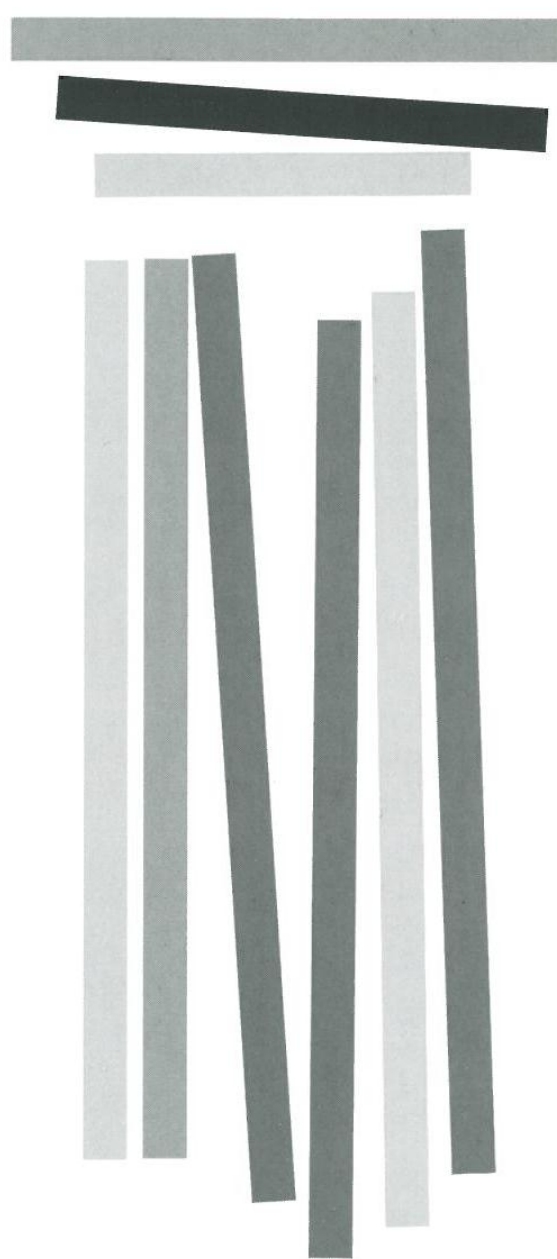


SPROGFORUM

Tidsskrift for sprog- og kulturpædagogik · Maj · 2002 · Nr. 23

Evaluering



SPROGFORUM

Tidsskrift for sprog- og kulturpædagogik · xxx · 2002 · Nr. 23

Nr 23, 2002: Tema: Evaluering - hvad skal vi kunne?



In English
Please! 

Redaktionen for nr. 23: Bodil Bjerregaard, Birte Dahlgreen,
Marianne Ledstrup, Elsebeth Rise, Peter Villads Vedel

Indhold

Forord

[Evaluering - hvad skal vi kunne?](#)

Kronikken

[Michael Svendsen Pedersen: Hold på briller, professionalisme og arbejde](#)

[Sari E. Luoma: Hvad sker der inden for sprogtestning?](#)

[Marianne Hirtzel: Den Europæiske Sprogportfolio](#)

[Aase Brick-Hansen: Folkeskolens afsluttende prøver i engelsk, tysk og fransk](#)

[+ bilag](#)

[Jens Peder Weibrecht: Prøveformer med IT i fransk og spansk](#)

Elana Shohamy: Demokratiske og udemokratiske dimensioner i evaluering

[Else Lindberg: Studieprøven i dansk som andetsprog - en ny prøve](#)

[Else Brendholdt, Tage Øe & Anders Hove: Skriftlig eksamen i engelsk i læreruddannelsen](#)

[+ bilag](#)

[Nick Saville: Kvalitet og retfærdighed](#)

[Karen Sonne Jakobsen: Evaluering og prøver i projektorganiserede sprogstudier](#)

[Vibeke Toft Jørgensen: Genren hedder eksamensopgave](#)

[Ordliste](#)

[Godt Nyt](#)

Forord

Den danske pædagogiske tradition, inkl. sprogpædagogikken, har ikke interesseret sig særlig meget for spørgsmål vedrørende evaluering og testning. Kan det være at man i Grundtvigs fædreland har været tilbageholdende med at beskæftige sig med evaluering og evalueringsformer, et felt der forbindes med kontrol og for ensidig kvantificering?

Imidlertid har den moderne evalueringsforskning udviklet sig kraftigt i de sidste årtier, og visse steder i udlandet har man udviklet nogle nye perspektiver som det vil være meget spændende at inddrage i den danske diskussion om udvikling af sprogundervisning og kommunikativ kompetence i fremmedsprogene og i dansk som andetsprog. Derfor har vi valgt at udgive et nummer af **Sprogforum** der fokuserer på evaluering af kommunikativ kompetence, og hvor vi bl.a. trækker på nogle udenlandske evalueringsforskere.

En af de nye udviklinger er læringsrelateret evaluering, og dette syn på evaluering er også ved at brede sig i Danmark: Hvor en traditionel prøve ligger ved afslutningen af et forløb og kontrollerer/måler de opnåede resultater, er læringsrelaterede evalueringsformer led i læreprocessen: Læreren bruger evalueringen til sproglig vejledning, sprogtilegneren evaluerer sig selv (*self-assessment*), eller sprogtilegnerne evaluerer hinanden (*peer-assessment*). De pågældende kan f.eks. evaluere sig selv ved at føre dagbog/logbog eller ved at arbejde på deres sprogportfolio. De kan f.eks. evaluere hinanden i forbindelse med projektarbejde m.v.

Noget relativt nyt er også computerbaseret fleksibel evaluering: Sprogtilegneren arbejder sig gennem et softwareprogram der er indrettet sådan at det afpasser sig efter brugerens niveau: Hvis programmet konstaterer at sprogtilegneren (test-tageren) har gode scorere, begynder det at stille vanskeligere spørgsmål; hvis det konstaterer at brugeren scorer lavt, sænker den niveauet for spørgsmålene.

Nogle evalueringsforskere, bl.a. israeleren E. Shohamy, søger at sætte brugen af evaluering, prøver og tests ind i en større samfundsmæssig sammenhæng: Hvordan bruger man prøver i praksis? Hvilken indvirkning har prøver på sprogundervisningen og på sprogtilegnerne? Hvilken politisk og sprogpolitisk betydning har prøverne, både indholdsmæssigt og formmæssigt? Hvad kan man forstå ved en evalueringsetik? Denne diskussion ville det også være spændende at have i Danmark.

Gennem Europarådets forskellige sprogprojekter er der ved at tegne sig en fælleseuropæisk politik vedrørende kriterier for evaluering af kommunikativ kompetence. Den har bl.a. udmøntet sig i publikationen *Common European Framework of Reference for Languages* og i udviklingen af Den Europæiske Sprogportfolio. Desuden arbejder den europæiske sammenslutning af sprogtestere (ALTE) med at udvikle moderne sprogprøver m.v. Disse tiltag kommer også til at få betydning for sprogevaluering og sprogundervisning i Danmark.

Vi lægger ud med **Sari E. Luomas** artikel *Hvad sker der inden for sprogtestning?* Luoma tager afsæt i de

senere års fokus på sprogevaluering og opridser nogle centrale ændringer, herunder især udviklingen inden for anvendelsen af computerbaserede fleksible test- og evalueringsinstrumenter.

Marianne Hirtzel beskriver i sin artikel *Den Europæiske Sprogportfolio* portfolioprojektets forbindelse til *Common European Framework* og bestræbelserne på at udarbejde et fælleseuropæisk system.

Aase Brick-Hansen beskæftiger sig i sin artikel *Folkeskolens afsluttende prøver i engelsk, fransk og tysk* med prøvebekendtgørelsen og behovet for en revidering af denne ud fra de hidtidige erfaringer med prøverne.

Jens Peder Weibrecht afsøger i sin artikel *Prøveformer med IT i fransk og spansk* internettets rolle for udvikling af nye prøveformer med udgangspunkt i eksempler fra forsøg ved årsprøver og studentereksamen.

Elana Shohamy tager i artiklen *Demokratiske og udemokratiske dimensioner i evaluering* afsæt i et brugerorienteret perspektiv der lægger vægt på kort- og langsigtede samfundsmæssige konsekvenser af anvendelsen af tests inden for uddannelse, og nyere bestræbelser på at fremme testningens demokratiske dimensioner via kritisk testning og ud fra demokratiske bedømmelsesmodeller.

Else Lindberg giver i sin artikel *Studieprøven i dansk som andetsprog - en ny prøve* en karakteristik af den nye studieprøve i dansk som andetsprog der svarer til den hidtidige Danskprøve 2.

Else Brendholt, Tage Øe og Anders Hove skildrer i deres artikel *Skriftlig eksamen i engelsk i læreruddannelsen* opgaveudvalgets arbejde med at fastlægge kriterier og udarbejde modelsæt i forbindelse med formuleringen af eksamensopgaver i engelsk og fremlægger og begrundet resultaterne herfra.

Nick Saville skildrer i artiklen *Kvalitet og retfærdighed. ALTE's Code of Practice* det europæiske samarbejde omkring sprogevaluering der er forankret i ALTE . Foreningens bestræbelser på at fastlægge et fælles kodeks for praksis behandles indgående inden for de fire hovedområder: Udvikling af eksaminationer, tolkning af eksamensresultater, bestræbelse på retfærdighed og informering af eksaminander.

Karen Sonne Jakobsen gennemgår i sin artikel *Evaluering og prøver i projektorganiserede sprogstudier* projektevalueringen som kompleks og kommunikativt krævende form og plæderer for at principper fra denne evalueringsform bliver anvendt ved evaluering af sprogfagenes færdighedsside.

Vibeke Toft Jørgensen kredser i sin artikel *Genren hedder eksamensopgave. Om den skriftlige prøve i engelsk til lærereksamen* om eksamensopgaven som genre og problematiserer genrebevidsthedens begrænsninger ud fra opgave- og bedømmelseseksempler i relation til den nye skriftlige lærereksamen.

For at præsentere og definere en række centrale begreber har **Michael Svendsen Pedersen** udarbejdet en *Ordliste*.

Kronikken *Hold på briller, professionalisme og arbejde* i dette nummer er skrevet af **Michael Svendsen Pedersen** og diskuterer det implicitte sprogtilegnelsessyn der ligger i de aktuelle integrationspolitiske udspil.

Redaktionen

No. 23, 2002: Theme: Assessment - what should we know?



Editors for no. 23:

Bodil Bjerregaard, Birte Dahlgreen, Marianne Ledstrup, Elsebeth Rise, Peter Villads Vedel

Contents

Foreword: [Assessment - what should we know?](#)

Chronicle: **Michael Svendsen Pedersen:** Keeping sight of professionalism and work

[Sari E. Luoma: What is happening in language testing?](#)

[Marianne Hirtzel: The European Language Portfolio](#)

Aase Brick-Hansen: 'Folkeskole' final exams in English, German and French

Jens Peder Weibrecht: Types of tests in French and Spanish involving IT

Elana Shohamy: Democratic and undemocratic dimensions of assessment

Else Lindberg: Study test in Danish as a second language a new test

Else Brendholdt, Tage Øe & Anders Hove: The written exam in English at colleges of education in Denmark

[Nick Saville: Quality and Fairness - the ALTE Code of Practice and Quality Management Systems](#)

Karen Sonne Jakobsen: Assessment and tests in project-organised language studies

Vibeke Toft Jørgensen: The name of the genre is exam questions

Word List

Good news

Foreword

Assessment - what should we know?

The Danish pedagogical tradition, including language pedagogics, has not been particularly interested in questions related to assessment and testing. Can it be that people in Grundtvig's native land have been reluctant to deal with various forms of assessment a field connected in people's minds with control and one-sided quantification?

Modern assessment research has, however, developed by leaps and bounds over the past decades, and at certain places outside Denmark some new perspectives have been developed that it would be very exciting to include in the Danish discussion concerning the development of language teaching and communicative competence in foreign languages and in Danish as a second language. For this reason, we have chosen to publish an issue of Sprogforum that focuses on the assessment of communicative competence, in which we i.a. include the findings of a number of foreign researchers into assessment.

One of the most recent developments is learning-related assessment, a view of assessment that is also becoming more widespread in Denmark: Where a traditional test marks the conclusion of a course and controls/measures the achieved results, learning-related forms of assessment are a part of the learning process: The teacher uses the assessment for language guidance, the language learner assesses himself (self-assessment), or the language learners assess each other (peer assessment). Those involved can, for example, assess themselves by keeping a diary/logbook, or by working on their language portfolio. They can, for example, assess each other in connection with project work, etc.

Something relatively new is also computer-based flexible assessment: The language learner works his way through a software program that is so designed that it adapts itself to the learner's level: If the program registers that the language learner (test-taker) has good scores, it begins to ask more difficult questions; if it registers that the user is getting poor scores, it lowers the level of the questions.

Some assessment researchers, including the Israeli E. Shohamy, attempt to place the use of assessments, trials and tests in a larger societal context: How are tests used in practice? What effect do they have on language teaching and on language learners? What political and linguistic importance do they have in terms of both form and content? What can assessment ethics be said to mean? It would also be exciting to have this discussion in Denmark.

