du vill delge oss Dina synpunkter, gärna utifrån pilotstudiens syfte och mål nedan, men också kring annat som Du vill lyfta fram eller ytterligare information som kan vara till stöd för vår tolkning av pilotstudien. **Du enkelt skriva direkt i enkäten, om Du sparar den på din dator först, och bifoga den i ett mail tillbaka. Vi behöver Ditt svar i vecka 14.**

Pilotstudiens syfte och mål beskrivs på följande sätt i uppdragsbeskrivning respektive planering:

I takt med den tekniska utvecklingen och ökande datortäthet och IT-kunnande inom skolans värld har man i många länder under en längre tid startat olika projekt med datorbaserade prov, däribland adaptiva prov, inom olika ämnen. Det känns därför angeläget att man i Sverige börjar med att bygga upp en kunskapsbas för att sedan med hjälp av en pilotstudie i Sverige inhämta erfarenheter för att kunna skapa ett diskussionsunderlag för framtida beslut.

*Tyngdpunkten ligger på **attityder, uppfattningar och reaktioner** hos elever och lärare och eventuellt skolledare och har fokus på beredskapen för användning av datormediet som sådant i relation till prov och bedömning, dels på ett allmänt plan och dels med fokus på den datorbaserade formen.*

(Ur Överenskommelse och uppdragsbeskrivning, Skolverket)

*Syftet är att i någon mån undersöka **den tekniska och pedagogiska beredskapen** i svenska skolor att genomföra datorbaserade prov i engelska.*

Vi söker svar på frågor kring

- användarmöjlighet och användarvänlighet, det vill säga hur man på olika nivåer upplever att i skolan göra engelskprov på dator.

- förhoppningar och farhågor, känsla av förtrogenhet, förståelse, eventuell ängslan inför och tillfredsställelse med provsituationen, ämnet och datormediet.

(Ur Preliminär planering av pilotstudie i engelska, Enheten för språk och litteratur)

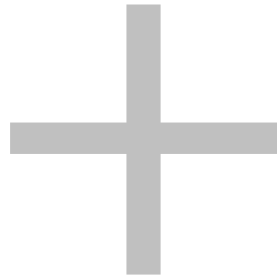
AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedago-gen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>



AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedagogen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>

	Ja, absolut			Nej, absolut inte		
Jag använder ofta datorer i min engelskundervisning						
Jag kommer att använda fler datorbaserade övningar och prov i min undervisning						
De datorbaserade proven var lätta att förstå sig på och använda						
På min skola finns tillräcklig beredskap att genomföra datorbaserade prov						
Jag tycker det finns pedagogiska fördelar med datorbaserade prov						
Jag skulle ställa mig positiv till datorbaserade nationella prov						

Information, till exempel om elevgruppen, som hjälper oss tolka pilotstudien:

Namn: _____



Stort TACK för Din medverkan!

AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedagogen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>



GÖTEBORGS UNIVERSITET INSTITUTIONEN FÖR PEDAGOGIK OCH DIDAKTIK

Datorbaserade prov i engelska - skollädaenkät

Din skola har under vårterminen deltagit i en pilotstudie kring datorbaserade prov i engelska för Skolverket. Det är mycket värdefullt för oss att få veta hur Du som skolläda ställer dig till och funderar kring datorbaserade prov i engelska i grundskolan. Vi hoppas att Du vill delge oss Dina synpunkter, gärna utifrån pilotstudiens syfte och mål nedan, men också kring annat som Du vill lyfta fram eller ytterligare information som kan vara till stöd för vår tolkning av pilotstudien. **Du kan skriva direkt i enkäten, om Du sparar den på din dator först, och bifoga den i ett mail tillbaka. Vi behöver Ditt svar i vecka 15.**

Pilotstudiens syfte och mål beskrivs på följande sätt i uppdragsbeskrivning respektive planering:

I takt med den tekniska utvecklingen och ökande datortäthet och IT-kunnande inom skolans värld har man i många länder under en längre tid startat olika projekt med datorbaserade prov, däribland adaptiva prov, inom olika ämnen. Det känns därför angeläget att man i Sverige börjar med att bygga upp en kunskapsbas för att sedan med hjälp av en pilotstudie i Sverige inhämta erfarenheter för att kunna skapa ett diskussionsunderlag för framtida beslut.