Via the various projects of the European Council a common European policy is emerging concerning criteria for the assessment of communicative competence. This has i.a. resulted in the publication Common European Framework of Reference for Languages and in the development of The European Language Portfolio. In addition, the European association of language testers (ALTE) is now developing modern language tests, etc. These measures will also be of importance for language assessment and language teaching in Denmark.

We begin with Sari E. Luoma's article What is happening in language testing? Luoma takes as her point of departure the focus of recent years on language assessment, outlining certain key changes, including in particular the development of the use of computer-based flexible instruments of testing and assessment.

In her article The European Language Portfolio, Marianne Hirtzel describes the connection of the portfolio project with Common European Framework and the efforts to devise a common European system.

Aase Brick-Hansen, in her article 'Folkeskole' final exams in English, German and French, deals with the test executive order from the Ministry of Education and the need to revise this on the basis of examinations so far.

In his article 'Types of tests in French and Spanish involving IT', Jens Peder Weibrecht looks at the role of the Internet in the development of new types of tests, using examples from experiments with annual exams and upper secondary leaving exams as his point of departure.

In her article 'Democratic and undemocratic dimensions in assessment', Elana Shohamy has as her starting point a user-oriented perspective that stresses short-term and long-term societal consequences of the use of tests within education as well as recent attempts to promote democratic dimensions of testing via critical testing, on the basis of democratic models of assessment.

Else Lindberg, in her article Study test in Danish as a second language a new test, outlines the new study test in Danish as a second language that corresponds to the present Danish Test 2.

In their article The written exam in English at colleges of education in Denmark, Else Brendholt (før med - dt! we will have to check with Annegret. Both varieties have, unfortunately, been used in the issue)), Tage Øe and Anders Hove give an impression of the work of the exam committee in laying down criteria and constructing sets of questions in connection with the formulation of examination assignments in England, presenting and motivating the results from this work.

In his article Quality and Fairness - the ALTE Code of Practice and Quality Management Systems, Nick Saville outlines the cooperation taking place within language assessment that is anchored in ALTE. The association's efforts to determine a common codex for practice is dealt with in detail within the four main areas: Development of examinations, interpretation of exam results, efforts to ensure impartiality and informing of candidates.

In her article Assessment and tests in project-organised language studies, Karen Sonne Jakobsen outlines the project assessment as a complex and communicatively demanding form, arguing in favour of principles from this form of assessment being used when assessing the proficiency side of language subjects.

In her article The name of the genre is exam questions. Concerning the written exam in English at colleges of education, Vibeke Toft Jørgensen looks at the exam question as a genre, discussing the limitations of genre awareness on the basis of assignment and assessment examples in relation to the new written exam at colleges of education.

To present and define a number of key concepts, Michael Svendsen Pedersen has drawn up a Word List.

The feature article Keeping sight of professionalism and work in this issue has been written by Michael Svendsen Pedersen. It discusses the implicit view of language acquisition that underlies the present proposals concerning integration policy.

The Editors

Translated by John Irons

Tema: Evaluering - hvad skal vi kunne?

kronikken



Hold på briller, professionalisme og arbejde

Michael Svendsen Pedersen
Lektor ved Institut for Uddannelsesforskning,
Roskilde Universitetscenter.

Hvordan lærer man bedst dansk?

Dette *pædagogiske* spørgsmål, som undervisere i dansk som andetsprog dagligt stiller sig i forbindelse med planlægningen af deres undervisning, er blevet et *politisk* spørgsmål efter at integrationsminister Bertel Haarder har spillet ud med forslag til en ny integrationspolitik der indebærer at flygtninge og indvandrere hurtigst muligt skal i arbejde/praktik. Forslaget har ganske vist selvforsørgelse som omdrejningspunkt: "Når jeg er integrationsminister, så er jeg minister for at få folk til at klare sig selv og ikke for noget som helst andet." [note 1](#), men "ikke noget som helst andet" kommer faktisk til også at omfatte et syn på sprogtilegnelse og sprogpædagogisk faglighed inden for danskundervisningen.

Med disse emner på den politiske dagsorden er det vigtigt at dem der arbejder med dansk som andetsprog for voksne, gør sig deres sprogfaglige position klar og blander sig i debatten, dels for at forhindre at der bliver truffet politiske beslutninger på grundlag af hverdagsforestillinger om sprogtilegnelse; dels for at pege på mulighederne for en fortsat professionalisering af undervisningen inden for de nye rammer der er ved at blive udstukket. [note 2](#)

Om tilegnelsen af dansk siger Bertel Haarder: "Man lærer mere om et fremmed samfund inklusive sproget ved at deltage i arbejdslivet og tale med danskerne frem for at sidde i klasselokalerne på sprogcentre med en hel masse andre udlændinge." [note 3](#) Om sprogcentre hedder det at "sprogskolernes faktiske monopol på danskundervisning [brydes]". [note 4](#) Om underviserne antydes det at de kunne være pensionerede dansklærere fra folkeskolen. [note 5](#)

Hermed ligger et udspil til diskussion af hvordan man lærer sprog, og til diskussion af hvad det er for en faglighed lærerne skal have for at udføre et stykke professionelt arbejde.

Sprogtilegnelsessyn

Lærer man sprog ved at bruge det på arbejdspladsen - og i andre daglige sammenhænge? Ja, det kan man godt, men det er samtidig vigtigt at understrege at et "bad i sproget" - måske midt i maskinstøj eller høje barnestemmer - langt fra er tilstrækkeligt til at tilegne sig et velfungerende sprog som man kan klare sig tilfredsstillende med på en dansk arbejdsplads og i det danske samfund som helhed. Mange sproglærere har deres hyr med at bryde et 'flydende forkeret' sprog op som flere af deres kursister har tilegnet sig på en arbejdsplads. For at *brugen* af sprog på arbejde kan føre til *tilegnelse* af et udviklet sprog, er det nødvendigt at sprogbrugeren får mulighed for at gøre arbejdssituationen - eller alle andre situationer i hverdagslivet for den sags skyld - til *læringssituationer*. Sproget på arbejde skal være et sprog der sættes i arbejde, sprogbrugeren skal blive *sprogtilegner*, og det forudsætter at han/hun kan planlægge, bearbejde og reflektere over sine kommunikative erfaringer. Hertil kræves for langt de fleste sprogtilegneres vedkommende hjælp fra professionelle lærere i dansk som andetsprog.

Der vil sikkert være mange praktiske problemer der skal løses i forbindelse med at gennemføre sprogundervisning på arbejdspladsen, og i mange tilfælde vil det nok også være relevant med sprogundervisning som *forberedelse* til arbejdsmarkedet, som et *supplement* til arbejdet osv., men nok så afgørende er det at tage stilling til hvad det er for nogle sproglige kompetencer kursisterne skal have gennem undervisningen.

Det moderne arbejdsmarked kræver omfattende kompetencer - både af den faglige og af den 'bløde' slags - og skal flygtninge og indvandrere have en chance her, rækker det ikke med 'pizza-dansk'; der er brug for sproglige kompetencer af et vist niveau, og de skal omfatte alle aspekter af livet på arbejdspladsen: kantinesnak, deltagelse i møder, sikkerhed, fagforening osv.

Det kan desuden ikke være perspektivet at livet og integrationen skal stoppe ved arbejdspladsen. For mange har det indtil nu været vanskeligt at komme *ind* på en arbejdsplads. Hvis intentionen i den nye integrationspolitik med at skaffe arbejds- og praktikpladser lykkes, kan det (heller) ikke være meningen at livet, integrationen og sprogundervisningen ikke skal pege *ud over* arbejdspladsen. Også her er der brug for en professionel sprogundervisning der kan give flygtninge og indvandrere mulighed for at klare mere end dagen og vejen i mange forskellige hjørner af et moderne og kompliceret samfund som det danske, og rammerne for en sådan undervisning ligger allerede i de eksisterende fagbeskrivelser: Undervisningen på sprogcentre skal - ud over at bidrage til at kursisterne kan varetage et erhvervsarbejde - give dem mulighed for at fungere som samfundsborgere og deltage i uddannelse.

Professionalisme

Den professionalisme der er brug for, findes rundt om på landets sprogcentre. Med 'Lov om dansk som andetsprog for voksne udlændinge m.fl. og sprogcentre', der trådte i kraft 1999, er undervisningen i dansk som andetsprog blevet en integreret del af det danske uddannelsessystem; og lærerne, der har dansk som linjefag eller er cand. mag. i mindst ét sprog, har en 1-årig overbygningsuddannelse fra universitetet. Derudover deltager mange lærere på sprogcentre i kurser i et omfattende efteruddannelsesprogram som er specielt rettet mod dansk som andetsprog for voksne. Til efteråret bliver der således tilbudt et kursus der har "sproget på arbejde" som sit fokus.

Den pædagogiske diskussion og udvikling der foregår inden for området, trækker på den ene side på mange års erfaringer med danskundervisning for voksne udlændinge i de forskellige sammenhænge hvor denne undervisning indtil nu er foregået, og på den anden side på den nyeste viden om sprogtilegnelse og sprogpædagogik - både inden for andetsprog og fremmedsprog.

Når sprogcentrenes monopol på danskundervisning brydes, bliver det helt afgørende at sikre at der ikke gives køb på kravene til undervisernes professionalisme, og med al respekt for pensionerede dansklærere fra folkeskolen, er det en anden professionalisme de har.

En udfordring

Nogle af nøgleordene i sprogcentrenes pædagogik er fleksibilitet, virkelighedsorientering, kommunikative problemløsningsopgaver, sproglig opmærksomhed, ansvar for egen læring - alt sammen det bedste udgangspunkt for også at sætte sproget i arbejde på arbejdet. Men med integrationspolitikken forankring i arbejdslivet skal mange sproglærerne tilegne sig ny viden og nye kompetencer oven på de eksisterende, og det kan helt givet blive en anledning til en inspirerende videreudvikling af eksisterende pædagogiske kommunikative praksisser.

For nyligt kunne nyhedsmedierne meddele at en del udlændinge har svært ved at følge undervisningen på sprogcentre fordi de ser og hører dårligt. Det kan klares med briller og høreapparater. Derudover kræver opbygningen af en velfungerende undervisning i dansk som andetsprog professionelle sproglærere - og meget gerne et arbejde.

Noter

- 1 Borberg, Karin Ryø: Integration handler om at klare sig selv. i: Månedsmagasinet Lederne. Nr. 4, april 2002. Udgives af Ledernes Hovedorganisation.
- 2 Pedersen, Michael Svendsen: Sproget på arbejde. i: Sprog & Integration, nr. 1 maj 2002. Udgives af UC2. www.uc2.dk
- 3 Haarder, Bertel (2002): Integration er hårdt arbejde. Kronik i Politiken 5. marts.
- 4 Regeringen: På vej mod en ny integrationspolitik. 5. marts 2002.
- 5 Samme kilder som i note 1 og 2.

Tema: Evaluering - hvad skal vi kunne?



Hvad sker der inden for sprogtestning?

Sari E. Luoma

**Forsker på Jyväskyläs Universitet i Finland
og evalueringsspecialist ved Defense Language
Institute Foreign Language Center.**

Sprogundervisningens og sprogtestningens verdner har ændret sig radikalt inden for de sidste 20-30 år. Med den kommunikative revolution blev sprogtilgængers erfaringer med at lære sprog i institutionaliserede sammenhænge eksplicit forbundet med *brugen af sproget*, og lige siden er der blevet introduceret nye måder til at hjælpe sprogtilgængere med at opnå kommunikative færdigheder. Vi har set hvordan der bl.a. er blevet lagt vægt på dialoger og samarbejde, anvendelse af sproget i bestemte kontekster, strategi-træning, og udvikling af interkulturel bevidsthed.

Med en vis forsinkelse og i modificeret form har de samme træk fundet vej til praksis inden for testning og evaluering. Forsinkelsen kan delvis forklares med den iboende konservatisme der ligger i testning som en normbaseret aktivitet. Men den skyldes også til en vis grad lærer-bedømmerens afgørende rolle i planlægningen og gennemførelsen af evaluering, kombineret med en bevidst eller ubevidst frygt for de strenge krav til korrekthed og retfærdighed som ligger i evaluering. Det er nemmere at følge en gammel praksis og tro på at den har stået sin prøve, end det er at se sin frygt og usikkerhed om evaluering i øjnene. Denne situation er imidlertid ved at ændre sig.

Mens hovedstrømmen stadig lægger vægt på måling og standarder, er der inden for det seneste årti sket ændringer i evalueringsteorier og -praksisser som berører alle områder inden for metode, teori og aktiviteter (en tilgængelig oversigt findes f.eks. i Birenbaum 1996), og det bliver stadig tydeligere at der i dag er flere forskellige strømninger inden for sprogtestningen. Fra at have været begrænset til *måling* har sprogtestning udvidet sig til *sprogevaluering*, et område med funktionelle tilknytninger til sproguddannelse, en række områder inden for anvendt lingvistik og uddannelsesevaluering.

I denne artikel vil jeg diskutere nogle af de ændringer der har fundet sted inden for sprogtestning og -evaluering i den 'traditionelle' betydning og i den nyere, bredere betydning. Artiklen fokuserer på problemstillinger og implikationer der er relevante for sproguddannelse, men jeg vil også omtale nye artikler og oversigter som giver en mere detaljeret forklaring af særlige udviklinger.

Teknologisk udvikling: Computer-baseret fleksibel testning

Computere og internet udgør en vigtig kontekst for sprogtilgneres brug af sprog i dag. Dette gælder især for sprogundervisning i engelsk, skønt anvendelsesområdet, især med hensyn til specialiseret sprogindlæring, også er blevet udviklet til andre sprog. De spændende muligheder for sprogevaluering som anvendelsen af computere giver, ligger i udviklingen af computer-baserede fleksible test- og evalueringsinstrumenter.

Computerbaserede fleksible evalueringsværktøjer indeholder for-programmerede opgavesæt som kan organiseres efter emne, færdighed der skal testes, og/eller særlige emneområder som skematiske spørgsmål-svar-sekvenser. Derudover gives der oplysninger om måling i forbindelse med opgaverne. Disse oplysninger angiver hyppigst hvor vanskeligt et item har vist sig at være i præ-tests og analyser.

Idéen med computer-baserede fleksible tests er at opgavegivingen bliver tilpasset den enkelte sprogtilgneres svar. Tilegneren får først en opgave - også kaldet et *item* - på mellemniveau. Hvis han eller hun har løst opgaven rigtig, giver programmet en sværere opgave. Hvis også denne besvarelse er korrekt, vælger programmet et endnu sværere item. Hvis tilegneren nu laver fejl, giver programmet et lettere item og begynder at indkredse tilegnerens færdighedsniveau. En anden tilegner kan få den samme opgave til at begynde med, men kan - afhængig af sit besvarelsesmønster - få helt andre sæt af items. Programmet kan ikke desto mindre bruge de målings-oplysninger der er knyttet til de enkelte items, til at give en endelig karakter til begge tilegnere på den samme skala.

Det statistiske grundlag for at lave computer-baserede fleksible tests er for det meste en eller anden form for *Item Responsteori* (IRT). IRT-modeller er baseret på sandsynligheden for korrekt svar, med andre ord sandsynlighedsteori; [note 1](#).

Fordelene ved computer-baseret fleksibel testning er bl.a. at tests kan være kortere når den ekstra tid der går med at besvare items som er for nemme eller for vanskelige for sprogtilgneren, spares, og at resultater fra tests der er blevet givet på forskellige tidspunkter, er sammenlignelige. Når først et lager af items er blevet afprøvet og analyseret således at item-værdierne er kendte, kan enhver kombination af dem gives til nye test-tagere. Test-sikkerheden bliver også forbedret når indholdet i en hvilken som helst ny test er uforudsigelig.

Der er imidlertid også nogle ulemper. Den vigtigste praktiske ulempe er omfanget af det arbejde der kræves for at oprette et item-lager til en fleksibel test. Det antal af sprogtilegnere der skal bruges til at afprøve et item-lager ordentligt, kan tælles i hundreder, eller for nogle statistiske modellers vedkommende i tusinder. Udviklingen af fleksible tests er med andre ord ikke noget en hvilken som helst lærer umiddelbart kan gå i gang med.

Hvis brugbare fleksible evalueringssystemer imidlertid allerede eksisterer, kan lærere afprøve dem på deres sprogtilegnere og evaluere deres brugbarhed. Tilegnerne vil sandsynligvis finde det interessant at finde ud af hvordan de bliver indplaceret af forskellige programmer, men man kan også bede dem reflektere over andre aspekter af evalueringsprocessen, som f.eks. hvor sjov den har været, og hvor nyttig den har været som en støtte til deres læring. [note 2](#)

Foruden *Item Responsteori* og dens anvendelse i fleksibel testning har der også været andre tekniske og statistiske udviklinger inden for sprogtestning i de seneste 30 år. For dem som er interesseret i disse aspekter af sprogevaluering, findes korte og præcise opsummeringer af udviklingerne i Bachman & Eignor 1997 og Bachman 2000. Ud over information giver disse artikler også yderligere henvisninger til specifikke statistiske teknikker.

Teoretiske udviklinger: Fremgang for validering

Inden for sprogtestningsteori er de vigtigste udviklinger i de seneste årtier sket inden for validering. Validitet har altid været et af de centrale kriterier for kvalitetsmåling i testning, men i første halvdel af det 20. århundrede blev det normalt opfattet som et uproblematisk begreb. Spørgsmålet var om testen udfyldte sin opgave, og dette

spørgsmål blev besvaret ved at sammenholde test-resultaterne med evaluering af test-tagerens præstation i forbindelse med den opgave som testen rettede sig mod i den virkelige verden. Nutidens validitets-begreb er imidlertid meget bredere. Det drejer sig om hvorvidt testen kan forsvares videnskabeligt og socialt. Dette lyder flot, men hvad test-udviklerens arbejde angår, drejer det sig faktisk om nogle ret praktiske spørgsmål. Validitet har med test-resultaternes betydning at gøre, og det centrale spørgsmål er: "Hvad er test-måling?" efterfulgt af: "Hvordan kan man vide det?".

Valideringsarbejdet begynder med test-udviklerens definition af den færdighed som skal evalueres i testen. Rent operationelt bliver færdigheden indarbejdet i opgaverne, evalueringskriterierne og de faktiske aktiviteter der udføres under evalueringen. Disse aktiviteter må nødvendigvis undersøges med henblik på at tjekke at den færdighed man regner med bliver evalueret i testen, også rent faktisk er den som testen retter sig mod. Men i nutidens teorier bliver det understreget at udviklerne må gå længere end det, at de må være i stand til at sige hvad resultaterne betyder.

Skønt dette er et ganske udfordrende krav, er test-udviklerne i en ideel position hvad angår beskrivelsen af de færdigheder de evaluerer, fordi de under udviklingen arbejder med test-opgaverne og evalueringskriterierne i lang tid. I dette arbejde er udviklerne bevidst opmærksomme på *hvordan* den færdighed *det er hensigten* at evaluere, bedst kan operationaliseres inden for evalueringssituationens praktiske begrænsninger. 'Det eneste' de behøver at gøre for at dette arbejde kan blive til en del af valideringen, er at nedskrive deres idéer og grundene til deres beslutninger. Arbejdet fortsætter i form af empiriske undersøgelser af om opgaverne og evalueringskriterierne faktisk indarbejder disse intentioner. Validering indebærer med andre ord mange forskellige metodologier, der alle har som mål at vise hvad testen tester.

Da valideringen imidlertid ikke alene fokuserer på hvad testen tester, men også på betydningen af scoren, bliver de sociale dimensioner i evaluering og score også relevante. Fra en filosofisk synsvinkel indebærer dette en erkendelse af at evaluering er en menneskelig aktivitet der udføres på grundlag af velovervejede vurderinger. Fra en samfundsmæssig synsvinkel indebærer det en erkendelse af at test-scoring ofte udgør et grundlag for samfundsmæssige beslutninger. Når scoren bliver brugt - til at give karakter eller til at beslutte hvorvidt en person er kvalificeret til at studere på et universitet f.eks. - har dette konsekvenser for test-tagerne, og evalueringsudviklere er ifølge den aktuelle validitetsteori delvis ansvarlige for disse konsekvenser. De deler ganske vist ansvaret med dem der bruger scoren, f.eks. de universitetskommissioner der udvælger studerende, men de kan ikke fuldstændig undgå at have et ansvar.

Valideringsarbejde er vigtigt fordi det handler om kvalitet og om ansvarlighed. Hvis selvrefleksionsaspektet af validering tages alvorligt, hjælper dette testudviklerne med at forbedre deres tests. I sin nuværende komplekse form står valideringsteori imidlertid over for en interessant udfordring. Det er nemlig ikke klart hvem det er der har brug for de komplekse valideringsrapporter og -studier som den nuværende teori synes at stille krav om. Hvem andre end en anden tester vil være i stand til at læse rapporterne og være i stand til at vurdere om testen er omhyggeligt udarbejdet og tester det den siger den tester? Den nuværende tendens inden for valideringsteori er derfor at gøre konceptet klart og operationaliserbart [note 3](#).

Når validering bliver opfattet som en proces - hvor man bevæger sig fremad ét skridt ad gangen - fører den til ærlige forsøg på at gøre et godt stykke bedømmelsesarbejde - med andre ord give en ansvarlig (*accountable*) bedømmelse. Argumentet for at fortsætte med forskellige former for valideringsundersøgelser er at kvalitet er vigtig. Desuden ville afskaffelse af tests ikke afskaffe behovet for at træffe sociale beslutninger. De ville bare blive truffet på et andet grundlag. Målet med ansvarlig bedømmelse er at fremskaffe et principielt grundlag for kvalificerede beslutninger. Når beslutninger træffes på grundlag af scorerne i sprogtests, er det normalt dem der har de bedste scorere der får gavn af det. Her er det at validering betyder noget. Det er vigtigt at vide hvad forskellen er på dem med en lavere score og dem med en højere score, således at det kan blive tjekket at beslutningerne er truffet på et forsvarligt grundlag.

Ændringer i omfang og filosofi: Fra testning til bedømmelse

Tidligere havde man den opfattelse at det først og fremmest drejede sig om tests i stor skala med særlig vægt på

statistisk analyse. Mens det stadigvæk er ét af de vigtigste aktivitets- og forskningsområder, er et andet lige så vigtigt område læringsrelateret bedømmelse. De bedømmelsesformater der her er relevante, omfatter forskellige former for selvevaluering og partnerevaluering (*peer evaluation*), portfoliobedømmelse, læringsdagbøger osv. Hvad angår testteori, har udviklingen betydet at sprogtestere har måttet revurdere deres antagelser om dette at bedømme, såvel som deres antagelser om de kvalitetskriterier der gælder i forskellige sammenhænge.

Som aktivitet er læringsrelateret bedømmelse mere kompleks og fleksibel end traditionel testning. Afhængig af behovene i de forskellige bedømmelsessituationer kan organiseringen variere med hensyn til:

- *Hvem bedømmer?* Læreren, andre studerende, sprogtilegnerne selv?
 - *Hvad bedømmes?* Det enkelte svar på et spørgsmål eller en opgave? Den generelle præstation i klasselokalet? Indstilling til sprogtilegnelse og sprogbrug? Erfaringer med fremmedsprog uden for klasseværelset?
 - *Hvem udvikler opgaven?* Læreren alene? Læreren og de studerende i fællesskab? De studerende i fællesskab? Én studerende til en anden? Den enkelte studerende til sig selv?
 - *Hvem træffer beslutning om kriterierne for præstationen?* Læreren alene? Læreren og de studerende i fællesskab? De studerende i fællesskab?
 - *Hvad sker der med bedømmelsesresultaterne?* Bliver de brugt som information til læreren og de studerende? Bliver de anvendt som støtte til sprogtilegnernes bevidsthed om sprogets natur og deres brug af sproget? Bliver de anvendt som grundlag for karaktergivning? Bliver de anvendt til vurdering af kursusplanens hensigtsmæssighed?
- Andre ting?

Den type af kvalitetskriterier der er drivkraften i de valg der træffes inden for læringsrelateret bedømmelse, omfatter ifølge McNamara (2001) at undervisningsprocessen er vedkommende, at læringen bliver faciliteret på en lang række forskellige måder, at kvaliteten af undervisningen forøges, at lærernes administrative byrder nedsættes. Stillet op på denne direkte måde er disse kriterier selvindlysende for lærerne, men de står helt tydeligt i modsætning til de bedømmelseskriterier som testere og uddannelsesadministratorer, såsom undervisningsministerier, har opstillet.

For testteoretikere er det vigtigste kvalitetskriterium det aktuelle brede validitetskoncept. Det betyder at de færdigheder der bliver bedømt, bør defineres, og at definitionen bør kunne forsvares intellektuelt; at der bør være bevis på reliabilitet; at der også bør være andre former for dokumentation og data der støtter testens validitet - såsom bevis på kvalitetskontrol fra testsudviklingsprocessen og analyser af de studerendes præstationer - for at vise at opgaven faktisk tester de færdigheder som det var hensigten at teste. Desuden bør konsekvenserne af at anvende testen overvejes før proceduren bliver anvendt og bagefter kontrolleret. For administratorer giver bedømmelse og evaluering nogle mål for ansvarlighed. De sætter ministerier i stand til at se hvor store fremskridt der gøres i læring og undervisning, og dermed hvad regeringer og stater får ud af deres investeringer i uddannelse. McNamara påpeger at de tre gruppers behov delvis er i konflikt med hinanden.