*Tyngdpunkten ligger på **attityder, uppfattningar och reaktioner** hos elever och lärare och eventuellt skolläda och har fokus på beredskapen för användning av datormediet som sådant i relation till prov och bedömning, dels på ett allmänt plan och dels med fokus på den datorbaserade formen.*

(Ur Överenskommelse och uppdragsbeskrivning, Skolverket)

*Syftet är att i någon mån undersöka **den tekniska och pedagogiska beredskapen** i svenska skolor att genomföra datorbaserade prov i engelska.*

Vi söker svar på frågor kring

- användarmöjlighet och användarvänlighet, det vill säga hur man på olika nivåer upplever att i skolan göra engelskprov på dator.*
- förhoppningar och farhågor, känsla av förtrogenhet, förståelse, eventuell ängslan inför och tillfredsställelse med provsituationen, ämnet och datormediet.*

(Ur Preliminär planering av pilotstudie i engelska, Enheten för språk och litteratur)

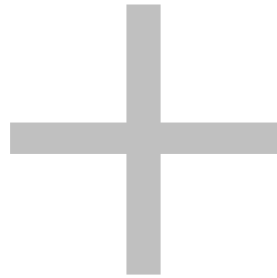
AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedagoen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>



AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedagogen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>

	Ja, absolut			Nej, absolut inte		
På min skola finns tillräcklig beredskap att genomföra datorbaserade prov						
Jag tycker det finns pedagogiska fördelar med datorbaserade prov						
Jag skulle ställa mig positiv till datorbaserade nationella prov						

Ev. övrig information om Din skola som hjälper oss tolka pilotstudien:

Namn: _____



Stort TACK för Din medverkan!

AnnaKarin Lindqvist

Enheten för språk och litteratur

Box 300, SE 405 30 Göteborg (Besöksadress: Pedagogen hus A, Västra Hamngatan 25)

Tel: 031-786 2157 Fax: 031-786 2380 (fax)

<http://www.ipd.gu.se/enheter/sol/> <mailto:annakarin.lindqvist@ped.gu.se>



GÖTEBORGS UNIVERSITET INSTITUTIONEN FÖR PEDAGOGIK OCH DIDAKTIK

Till Rektor och berörda lärare i engelska

Göteborg oktober 2008

Pilotstudie gällande datorbaserade prov i engelska

På uppdrag av Skolverket görs under läsåret 2008/2009 en kunskapsöversikt över forskning om och användning av datorbaserade, adaptiva prov inom och utom Norden. Studien genomförs av Umeå universitet, Institutionen för beteendevetenskapliga mätningar samt Göteborgs universitet, Institutionen för pedagogik och didaktik – Enheten för språk och litteratur <http://www.ipd.gu.se/enheter/sol/>. Uppdraget är ett led i den fortlöpande forskningen kring och utvecklingen av det nationella provsystemet.

Många länder har, i takt med den ökande tekniska utvecklingen, startat projekt med datorbaserade, i en del fall adaptiva, prov i olika ämnen, så kallad CBT (Computer Based Testing) respektive CAT (Computer Adaptive Testing). Adaptivitet innebär här att datorn efter hand, och i någon mån, anpassar svårighetsgraden efter elevens prestation på de tidigare uppgifterna i provet, och således att elever får individuella prov. Inom ramen för denna kunskapsöversikt har erfarenhetsutbyte skett med bl.a. Danmark, Norge och Finland. Exempeluppgifter till Norges datorbaserade nationella prov i engelska finns på http://www.utdanningsdirektoratet.no/templates/udir/TM_Artikkel.aspx?id=3548. Dessa ska dock inte delges elever eftersom vår studie delvis handlar om elevers beredskap inför datorbaserade prov.

Inom NAFS-projektet (Nationella Prov i Främmande Språk) vid Göteborgs Universitet, som bland annat utvecklar de nationella provmaterialen i engelska, franska, spanska och tyska, finns en lång tradition av samarbete med elever och lärare kring frågor som användbarhet, användarvänlighet och likvärdighet. När det gäller den här aktuella pilotstudien i engelska har vi valt att ställa en förfrågan om deltagande till åtta grundskolor. Målgruppen för studien är elever i såväl grundskolans äldre som yngre åldrar, samt deras lärare och skolledare.

I praktiken kommer vi alltså ut till er skola under en eller två halvdagar i februari eller mars 2009 och genomför pilotstudien - men vi behöver få hjälp med att planera, låna elever samt få ta del av er erfarenhet och era synpunkter.

Preliminär planering:

Enkäter före genomförandet med elever/lärare/skolledare (ca 20 min per enkät)

Genomförande av datorbaserade provuppgifter (CBT och/eller CAT) i två klasser (ca två lektioner per klass/elev)

Iakttagande studier under genomförandet

Enkäter och ev. samtal efter genomförandet (ca 30 min/enkät eller samtal)

Vi ser mycket fram emot att få samarbeta med er och kommer att kontakta er för ett inledande planeringsbesök under december eller januari.

Med vänliga hälsningar

AnnaKarin Lindqvist

forskare

Kontaktperson

annakarin.lindqvist@ped.gu.se

031-786 21 57

Gudrun Erickson Biträdande

Universitetslektor

Projektledare

gudrun.erickson@ped.gu.se

031-786 24 53

Enheten för språk och litteratur
Pedagogen hus A, Västra Hamngatan 25, Box 300, SE 405 30 Göteborg