Tidligere har praktikere og teoretikere undgået konflikten ved ikke at betragte læringsrelateret bedømmelse som en del af 'test-' og 'administrations'-verdenen. De to grupper har - bevidst eller ubevidst - haft en tendens til at støtte hinanden. Når læringsrelateret bedømmelse nu er ved at blive en del af test-/bedømmelsesverdenen, er der brug for at sikre at foreningen sker i begge parter interesse. Dette indebærer både udfordringer og løfter. Udfordringerne på test-siden omfatter teoretikernes behov for at finde ud af hvilken forskning der er nødvendig for at forstå principperne i læringsrelateret bedømmelse og støtte dens berettigelse. McNamara (2001) henleder f.eks. opmærksomheden på at vi ved meget lidt om de refleksions- og analyseprocesser lærere og studerende gennemgår når de bedømmer en students arbejde, og at processerne kan adskille sig fra hinanden i forskellige bedømmelsessituationer. Udfordringerne på undervisningssiden består bl.a. i at der er behov for eksplicit refleksion over karakteren af de studerendes færdigheder på forskellige færdighedsniveauer, og over de træk ved præstationen (og muligvis bedømmelsessituationen) der fører til forskellige bedømmelsesresultater. Den nye tilgang inden for forskning, refleksion og analyse giver imidlertid løfte om at der vil blive produceret ny viden om de principper der ligger bag vores daglige handlen. Dette vil forhåbentlig forbedre vores arbejde og muligvis også mindske kløften mellem teori og praksis, en kløft som nogle gange synes at være ret bred.

Refleksioner over bedømmelsesprincipper og -praksis kan hjælpe lærerne med at sætte ord på deres implicitte forestillinger om og forståelse af sprogfærdighed. Det er noget lærerne kan gøre alene eller i små grupper. En del af arbejdet består i selv-analyse, men lærerne kan måske komme videre med deres refleksioner og analyser hvis de præsenterer deres tanker for en kollega eller to. Processen består af mindst to trin: beskrivelse og analyse/refleksion.

Som udgangspunkt for beskrivelsen bør læreren vælge én test- eller bedømmelsesprocedure som han eller hun har anvendt for nylig, og stille sig spørgsmål om den. Det er nyttigt at skrive svarene ned, men det er tilstrækkeligt at gøre det i skitseform; der er ikke brug for færdigpolerede formuleringer i denne fase. En grundlæggende liste over spørgsmål kunne se således ud:

-
- Hvad er 'sprogfærdighed' for disse sprogtilegnere i lyset af denne bedømmelsesprocedure?
- Hvordan ville du karakterisere en god præstation i forbindelse med løsningen af disse opgaver?
- Hvordan ville du karakterisere en dårlig præstation i forbindelse med løsningen af disse opgaver?
- Hvordan ser en gennemsnitspræstation ud i forbindelse med disse opgaver?
- Er det nyttigt i denne bedømmelse at overveje om de(n) bedømte sprogfærdighed(er) skal bestå af underfærdigheder? Hvis ja, hvilke?
- Hvilke andre færdigheder kræves der i bedømmelsen ud over sprogfærdigheder?

I analyse-/refleksionsfasen bearbejder man det tekstudkast der blev produceret i den foregående fase. Idéen er at fokusere på lærerens opfattelse af sprogfærdighed a) i relation til et særligt bedømmelsesinstrument, og b) generelt. Trinene er:

- Gennemgå de tekster du har produceret, og undersøg hvilke begreber du bruger til at beskrive sprogfærdighed med.
- Hvilke begreber bruger du til at skelne mellem færdigheder på et højere og et lavere niveau? Er disse begreber opgaverelaterede eller generelle (dvs. anvendelige på tværs af en række opgaver)?
- Hvilke tanker gør du dig om sprogfærdighed i denne test/bedømmelse i forhold til dine forestillinger om hvad sprogfærdighed er i mere almen forstand? (F.eks. i andre tests/bedømmelser? I undervisning generelt?)

Der er desuden behov for at analysere de filosofiske principper der ligger bag de kvalitetskriterier der bliver opstillet for forskellige slags bedømmelser. Der er allerede udført en del teoretisk arbejde inden for dette område, og det næste der er brug for, er måske analyser udført af lærere og testere af hvad dette betyder i praksis. Et godt udgangspunkt for dette arbejde er Moss' (1992) analyse af psykometriske og hermeneutiske filosofier om bedømmelse.

Theme: Assessment - what should we know?



What is happening in language testing?

Sari E. Luoma

European Project Officer.

**University of Cambridge Local Examination Syndicate
(UCLES), UK.**

The worlds of language teaching and language testing have changed a lot in the past 20-30 years. With the communicative revolution, language learners' experiences of what it is like to learn a language in institutional settings became explicitly connected with the use of language, and new procedures for assisting learners with the acquisition of skills required by language use have been introduced ever since. We have seen emphasis on dialogues and pairwork, situated contextual practice, strategy training, elaboration, and intercultural awareness raising, among others.

In a muted and modified way, and with some delay, these same features have also found their way into testing and assessment practices. The delay may partly be explained by the inherent conservativeness of testing as a norm-based activity. Partly, it is also due to the teacher-assessor's decisive role in initiating and planning assessment, combined with a conscious or subconscious fear of the strong demands of correctness and fairness that assessment involves. It is easier to follow old practices and believe that they are tried and trusted than to confront one's own fears and uncertainties about assessment. However, this situation is now changing.

Underneath the mainstream emphasis on measurement and standards, changes in assessment theory and practice have been going on in the past decades, and similarly to teaching, they have concerned all areas of approach, philosophy, and activity (an accessible overview of the developments is given e.g. by Birenbaum 1996). Though this may not have been in the theoretical headlines thirty years ago, the situation has changed in the past ten years, and it is becoming increasingly clear that there are several diverse currents in the core of language testing today. From the narrow confines of measurement, language testing has broadened into language assessment, a field with functional links with language education, several fields of applied linguistics, and educational assessment. Apart from topics of discussion, the change is also reflected in publication titles. The main journal in the area is still called Language Testing, but the volume on these topics in a recent encyclopedia was subtitled Language testing and assessment (Clapham and Corson (Eds.) 1997), and an ongoing series of current textbooks is called the Cambridge Language Assessment Series.

In this article, I will discuss some of the changes that have been going on in language testing and assessment in the "traditional" and the newer, broader sense. The paper has a practical focus on issues and implications that are relevant for language education. However, I will also include references to recent articles and reviews, which explain particular developments in more detail. As there is a whole article on the social and political dimensions

of language testing in this issue, I will not discuss the developments on this topic here regardless of its importance. Instead, I will focus on the other changes of philosophy and practice that recent developments in language testing have brought about.

Developments in technology: computer-adaptive testing

Computers and the Internet provide an important context of language use for today's language learners. This is especially so in English language education, though applications, particularly specialized language learning applications, have also been developed for other languages. The exciting opportunity for language assessment that computerisation offers is the development of computer-adaptive tests and assessment instruments.

Computer-adaptive assessment tools contain pre-programmed sets of tasks, which may be organized according to topics, skills tested, and/or specific content areas such as formulaic question-response sequences or singular-plural distinctions. Additionally, the tasks are tagged with some measurement information. This most often indicates how difficult the item has been found to be in pretesting and analysis, and possibly also how effective it is in telling apart those learners whose general ability level is higher than the difficulty of the item from those whose general ability is lower (i.e., item discrimination).

The idea in computer adaptive tests is that task delivery adapts to an individual learner's responses. The learner first gets an intermediate task (or item, as they are frequently called in testing terminology). If he or she gets it right, the program delivers a harder item. If the response is correct, the program chooses a harder item still. If the learner now gets this wrong, the program gives an easier item, beginning to home in on the learner's ability level. Another learner may get the same task at the beginning, but depending on his or her pattern of responses, may see quite a different set of items. Nevertheless, the program can use the measurement information attached to the items to give a final grade to both learners on the same scale.

The statistical basis for making computer adaptive tests is most often some form of Item Response Theory (IRT). IRT models are based on likelihood of correct response, in other words, probability theory. An accessible introduction to the use of item response theory in language testing is provided in Tim McNamara's 1996 book *Measuring second language performance*. There are also several introductory textbooks in statistics, which explain the theory that underlies the model; the search word 'item response theory' or 'IRT' will help locate some of them in a library near you. Moreover, the basics are also explained on some sites on the Web, such as ericae.net/scripts/cat/.

The advantages of computer adaptive testing include that tests can be shorter when the extra time for answering many items that are too easy or too difficult for the learner is avoided, and that results from tests given at different times are comparable. Once a pool of items has been piloted and analysed so that the item values are known, any sub-combination of them can be given to new test takers and their ability estimated. Test security is also improved when the content of any new test is unpredictable.

However, there are also drawbacks. The most important practical one is the amount of work needed to construct an item pool for an adaptive test. The numbers of learners that are needed for piloting items properly are in the hundreds, or for some statistical models in the thousands. In other words, construction of adaptive tests is not something that every teacher can easily do.

However, if potentially useful adaptive assessment systems already exist, teachers can try them out with their learners and evaluate their usefulness. The learners are likely to find it interesting to learn how different programs score them, but they might also be asked to reflect on other aspects of the assessment process such as enjoyment and experienced usefulness for supporting learning. One system that learners might be asked to try out is DIALANG, which is going to be available through www.dialang.org in an increasing number of languages in the course of year 2002. The first languages will be available as soon as the beta version is technically steady enough for public testing. More computer-mediated assessment alternatives can be found for example through Glenn Fulcher's Resources in language testing page at www.surrey.ac.uk/ELI/ltr.html under the "links" section. The pages also contain a wealth of other links to other language testing resources on the Internet, for example

videotaped explanations of basic concepts in language testing.

Besides Item Response Theory and its application in adaptive testing, there have also been several other technical and statistical developments in language testing in the past 30 years. For those who are interested in these aspects of language assessment, the developments are summarised succinctly in articles by Bachman and Eignor (1997) and Bachman (2000). Apart from information, the articles also provide further references on particular statistical techniques.

Theoretical developments: The rise of validation

In the area of language testing theory, the most significant developments in recent decades have happened in the area of validation. Validity has always been one of the central criteria of measurement quality in testing, but in the first half of the 20th century, it used to be considered an unproblematic concept. The question was whether the test was doing its job, and it was answered by correlating the scores with evaluations of the test takers' performance on the real life task that the test was expected to predict. The current concept of validity, however, is much broader than that. It is about the scientific and social defensibility of the test. This sounds grand, but in terms of test developers' work, it is actually quite practical. Validity is concerned with the meaning of the test scores. The central question is "What is this test measuring?" followed by "How do you know?"

Validation work begins from the test developers' definition of the skill that is to be assessed in the test. Operationally, the skill is implemented in the tasks, the assessment criteria, and the actual activities that result in giving the assessment. These need to be investigated to check that the skill that is supposedly assessed in the test is also actually addressed by the tasks. But current theory emphasizes that the developers need to go further than that, they need to be able to say what the results mean.

While this is quite a challenging demand, test developers are in an ideal position to describe the skills that they are assessing, because during development, they work with the test tasks and assessment criteria for a long time. In this work, a focus on the skill being measured, which is technically known as 'the construct', means that the developers pay conscious attention to how the skill that should be assessed can best be operationalized within the practical limitations of the assessment situation. "All" they need to do to make this work part of validation is write down their ideas and the reasons for their decisions. The work then continues by empirical investigations of whether the tasks and assessment criteria actually implement these intentions. In other words, validation involves several studies and many different methodologies, all aimed at showing what the test is testing.

However, since the focus of validation is not only on what the test is testing but also on the meaning of the scores, the social dimensions of assessment and scoring also become relevant. In a deep philosophical sense, this means recognising that assessment is a human activity, it is conducted by people on the basis of considered judgement. In a societal sense, it means recognizing that test scores often form bases for social decisions. When scores are used for giving grades or for deciding whether someone is eligible for study at a university, for example this has consequences for the test takers, and assessment developers are partly responsible for these consequences, according to current validity theory. They share the responsibility with the score users, for example university student selection boards, but they cannot avoid it completely.

Validation work is important because it is about quality and accountability. If the self-reflection aspect of validation is taken seriously, this helps developers improve their tests. However, in its current, complex form, validity theory is facing an interesting challenge. It is namely not clear who needs the complex validation reports and studies that current theory seems to call for. Who, other than another tester, is going to be able to read the reports and be able to evaluate whether they mean that the test is carefully developed and that it tests what it says it tests? The current trend in validity theory is therefore to make the concept clear and operationalizable. Good examples of this work are Carol Chapelle's (1998, 1999) articles, which provide good reading especially for those who work with more formal tests such as end-of-course examinations or graduation examinations. Other strands in the current discussion focus on the social dimension of test use, and suggest that perhaps the most legitimate use of language assessments is the assistance of further language learning (e.g. McNamara 2001).

When validation is considered a process, which is undertaken one step at a time, it leads to honest attempts at trying to do a good job at assessment in other words, accountable assessment. The argument for continuing with simple and more complex validation studies is that quality matters. Moreover, doing away with tests would not do away with the need to make social decisions. They would just be made on some other basis. The aim with accountable assessment is to provide a principled basis to inform decisions. When decisions are informed by scores from language tests, the ones with the better scores usually get the benefits. This is where validation counts. It is important to know what is different between those with lower scores and those with higher scores so that it can be checked that the decisions are made on a justifiable basis.

Changes of scope and philosophy: From testing to assessment

As mentioned at the beginning of the article, the biggest changes in language assessment in recent decades have been about the scope of the field of language testing. Earlier, it was considered to be mostly about large-scale tests with strong emphasis on statistical analysis. While that is currently one of the main areas of activity and research, another equally important strand is learning-related assessment. The formats of assessment that are relevant here include various kinds of self-evaluation and peer evaluation, portfolio assessment, learning diaries, etc. In terms of testing theory, the development has meant that language testers have had to rethink their assumptions about assessment as an activity, as well as the quality criteria that apply in different contexts.

As an activity, learning-related assessment is more complex and flexible than traditional testing. According to the needs of different learning-assessment situations, arrangements can vary about

- Who assesses? Teacher, fellow student, the learners themselves?
- What is assessed? An individual answer to a question or task? Classroom performance in general? Feelings about language learning and language use? Experiences of foreign language use outside the classroom?
- Who creates the tasks? The teacher alone? The teacher and the students together? The students in collaboration? One student for another? Each student for him/herself?
- Who decides the performance criteria? The teacher alone? The teacher and the students together? The students in collaboration? (In contrast to tasks, teachers still probably have more responsibility for guiding decisions about criteria. They have more background knowledge in the area, which makes this justified, but negotiation with students can sometimes be helpful.)
- What is done with the assessment results? Recorded for information by the teacher and the students? Used to support the learners' awareness of the nature of the language and their use of it? Used for giving grades? Used for judging the appropriacy of the course plan? Other?

The kinds of quality criteria that drive the choices in learning-related assessment, according to McNamara (2001) include meaningfulness in the instructional process, facilitation of learning in a multitude of different ways, enhanced quality of teaching, and minimization of administrative burden on teachers. Spelled out, these criteria are self-evident to teachers, but they contrast quite clearly with the criteria that testers or educational administrators such as ministries of education have posed for assessment.

For testing theorists, the most important quality criterion is the current, broad concept of validity. This means that the skills being assessed should be defined and that the definition should be intellectually defensible, that there should be evidence of reliability, that there should also be other records and data supporting the validity of the test such as evidence of quality control from the test development process and analyses of student performances to show that the intended skills are actually assessed in the tasks, and that the consequences of using the assessment should be considered before the procedure is used and monitored afterwards. For administrators, assessment and evaluation provide measures of accountability. They enable ministries to see how much progress is being made in teaching and learning, and thus, what governments and states are getting for their investment in education. McNamara points out that the needs the three groups are partially in conflict.

In the past, practitioners and theorists have avoided the conflict by not considering learning-related assessment

part of the "testing" or the "administrative" world. The latter two have consciously or subconsciously tended to support each other. When learning-related assessment is now beginning to be part of the testing/assessment world, some work is needed to make sure that the unification serves the interests of both parties. This involves both challenges and promises. The challenges on the testing side include that theorists need to consider what research is needed to understand the principles of learning-related assessment and support its legitimacy. McNamara (2001) suggests that, for instance, we know very little about the processes of reflection and analysis that teachers and students go through when they evaluate a piece of student work, and the processes may be different in different assessment situations. The challenges on the teaching side include that there is a need for explicit reflection on the nature of student skills at different levels of ability and on the features of performance (and possibly assessment situation) that lead to different assessment outcomes. The promise is that with the new kinds of research, reflection and analysis, new knowledge about the principles that underlie our daily action is uncovered. This will hopefully improve our work and possibly also reduce the gap between theory and practice, which seems quite considerable at times.

Reflection of assessment principles and practices helps teachers put their implicit ideas and understandings about language ability into words. This is something that teachers can do alone or in small groups. Part of the work is self-analysis, but teachers may be able to get further in their reflection and analysis if they present their thoughts to a colleague or two. Minimally, the process consists of two steps, description and analysis/reflection.

To begin the description, the teacher should choose one test or assessment procedure that he or she has used recently and ask questions about it. It is helpful to write down the answers, though draft form is quite enough; no polished formulations are needed at this stage. A basic list of questions might be:

- What is "language ability" for these learners, in the light of this assessment procedure?
- How would you characterise a good performance on these tasks?
- How would you characterise a weak performance on these tasks?
- What is an average performance like on these tasks?
- Is it useful for this assessment to consider the ability (abilities) assessed to consist of sub-skills? If so, what?
- What other skills do the assessment procedures require in addition to language skills?

The analysis/reflection stage works on the draft texts produced in the previous stage. The idea is to focus on the teacher's concept of language ability a) related to a particular assessment instrument, and b) in general. The steps are:

- Go through the texts that you produced and look for the concepts that you use to describe language ability.
- What concepts do you use to distinguish between more advanced and less advanced skills? Are these concepts task-related or general (applicable across a variety of tasks)?
- How are your thoughts about language ability in this test/assessment related to your ideas about the nature of language ability in more general terms? (for example, in other tests/assessments? In teaching in general?)

Moreover, there is a need to analyse the philosophical principles that drive the quality criteria posed for different kinds of assessments. Some theoretical work in this area has already been done, and perhaps what is needed next is practitioner analysis by teachers and testers on what this means in practical terms. A good starting point for this work is Moss's (1992) analysis of psychometric and hermeneutic assessment philosophies.

Psychometric principles are what drive traditional testing. The aim in psychometrically driven testing is to predict future performance as accurately as possible. Thus, the tester looks for evidence of behavioural consistencies to support the prediction. Consistency can only be found if one thing is tested many times, and within limited testing times, this means that each individual points that is tested is quite small. The whole, i.e. language ability, is then assessed by summing up the small components. If some components are variable or context-dependent, they might be left out of the test because they do not contribute to the discovery of consistencies. Similarly, features of performance that do not appear in every test taker's performance might be

ignored because they do not offer a reliable basis for prediction. Consistency and equality are valuable properties also in test administration. Premium is placed on comparability. The tasks, performance conditions, and assessment procedures must be the same for all test takers. This guarantees fairness of measurement in psychometric assessment.

In hermeneutic assessment, the overall aim is to understand the whole (that is, the nature of an individual learner's language ability) in the light of its parts. Case by case, the interpretation and evaluation is modified until all the features of the performance are accommodated. The tasks, rather than operationalizing a theoretical notion of skill, are representative of the individual, and different individuals might perform different tasks. Premium is placed on finding the most appropriate descriptions of individuals' skills, and therefore, rational argumentation among evaluators is valued. The evaluations of those who know the examinees well, such as teachers or long-time peers, are respected in particular, and the justification for the fairness of hermeneutic assessment comes from versatility of performance data and defensibility of the final, individualised assessment.

Clearly, the quality criteria associated with the above approaches to assessment are very different. While it may not be common to find extreme applications of either philosophy in the real world of language education today, it might be useful to keep the distinctions in quality criteria in mind, particularly in order to avoid applying the criteria of one mode to examples of the other, at least without questioning why one should do so. Thus, when we use large-scale tests that aim to predict future performance, it would be prudent to make sure that these instruments are truly standardized: reliable, comparable, consistent, and impartial. This can be shown through quantitative (and some qualitative) analyses. In the light of current validity theory, the consequences of test use should also be considered. In individualised, hermeneutically motivated assessment, the most important quality criteria are transparency and justification. The examinees should know what will be evaluated, and they should be aware of the range of possible tasks that they can choose from. From the evaluation process, there should be evidence that in each case, the evaluators put effort into accommodating the assessment interpretation to the individual, and that the final assessment is rationally defensible. Because of the likely context where hermeneutically motivated assessments would be used as well as the type of assessment information to be produced, statistical analyses are hardly relevant in quality assurance. Rather, teachers might consider presenting a series of assessment cases to a colleague, or in some contexts perhaps to the examinee, to see whether their assessments are defensible.

The above is the current stage of development in the theory of philosophies of assessment. The ideas now need application into the reality of educational assessment to see whether or to what extent they are helpful, and what more is needed to understand the values that guide the newly redefined world of language assessment.

Conclusion

There are a large number of advances and activities going on in language assessment today. The good news for the teaching world is that quite a large part of it is beginning to be relevant for teachers, and serious attempts are being made to bring the worlds of teaching and testing closer together. On these arenas, contributions from teachers would be very welcome.

References

- Bachman, L. F.** 2000. Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17, 1-42.
- Bachman, L. F.** and **Eignor, D. R.** 1997. Recent advances in quantitative test analysis. In **Clapham, C.** and **Corson, D.** (Eds.), 227-242.
- Birenbaum, Menucha** 1996. Assessment 2000: Towards a pluralistic approach to assessment. In **Birenbaum** and **Dochy** (Eds.), *Alternatives in assessment of achievement, learning processes and prior knowledge*. Boston: Kluwer Academic Press, 3-29.
- Clapham, C.** and **Corson, D.** (Eds.) 1997. *Encyclopedia of language and education*. Volume 7: Language testing and assessment. Dordrecht: Kluwer Academic Press.

- Chapelle, C. A.** 1999. Validity in language assessment. *Annual Review of Applied Linguistics* 19, 254-272.
- Chapelle, C. A.** 1998. Construct definition and validity inquiry in SLA research. In L. F. Bachman and A. D. Cohen (eds.) *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press, 32-70.
- McNamara, Tim** 2001. Language assessment as social practice: challenges for research. *Language Testing* 18 (4), 333-349.
- Moss, Pamela** 1992. Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research* 62(3), 229-258.

Tema: Evaluering - hvad skal vi kunne?



Den Europæiske Sprogportfolio

Marianne Hirtzel

European Project Officer.

University of Cambridge Local Examination Syndicate (UCLES), UK.

Den Europæiske Sprogportfolio (*European Language Portfolio (ELP)*) blev officielt lanceret af Europarådet i Det Europæiske Sprogår 2001 efter flere år med pilotprojekter og udviklingen af mange forskellige versioner rundt om i Europa. Mange versioner er blevet fremstillet; de fleste af dem målrettet mod bestemte sektorer og/eller aldersgrupper, men de har visse træk til fælles.

Portfolioprojektet er meget tæt forbundet med *Common European Framework (CEF)*, og de er begge udsprunget af arbejdet med Europarådets projekt Moderne Sprog i 1990'erne. Portfoliomodellen anvender *Common European Frameworks* system med seks niveauer (A1, A2, B1, B2, C1, C2), og specielt *Passport*-afsnittet indeholder *Common European Frameworks* generelle skema med niveau- og færdighedsdeskriptorer. Dette er en meget vigtigt pointe og en styrke ved Portfolien da *Common European Framework* er ved at blive standard for kurser, materialer og prøver over hele Europa inden for skole- og erhvervsuddannelser samt højere uddannelser. Portfolien kan derfor indpasses i uddannelsessystemer og mere uafhængig læring og blive mere værdifuld for sin ejer fordi den henviser til et anerkendt fælleseuropæisk system.

Målene bag Portfolien

Portfolioprojektet er udviklet med henblik på at støtte den europæiske integration og de europæiske borgeres mobilitet ved at udstyre brugerne med et redskab der gør det muligt for dem at registrere deres formelle læring (kurser og eksaminer fra skolen), men det giver dem også et middel til at registrere

andre måder sprog kan læres og udvikles på. På denne måde får interkulturelle erfaringer og mindre formaliserede måder at udvikle sit sprog på tillagt betydning idet de giver et samlet billede af den enkeltes evner. Portfolien (og især *Passport*-delen) støtter derfor europæiske borgeres mobilitet - især for dem som søger arbejde eller uddannelse i et andet land - ved at indeholde en fuldstændig registrering af en persons færdigheder inden for rammerne af en model som er let at genkende fra det ene europæiske land til det andet.

Et stærkt element i Portfolio-projektet er dets støtte til livslang læring. Dem der har en Europæisk Sprogportfolio

bliver opfordret til at vende tilbage til den med regelmæssige mellemrum for at opdatere registreringen af deres færdigheder. Nogle starter måske i skolen med en Portfolio der specielt retter sig mod elever i folkeskolen, og bevæger sig derefter videre til udgaver der passer bedre til voksne, mens andre måske begynder med en udgave der passer til voksne sprogtilegnere. Skønt sprogbrugere sandsynligvis får deres Portfolier i forbindelse med mere formaliserede læringssituationer som f.eks. i skolen og på sprogkurser, er det et af målene at opmuntre dem til selvstændig læring. Portfolien er altid den enkeltes ejendom, og han eller hun bør anvende den som en hjælp til at registrere sine fremskridt og sætte sig nye læringsmål.

Portfolien motiverer europæiske borgere til at udvikle delkompetencer på flere sprog eftersom ethvert færdighedsniveau kan registreres. Borgerne bliver derved støttet i deres forsøg på i højere grad at blive flersprogede og flerkulturelle. Et af Europarådets mål er at opmuntre folk til at opnå delkompetencer på flere sprog i stedet for kun at beherske ét sprog og på den måde fremme en positiv holdning til kulturel diversitet og støtte en større forståelse af andre.

Strukturen i Den Europæiske Sprogportfolio

I 2000 nedsatte Europarådet en *Validation Committee* (valideringskomité) til at tage sig af ansøgninger om officiel godkendelse af forskellige udgaver af portfolier. En af betingelserne for dette er kravet om at officielt godkendte udgaver skal omfatte tre adskilte dele i hele portfolien. De tre dele er:

- *Passport* (pas): Denne del har en rapporteringsfunktion, og vil formentlig være den del der mest vil blive anvendt til eksterne formål;
- *Language Biography* (sprogbiografi): Bruges til at registrere erfaringer med læring af sprog;
- *Dossier*: Indeholder prøver på materiale der støtter de to andre dele.

Derudover er der nogle bestemte sider og andre elementer som bør være indeholdt i alle versioner. De bliver mere detaljeret behandlet nedenfor.

Officielt validerede versioner

Indtil nu er 23 versioner blevet officielt godkendt af Europarådet. Mange af disse er blevet lavet til bestemte lande eller sektorer såsom den nederlandske version til erhvervsuddannelser (18. 2001), den ungarske version til grundskolen (15.2001), og den version der er lavet af *Centre International d'Etudes Pédagogiques* i Frankrig (2.2000). Den Portfolio der er udviklet til voksne sprogtilegnere (dvs. aldersgruppen over 16 år) af EAQUALS og ALTE (version 6.2000), er en af de få der er lavet til denne sektor og med det mål af blive anvendt internationalt. Den bliver nærmere behandlet nedenfor med henblik på ræsonnementet bag indholdet og de anvendelsesmuligheder den kan have. [note1](#)

En Portfolio til voksne sprogtilegnere: EAQUALS-ALTE versionen

EAQUALS (*European Association for Quality Language Services*) begyndte sit arbejde med en Portfolio i 1999, og aftalte efterfølgende at indgå i et samarbejde med ALTE (*Association of Language Testers in Europe*) da man mente at de to foreningers mål supplerede hinanden fint. EAQUALS havde allerede udført nogle pilotstudier med den første Portfolio de udviklede, og dette viste sig at give værdifuld feedback til udviklingen af den nuværende version.

Den første offentliggjorte version af EAQUALS-ALTE Portfolien er netop nu ved at blive fremstillet og vil blive lanceret i begyndelsen af marts på fire sprog: Engelsk, fransk, tysk og italiensk (se detaljerne i slutningen af denne artikel). Det er nyttigt at se lidt nærmere på Portfolien for at sætte fokus på nogle af målene og principperne bag den.

Passport

Da det er mest sandsynligt at det er denne del af Portfolien der vil blive brugt til eksterne formål (som f.eks. når man skal søge job eller uddannelsesplads i udlandet), har den så vidt muligt fået en form hvor der ikke skal bruges ord. På de fleste af siderne skal man enten sætte kryds eller skravere nogle felter, og formatet er A5, således at dobbeltsider let kan fotokopieres på A4-ark.

De to første sider giver en introduktion til Europarådet, Portfolio-projektet og plads til at skrive sit navn. Det er også muligt, men ikke obligatorisk, at vedlægge et foto af ejeren.

Herefter følger et afsnit med overskriften *Profile of Language Skills* (Sprogfærdighedsprofil), hvori brugeren angiver sit/sine modersmål og andre sprog som han eller hun taler. Der er her plads til at registrere færdighed i op til 6 sprog, herunder modersmålene, som har samme vægt som de andre sprog i Portfolien. For at udfylde dette afsnit anvender Portfoliens indehaver selv-evalueringsskemaet fra *Common European Framework* (tabel 2 fra kapitel 3 i *Common European Framework of Reference for Languages: Learning, teaching, assessment*), som findes midt i Passport-brochuren. Skemaet giver generelle redegørelser for færdighed på seks niveauer i *Common European Framework* (A1 til C2) inden for fem færdighedsområder: Lytning, Læsning, Mundtlig Interaktion, Tale og Skrivning. Idéen er her at brugeren skal identificere sit generelle niveau inden for hver færdighed og hvert sprog ved at skravere de relevante felter i tabellerne for *Profile of Language Skills*. Denne del af *Passport* er ikke beregnet til daglig brug; brugeren vil sandsynligvis komme tilbage til denne tabel en gang hver måned eller så, afhængig af hvor intensivt og sammenhængende et sprog bliver lært og/eller brugt. I *Language Biography* (sprogbiografien) er der også plads til en nærmere og mere hyppig analyse af niveauerne.

Det næste hovedafsnit i *Passport* dækker *Language learning and intercultural experiences* (sprogindlæring og interkulturelle erfaringer), for det første i en tabel der ser på anvendelsen af sprog i lande eller områder hvor sproget ikke tales som modersmål, og derefter i områder hvor det bliver talt. Igen er det hensigten med disse to tabeller at man skal udfylde dem ved at afkrydse eller skravere et felt, og disse sider er udviklet for at give en generel sammenfatning som der bliver nærmere redegjort for i *Language Biography*. De sider der dækker anvendelse af sproget i et land eller et område hvor sproget ikke tales, angiver tidsperioder i år (op til 1/3/5 år og mere end 5 år), og inkluderer afsnit som skole, erhvervs- og voksenuddannelse, anvendelse på arbejdspladser og gennem direkte kontakt med andre der taler sproget. Disse sider kunne for eksempel udfyldes af én der bor i Tyskland og som regelmæssigt anvender fransk på sit arbejde. Der er også bagest plads til meget korte supplerende kommentarer, som kan bruges hvis man ikke synes at nogle af de andre sammenhænge er dækkende.

De sider der dækker ophold i et område hvor sproget tales som modersmål, er udformet på samme måde, men tidsfelterne her angiver perioder i måneder i stedet for år eftersom mange af sådanne ophold vil være af kortere varighed. Eksempler på sådanne ophold er: Deltagelse i et sprogkursus og anvendelse af sproget i forbindelse med studieophold eller kurser, og igen er der lidt plads til at indføje særlige detaljer.

Derefter giver *Passport* plads til at registrere *Certificates* og *Diplomas* (beviser og diplomer) man har fået, og til at angive karakterer eller uddannelser med niveauerne i *Common European Framework*. Dette afsnit kan indeholde prøver fra skolen eller andre steder fra.

Indtil dette punkt er de sider der er omtalt, Europarådets standardsider, og disse skal være med i enhver Portfolio der er blevet officielt godkendt. De sider og afsnit der følger, er blevet specielt udviklet til EAQUALS-ALTE versionen, herunder ALTE's *Framework of Language Examinations*. I øjeblikket omfatter den prøver der er lavet på 16 af de sprog der er repræsenteret i ALTE, og yderligere 8 sprog følger snart efter. Prøverne er lagt på 6 niveauer, således at man kan se hvilket niveau prøven er på, sammenlignet med mange andre der er lavet til andre sprog. Gennem arbejdet med ALTE's *Can Do projekt* er det empirisk blevet demonstreret hvordan det er muligt at sammenknytte prøver fra ALTE *Framework* og CEF, og prøverne har CEF's overskrifter. Ved at benytte de næste sider i *Passport* kan brugerne registrere alle de prøver de har taget på EAQUALS' medlemsskoler, og de prøver de har taget inden for ALTE *Framework*.

De sidste sider i *Passport* giver retningslinjer for hvordan man skal udfylde dokumentet sammen med nogle enkelte informationer om ALTE og EAQUALS og foreningernes Internetadresser.

Language Biography (sprogbiografi)

Som det blev nævnt ovenfor, er målet med *Language Biography*-delen af Portfolio at gøre det muligt for brugerne at give en mere detaljeret redegørelse for deres sprogtilegnelse og interkulturelle erfaringer. De første tre afsnit giver generelle retningslinjer, men er ikke alt for præskriptive hvad angår indhold og mængden af oplysninger. Der anvendes et løbsk system således at sprogbiografidelen er nem at opdatere og det kan gøres regelmæssigt, og dermed giver mulighed for stor fleksibilitet i anvendelsen. De tre afsnit er:

1. *My language learning aims* (mine sprogindlæringsmål): Denne del opmuntrer brugerne til at skrive om hvorfor de er ved at lære et bestemt sprog og at overveje hvilke færdigheder der er de vigtigste for dem. De kan bruge den til et særligt formål, eller det kan være vigtigt for dem at nå et bestemt niveau på grund af deres arbejde eller af andre grunde.
2. *My language learning history* (min sprogtilegnelseshistorie): Her bliver brugeren bedt om at give detaljer om sprogindlærings erfaringer i kronologisk rækkefølge, så som gennemførte kurser, anvendelse af sprog på arbejde og i forbindelse med ferier. Dette omfatter sprog som sprogtilegneren har mødt som barn, f.eks. inden for familien eller under ophold i udlandet.
3. *My most significant linguistic and intercultural experiences* (mine mest betydningsfulde sproglige og interkulturelle erfaringer): Fokus ligger her på hvordan disse erfaringer kan have ændret brugerens forhold til det pågældende sprog og den pågældende kultur og deres følelser i forbindelse med læringsoplevelser. F.eks. kan de være blevet mere motiveret til at lære efter at have besøgt et land hvor sproget tales som modersmål.

Som det ses gør afsnit 2 og 3 det muligt for brugeren i langt højere grad at uddybe de oplysninger som der kun blev sat kryds ved, eller som blev skraveret i felterne i *Passport* under *Language learning and intercultural experience*.

Det sidste afsnit i *Language Biography* giver mulighed for at uddybe de oplysninger der blev givet i *Profile og language skills*. Dette afsnit gav de generelle deskriptorer for niveauerne og færdighederne fra CEF. Her, i afsnittet *My current language learning priorities* (Mine aktuelle sprogtilægnelsesprioriteter), er der mere fyldige deskriptorer, som blev udviklet til CEF og ELP af Günther Schneider og Brian North til et *Swiss National Science Foundation* projekt. For hvert niveau og under færdighedsoverskriften er der mere detaljerede deskriptorer (omkring 40 deskriptorer pr. niveau) som gør det muligt for brugeren nøjere at præcisere sit niveau. Disse deskriptorer kan enten bruges til at vurdere ens aktuelle niveau eller til at udstikke læringsmål. Sprogtilegneren bliver bedt om at angive sit generelle niveau i *Passport*-skemaet og derefter arbejde sig igennem den relevante tjekliste for dette niveau idet han eller hun ser på hvilke deskriptorer der beskriver de ting som han/hun kan gøre nu, eller gerne ville kunne gøre i fremtiden. Efter at have gjort dette, kan det generelle niveau (som godt kan være forskelligt for hver færdighed) registreres i *Passport*.

Arbejdet med *Language Biography* kan gøres af sprogtilegneren selv eller sammen med en lærer, eventuelt i forbindelse med planlægningen af et kursus.

Dossier

I denne del af Portfolien samler brugeren udvalgte eksempler der støtter de oplysninger der er givet andetsteds. Der kan være tale om eksamensbeviser, rapporter eller dele af skriftlige arbejder såvel som kassettebånd eller videoer der giver prøver på mundtlige færdigheder eller præsentationer. Der er et skema hvori det udvalgte arbejde kan registreres, og hvori det kan angives om arbejdet er udført individuelt eller sammen med andre, og om det indeholder rettelser f.eks. fra en lærer.

Konklusion

Portfolien har altså både en rapporteringsfunktion og en pædagogisk funktion, ligesom den støtter Europarådets mål. Rapporteringsfunktionen er tydeligst i *Passport* der som nævnt sandsynligvis er den del af *Passport* der

hyppigst vil finde 'ekstern' anvendelse. Da *Language Biography* og *Dossier* også har en rapporteringsfunktion, giver de sprogtilegneren mulighed for at gå i gang med og fortsætte den udfordring det er at lære sprog som en selvstændig aktivitet som han eller hun selv er ansvarlig for.

Sprogtilegneren kan anvende elementerne i Portfolien som et middel til at holde styr på sine hidtidige fremskridt og sætte mål for sin fremtidige udvikling. Disse kunne omfatte en forbedring af de nuværende sprogfærdigheder eller tilegnelsen af nye sprog. Udover at blive anvendt til at sætte mål med, opmuntrer Portfolien sprogtilegneren til at tænke over hvordan han eller hun lærer sprog og til at se sprog i deres kulturelle kontekster. Man regner med at Portfolien vil have en indvirkning på læreplansudvikling eftersom sprogtilegnerne bliver mere bevidste om deres behov. Den skulle også kunne tilskynde sprogtilegnerne til at søge kvalitet i sprogkurser og eksaminer.

Noter & Henvisninger

1 Den danske version er under udarbejdelse af Ministeriet for Flygtninge, Indvandrere og Integration.

ALTE Secretariat: 1 Hills Road · Cambridge CB1 2EU · Tlf.: +44 1223 553925 · e-mail: alte@ucles.org.uk · www.alte.org

EAQUALS Secretariat: The British School of Trieste · Via Torrebianca 18 · 34132 Trieste · Italien · e-mail: Peter.Brown@eaquals.org · www.eaquals.org

Europarådet: www.culture.coe.portfolio

Litteratur

Swiss National Science Foundation projekt: Schneider, G. & North, B. (2000), Fremdsprachen können - was heisst das? Chur/Zürich, Rüegger. ©Editions scolaires du canton de Berne/Berner Lehrmittel- und Medienverlag 2001.

Common European Framework of Reference for languages: learning, teaching, assessment.

Engelsk udgave v/ Cambridge University Press. Fransk udgave v/ Didier. Tysk udgave v/Langenscheidt.

Oversat af Michael Svendsen Pedersen

Theme: Assessment - what should we know?



The European Language Portfolio: support and development for language learners

Marianne Hirtzel

European Project Officer.

**University of Cambridge Local Examination Syndicate
(UCLES), UK.**

The European Language Portfolio was officially launched by the Council of Europe during the European Year of Languages in 2001, following several years of pilot projects and the development of many versions around Europe. Many versions have been produced; most are targeted at particular sectors and/or age groups, but they have certain characteristics in common.

The Portfolio project is very closely linked to the Common European Framework and both have grown out of the work on the Council of Europe Modern Languages project during the 1990s. The Portfolio model uses the six level system from the Common European Framework (A1, A2, B1, B2, C1, C2) and in particular the Passport section includes the overall grid of descriptors by level and skill from the Common European Framework. This is a very important point and a strength of the Portfolio, as the Common European Framework is increasingly becoming the standard to which courses, materials and tests are linked all over Europe in school, vocational and higher education. For this reason the Portfolio can fit into education systems or more independent learning and provide added value for its owner by referring to a recognised pan-European system.

The underlying aims of the Portfolio

The Portfolio project is designed to support European integration and the mobility of European citizens by providing users with a tool which not only allows them to record formal learning (classes and examinations taken in school) but also provides a means of recording other ways in which languages may be learnt or developed. In this way intercultural experiences and less formal ways of language development are given value, providing an overall picture of an individual's abilities. The Portfolio, (and in particular the Passport element), therefore supports the mobility of European citizens by providing a full record of an individual's skills, within a transparent model which is easily recognisable from country to country in Europe, especially for those applying to work or study in another country.

A strong element of the Portfolio project is the provision of support for lifelong learning. Holders of European

Language Portfolios are encouraged to return to them at regular intervals in order to update the record of their skills. Some may start in school with a Portfolio specifically aimed at primary or secondary school students, and then move on to versions suitable for adult learners, others may start off with a version suitable for adult learners.

Although language users are likely to receive their Portfolios in more formal learning situations such as in school or on a language course, one of the aims is to encourage them in independent learning. The Portfolio is always the property of its owner, who should use it to help record his or her progress and identify new learning goals.

The Portfolio motivates European citizens to develop partial competence in several languages, as any level of ability may be recorded. Citizens are thus supported in becoming more plurilingual and pluricultural. One of the aims of the Council of Europe is to encourage people to achieve partial competence in several languages, rather than mastery of just one, in order to promote appreciation of cultural diversity and to support greater understanding of others.

The structure of the European Language Portfolio

In 2000 the Council of Europe set up a Validation Committee to deal with applications for official recognition of Portfolio versions. Among the conditions for this is the requirement that officially recognised versions must include three specific parts within the whole Portfolio. These three parts are:

- Passport: this has a reporting function and is likely to be the part most used for external purposes
- Language Biography: used to record language learning experiences
- Dossier: contains samples of materials supporting the other two parts

In addition to this there are certain pages and other elements which should be included in every version; these are examined in more detail in section 5 below.

Officially validated versions

To date 23 versions have been officially validated by the Council of Europe. Many of these have been produced for particular countries or sectors, such as the Netherlands version for upper secondary vocational education (18.2001), the Hungarian version for secondary education (15.2001) and the version produced by the Centre International d'Etudes Pédagogiques in France for young learners (2.2000). The Portfolio for adult language learners (that is the 16+ age group) produced by EAQUALS and ALTE (version 6.2000) is one of the few produced for this sector and with the aim of being used internationally. It is examined below in detail in order to look at the reasoning behind the contents and the potential uses it may have.

A Portfolio for adult language learners: the EAQUALS-ALTE version

EAQUALS (the European Association for Quality Language Services) started working on a Portfolio in 1999 and subsequently agreed on a collaboration with ALTE (the Association of Language Testers in Europe), as it was felt that the aims of the two associations complement each other well. EAQUALS had already carried out some pilot studies using the first Portfolio they developed and this provided valuable feedback for the development of the current version.

The first published edition of the EAQUALS-ALTE Portfolio is being produced at present and will be launched at the beginning of March in four languages: English, French, German and Italian (details at the end of this article). It is useful to look at the sections of the Portfolio in some detail in order to highlight some of the aims and principles behind it.

Passport

As this is the part of the Portfolio which is most likely to be used for external purposes (such as when applying for a job or for study abroad), as far as possible it is designed to be completed without using words. Most pages involve either ticking or shading in boxes and the format is A5 so that a double page can easily be photocopied onto A4 sheets.

The first two pages give an introduction to the Council of Europe, the Portfolio project and space to record the owner's name,. It is also possible, though not obligatory, to include a photo of the owner.

This is followed by a section entitled 'Profile of Language Skills', in which the user identifies his or her mother tongues(s) and other languages spoken. There is space here to record ability in up to 6 languages, including the mother tongue(s), which are given equal weight within the Portfolio. In order to complete this section the holder of the Passport makes use of the main self-assessment grid from the Common European Framework (table 2 from Chapter 3 in Common European Framework of Reference for languages: Learning, teaching, assessment), which appears in the centre of the Passport booklet. This grid gives overall statements of ability for the six levels of the Common European Framework (A1 to C2) in five skill areas: Listening, Reading, Spoken Interaction, Spoken Production and Writing. The idea here is for the user to identify his or her overall level for each skill and language, shading in the appropriate boxes on the 'Profile of Language Skills' tables. This part of the Passport is not designed for frequent use; the user would probably come back to this table once every month or so, depending on how intensively and continuously a language is being learnt and/or used. As will be seen, there is scope to analyse level in greater detail and more frequently in the Language Biography.

The next main section of the Passport covers 'Language learning and intercultural experiences', firstly on a table looking at use of languages(s) in a country or region where the language is not spoken, and then use in a region where it is. Again the aim with these two tables is for completion to be by means of ticking or shading boxes and these pages are designed to give an overall summary, which is expanded on in the Language Biography. The pages covering use of the language in a country or region where the language is not spoken provide for time periods in years (up to 1/3/5 years and over 5 years) and include sections such as school, vocational and adult education, use in the workplace and regular contact with other speakers of the language. These pages could for example be completed by someone living in Germany and using French regularly for their job. There is also a small space at the bottom for very brief further comment, which could be used if it is felt that none of the other contexts are suitable.

The pages covering stays in a region where the language is spoken are in a similar format, but the time boxes provide for periods in months rather than years, since many such stays would be of a shorter duration. Examples of the types of stays are: attending a language course and using the language for study or training, and again there is a small space for inclusion of particular details.

The Passport then provides space for the recording of Certificates and Diplomas obtained and identifying the marks or grades received with the levels in the Common European Framework. This section might include both school and other examinations.

Up to this point the pages mentioned are standard Council of Europe pages which should be included in every Portfolio which has been officially validated. The pages and other sections which follow were developed specifically for the EAQUALS-ALTE edition.

The ALTE Framework of Language Examinations appears on the next pages. At present this includes examinations produced in 16 of the languages represented in ALTE, with 8 further languages soon to be added. The examinations have been placed on 6 levels, so that it is possible to see which level an examination is at compared with many others produced for other languages. Through the work on the ALTE Can Do project, the linking between the examinations on the ALTE Framework and the Common European Framework has been demonstrated empirically and the examinations appear under the headings of the CEF. Using the next pages to appear in the Passport, users can record any courses they have taken at EAQUALS member schools and examinations they have taken which appear on the ALTE Framework.

The final pages of the Passport give guidelines for completing the document, together with brief information

about ALTE and EAQUALS, together with the associations' web site addresses.

Language Biography

As mentioned above, the aim of the Language Biography part of the Portfolio is to allow users to give greater detail about their language learning and intercultural experiences. The first three sections give general guidelines but are not too prescriptive as to the content or length of information given. The format is loose-leaf so that updating is easy and can be done regularly, allowing great flexibility of use. The three sections are:

1. My language learning aims: this section encourages users to write about why they are learning a particular language and to consider what skills are most important to them. They might be wanting to use it for a specific purpose or it may be important to reach a certain level for work or other reasons.
2. My language learning history: here the user is asked to give details in chronological order of language learning experiences, such as courses taken, using languages at work and on holidays. This includes languages the learner was exposed to as a child, for example within the family or living abroad.
3. My most significant linguistic and intercultural experiences: the focus here is on how these experiences may have changed the way the user relates to the language and culture concerned and their feelings about the learning experience. For example, they may feel more motivated to learn after visiting the country where the language is spoken.

It can be seen that sections ii. and iii. allow the user to develop in much greater detail the information which was simply ticked or shaded in boxes in the Passport under 'Language learning and intercultural experiences'.

The final section in the Language Biography gives the opportunity to develop in greater detail the information given in the 'Profile of Language Skills' section of the Passport. That section provided overall descriptors by CEF level and skill. Here, under the section 'My current language learning priorities' there is a much fuller set of descriptors. These were developed for the CEF and the ELP by Günther Schneider and Brian North in a Swiss National Science Foundation project. For each level and under the skill headings there are more detailed descriptors (about 40 descriptors per level), allowing the user to pinpoint more accurately his or her level. These may either be used to assess one's current level or to set learning objectives. The learner is asked to identify his or her overall level in the Passport grid, then work through the relevant checklist for that level, looking at which of the descriptors are things s/he can do now or would like to be able to do in the future. Having done this, the overall level (which may be different for each skill) can be recorded in the Passport.

Work on the Language Biography might be done by the learner independently or together with a teacher, possibly when planning a course.

Dossier

In this part of the Portfolio the user collects samples which support the information provided elsewhere. These may be certificates, reports or pieces of written work as well as cassettes or videos of spoken ability or presentations. There is a grid on which to record the work kept and to record whether it was work done individually or jointly with others, and whether it includes corrections from for example a teacher.

Conclusion

From the example given above it can be seen that the Portfolio has both a reporting and a pedagogical function, as well as supporting the aims of the Council of Europe. The reporting function is most evident in the Passport which, as has been mentioned, is likely to be the element of the Portfolio most used 'externally'. Whilst they also have a reporting role, the Language Biography and Dossier give the learner the opportunity to take up and continue the challenge of learning languages as an autonomous activity for which he or she is responsible.

The learner can use the elements of the Portfolio as a means of tracking progress to date and also as a means of setting targets for future development. These might involve improving the level of existing language knowledge

or learning new languages. As well as being used to set goals, the Portfolio encourages the learner to think about how they learn languages and to consider languages in their cultural contexts.

It is anticipated that the Portfolio will have an impact on curriculum design, due to learners becoming more aware of their needs. It should also encourage learners to seek quality in language courses and examinations.

For all these reasons the Portfolio helps to support the aims of the Council of Europe in developing better citizens on Europe, who are more aware of their roles in society.

References, further information and useful web site addresses

Council of Europe www.culture.coe.portfolio

References, further information and useful web site addresses

ALTE Secretariat: 1 Hills Road · Cambridge CB1 2EU · Tlf.: +44 1223 553925 · e-mail: alte@ucles.org.uk · www.alte.org

EAQUALS Secretariat: The British School of Trieste · Via Torrebianca 18 · 34132 Trieste · Italien · e-mail: Peter.Brown@eaquals.org · www.eaquals.org

Europarådet: www.culture.coe.portfolio

Litteratur

Swiss National Science Foundation projekt: Schneider, G. & North, B. (2000), Fremdsprachen können - was heisst das? Chur/Zürich, Rüegger. ©Editions scolaires du canton de Berne/Berner Lehrmittel- und Medienverlag 2001.

Common European Framework of Reference for languages: learning, teaching, assessment.

English version (publisher Cambridge University Press ISBN 0521 005310)

French version (publisher Didier ISBN 2278 050753)

German version (Publisher Langenscheidt ISBN 3468 494696)

Tema: Evaluering - hvad skal vi kunne?



Folkeskolens afsluttende prøver i engelsk, tysk og fransk

Aase Brick-Hansen

Læreruddannet, cand. pæd. i engelsk.

Lærer og skolekonsulent for fremmedsprog i Brøndby Kommune.

Medlem af Undervisningsministeriets opgaveudvalg for engelsk.

De afgangsprøver, vi kender i dag, bygger på folkeskoleloven af 1993, hvor der gives mulighed for at gennemføre afsluttende prøver i engelsk, tysk og fransk ved afslutningen af 9. og 10. klasse. Ved slutningen af 9. klasse kan eleverne vælge i samråd med forældre og skolen at indstille sig til folkeskolens afgangsprøve (FSA), ved slutningen af 10. klasse at indstille sig til 10. klasseprøven eller FSA. En forudsætning er, at eleverne har fulgt undervisningen i det pågældende skoleår.

Lidt historie

Folkeskolens afgangsprøver i engelsk og tysk, som vi kender dem i dag, har ikke ændret sig meget gennem årene; de er baseret på lignende prøver helt tilbage til 1978. Med 93-loven blev der mulighed for at aflægge prøve i fransk på linie med prøvekravene i tysk. I forbindelse med udviklingsarbejder i fransk havde denne mulighed været givet tidligere, men først fra 1991 med en skriftlig prøve på FSU niveau.

93-loven afskaffede muligheden for at niveaudele eleverne i grundkursus og udvidet kursus, hvilket resulterede i et ændret prøveniveau. Tidligere byggede FSA prøverne på læseplanerne for grundkursusniveau, hvilket betød, at mange elever fik tårnhøje karakterer, men med den nye lov blev dette ændret. Prøveniveauet bygger på læseplanerne for 9. klasse, dvs. det er en målrelateret prøve.

Indtil 1989 bestod de skriftlige prøver af en opgave i fri skriftlig fremstilling og en cloze-prøve, men efter dette tidspunkt består den kun af en fri skriftlig fremstilling. Eleverne får tre forskellige emner at vælge imellem, f.eks. et fabulerende emne, hvor de ud fra oplæggets tekst og billede skal sætte sig ind i situationen og 'digte' videre på den. Der kan være et emne, hvor eleverne skal bruge deres viden om omverdenen og det fremmedsprogede land, og endelig et oplæg, hvor eleverne kan inddrage deres personlige erfaringer. Med tiden har opgavesættene udviklet sig til farverige produktioner. De skriftlige prøver stilles centralt af ministeriet, der har nedsat et opgaveudvalg for hvert fag til at udarbejde opgaverne. En elevbesvarelse rettes af en beskikket

censor og elevens egen lærer.

Hvor står vi nu?

Siden august 2000 er den tidligere folkeskolens udvidede afgangsprøve (FSU) blevet erstattet af 10. klasse prøven, som eleverne nu kan vælge at aflægge mundtligt og/eller skriftligt. Det er endnu for tidligt at udtale sig om, hvorvidt denne ændring vil medføre, at færre elever ønsker at aflægge både den mundtlige og den skriftlige prøve, eller om nogle vil vælge f.eks. kun den mundtlige del. Elever på 10. klasseniveau kan også vælge at aflægge FSA-prøven for første gang eller at gentage den. I så fald vil kun den højeste opnåede karakter blive påført prøvebeviset.

Engelsk og tysk/fransk

Som det fremgår af prøvebekendtgørelsen til FSA [Bilag](#), er forskellen på engelsk og de to tilbudsfag tysk og fransk, at der til engelskprøven stilles krav om kendskab til kultur- og samfundsforhold, og at eleverne kan udtrykke sig mere præcist og nuanceret, men ellers er rammerne de samme.

Vedrørende 10. klasse prøvens mundtlige del stilles der i alle tre fag krav om kendskab til kultur og samfundsforhold; men i engelsk lægges der vægt på, at eleverne kan udtrykke sig præcist og nuanceret, hvorimod tysk og fransk stiller krav til, at eleverne kan udtrykke sig klart, herunder i det væsentlige sprogligt korrekt" - store krav til de to sidste sprog. Vedr. den skriftlige del stilles der til engelsk krav om at kunne udtrykke sig hensigtsmæssigt og sammenhængende, og der stilles krav om færdighed i at udtrykke sig forholdsvis præcist og nuanceret.

Tysk/fransk stiller krav om, at eleverne kan udtrykke sig forståeligt og hensigtsmæssigt i en sammenhængende skriftlig fremstilling, der i det væsentlige er sprogligt korrekt. At mange elever opfatter tysk og fransk som vanskelige sprog i forhold til engelsk, kan ses ud af tallene for de elever, der ønsker at aflægge den skriftlige del af 10. klasse prøven. Langt færre elever ønsker at aflægge prøve i tysk eller fransk på 10. klasse niveau. Tallene for 2001 er for engelsk: 21.563, for tysk: 11.133, for fransk: 402. Ligeledes ses der en tendens til, at nogle elever fravælger at aflægge prøve i tysk/fransk efter 9. klasse, men vil vente til 10. klasse, hvor de som tidligere nævnt har mulighed for at aflægge FSA-prøven.

Klare Mål

I foråret 2002 udsendte ministeriet *Klare Mål* for en række fag, bl.a. engelsk, tysk og fransk. Formålet er at gøre læseplanernes krav mere synlige og til dels mere målbare. Ved en sammenligning sprogfagene imellem fremgår det nu meget tydeligere, at der er forskel på de krav, der stilles til engelsk, og dem der stilles til tysk/fransk, ja man kan sige, at kravene til de to sidstnævnte er blevet nedtonet.

Næsten identiske krav

En revidering af prøvebekendtgørelsens krav kunne være på sin plads. Både lærere og elever giver udtryk for, at kravene til tysk/fransk er for store. Med et tilbageblik på læseplansarbejdet, der førte frem til nye læseplaner i 1995, må man konstatere, at kravene blev formuleret identisk med dem, der stilles til engelsk, hvad der er urealistisk set i forhold til timetal. Dertil kommer mediernes betydning for elevernes tilegnelse af engelsk. Fra 4.-9. klasse er det vejledende timetal for engelsk 17 timer, inkl. 10. klasse 21 timer. For tysk/fransk er det vejledende timetal fra 7.9. klasse 11 timer, inkl. 10. klasse 15 timer. I realiteten har eleverne færre timer, da der ikke mindst på 8.-10. klassetrin er mange aktiviteter i form af praktik, brobygning, lejrskole, afvikling af skriftlige prøver m.m., hvor den faglige undervisning er aflyst. Det kunne derfor være ønskeligt i stedet for at operere med et vejledende timetal pr. uge at nå frem til en timetildeling baseret på en årsnorm for fagene.

Erfaringer fra prøverne - ønsket om ændringer

Hvert år i november/december udsender ministeriet de såkaldte PEU-hæfter (Prøver-Evaluering-Undervisning), der også lægges på ministeriets hjemmeside. [note 1](#) I disse hæfter findes en sammenskrivning af de beskikkede censorers rapporter vedr. afvikling af prøverne i den forløbne prøvetermin. Det fremgår klart, at der har været stor tilfredshed med engelskprøvernes rammer og krav, hvorimod som nævnt tysk og fransk har følt sig klemte. F.eks. har de 20 minutters forberedelsestid (jfr. prøvebekendtgørelsen) ikke givet problemer for engelsk. Eleverne kan nå både at lytte og læse de ukendte tekster, tage notater samt slå de enkelte ord op, de har brug for, men dette er ikke tilfældet for tysk/fransk, hvor mange elever føler, at de ikke kan nå at sætte sig ind i teksterne.

Selvom prøvebekendtgørelsen giver mulighed for at søge om tilladelse til at indgå i et forsøgs- og udviklingsarbejde, er der meget få lærere, der gennem årene har benyttet sig af dette, hvorfor vi i dag kun står med få forsøg og erfaringer med ændrede prøveformer. Ganske vist er der sket en udvikling i indholdet af prøverne, men rammerne er i høj grad de samme, som vi kender fra 1978. Fra ministeriets side har der derfor været opfordret til både i 1999 og 2000 at indgå i sådanne udviklingsarbejder, men måske er lærernes pressede arbejdssituation medvirkende til, at kun meget få har fulgt opfordringen.

Projektarbejdsformen og prøverne

Med 93-loven fik projektarbejdsformen en central placering, idet der indførtes en obligatorisk projektopgave i både 9. og 10. klasse. For 10. klasses vedkommende er projektopgaven med den nye 10. klasse lov blevet fjernet, og i stedet skal eleverne udarbejde en obligatorisk selvvalgt opgave med udgangspunkt i deres uddannelsesplaner. Projektarbejdsformens mere fremtrædende plads i den daglige undervisning har hos nogle lærere og deres elever medført ønske om at afprøve en mundtlig prøveform, hvor eleverne har mulighed for at samarbejde, at arbejde projektorienteret i forberedelsesfasen og at kunne blive eksamineret parvis eller gruppevis. I engelsk har det fortrinsvis drejet sig om muligheden for en par- eller gruppeprøve med længere forberedelsestid, samt mulighed for at samarbejde og have adgang til alle hjælpemidler, adgang til Internettet mm. Erfaringerne med disse forsøg har hidtil været meget positive.

Hvad eleverne finder positivt i forhold til den traditionelle prøve er:

- at de kan arbejde i deres eget tempo
- at de har mere tid til at sætte sig ind i et materiale
- at de kan hjælpe hinanden
- at de har mulighed for at samarbejde og derved finde flere argumenter til deres fremlæggelse
- at de har tid til at tænke på fremlæggelsen
- at de kan arbejde sammen med en, de kender
- at de tør noget mere, når de kender censor
- at de er mindre nervøse ved fremlæggelsen
- at tidsrammen tre timer til forberedelsen er tilpas.

Det negative er, at der kan være en del ventetid mellem forberedelse og fremlæggelse. Lærere og censorer har også været positive, men især for læreren er der et stort forberedelsesarbejde. Til den kommende prøvetermin vil man undersøge muligheden for en to timers forberedelsestid og en kortere ventetid for eleverne.

I tysk er også den projektorienterede prøveform også blevet afprøvet, men resultatet er ikke entydigt positivt, da denne prøveform bl.a. kræver, at eleverne kan kommunikere frit på sproget, hvilket mange elever har vanskeligt ved.

Hvor er vi på vej hen?

I slutningen af januar 2002 har et lille antal skoler indsendt ansøgninger om at få tilladelse til at afvikle en anderledes prøveform på 10. klasseniveau. Ansøgningerne ligger tæt op ad de forsøg, der hidtil er foretaget og som er beskrevet udførligt i PEU-hæfterne. For tysk er der desværre kun en enkelt ansøgning, for fransk ingen.

Da man i ministeriet gerne vil have afprøvet flere forskellige prøveformer, har der været tanker om at opfordre skoler til at undersøge forskellige modeller for den mundtlige prøve i engelsk, f.eks. en prøveramme på to timers forberedelsestid med mulighed for at arbejde parvis/gruppevis eller en slags synopseprøve, hvor eleverne gennem lodtrækning får tildelt et af årets temaer til at uddybe og fremlægge ved den mundtlige prøve. Samtidig har der været planer om at opfordre nogle skoler til at afprøve en skriftlig prøve på 9. klassetrin, men disse tanker vil evt. først kunne føres ud i livet i løbet af det kommende skoleår.

Men hvis der indføres en skriftlig prøve for engelsk i 9. klasse (FSA), hvilke konsekvenser vil dette medføre for den nuværende skriftlige prøve i 10. klasse? Skal den f.eks. erstattes af en anden prøveform, hvor eleverne prøves i sproglig præcision og nuancering, forhold som formentligt ikke vil blive vægtet væsentligt ved FSA-prøven? Arbejdes der f.eks. hen imod at udvikle prøveformer, der kan rettes mekanisk/elektronisk? Disse spørgsmål er på ingen måde gennemdrøftet eller afprøvet, så der forestår et stort arbejde, inden en ændring kan træde i kraft.

Ministeriet har tidligere meldt ud, at der evt. ville komme en ændring af bekendtgørelsen til skoleåret 2002/2003, men ingen forventer, at denne kan være klar til d. 1. august 2002. I stedet for for tidligt at lægge sig fast på nye prøveformer kunne man ønske sig en overgangsordning, så der blev givet valgmulighed mellem den gamle prøveform og en af de nye, således at de indhøstede erfaringer kunne indgå i en ny bekendtgørelse.

Konsekvenser for ændringer af de skriftlige prøver

Skulle man beslutte, at der skal være mulighed for at aflægge en evt. skriftlig prøve i engelsk til FSA, må denne blive et tilbud til eleven, men ikke et krav for at kunne aflægge den mundtlige del af prøven. Ellers vil de svage elever slet ikke have mulighed for at aflægge prøve, da de kun kan klare kravene til den mundtlige del. Disse elever finder vi også i 10. klasse, som eleverne netop har valgt for at blive mere målrettet i deres uddannelsesplaner og at opnå et bedre fagligt standpunkt. Hvis disse elever ikke fortsat har mulighed for kun at aflægge den mundtlige del, vil de så ikke blive tabt i det nye system?

Vedr. den skriftlige del af 10. klasseprøven har der for tysk og fransk været tanker om et mere struktureret prøveoplæg, hvor eleverne skal besvare nogle bundne opgaver. Dette kunne måske bevirke, at flere elever også ville vælge at aflægge den skriftlige del af prøven.

Vejene må skilles

Vi må forvente, at vejene må skilles inden for de næste par år, så prøverne i engelsk ikke længere vil være de samme som i tysk/fransk. Eleverne har en langt bedre sprogfærdighed i engelsk og dette må komme til udtryk i de nye prøveformer. Det ville langt være at foretrække, at systematiske forsøg bliver iværksat, så vi får prøver i alle tre fremmedsprog, der passer til de enkelte sprog og det niveau, eleverne realistisk set kan nå.

Noter

Prøvebekendtgørelsen for de enkelte fag kan ses på [Brick-bilag.html](http://www.uvm.dk/brick-bilag.html)

1 Prøver, Evaluering, Undervisning, 1997, 1998, 1999, 2000, 2001, Undervisningsministeriet. Findes ligeledes på www.uvm.dk.

Bilag

FSA - Engelsk

§ 30. Prøven er mundtlig.

Stk. 2. Der opgives et alsidigt sammensat tekstudvalg på 30-40 normalsider af passende sværhedsgrad, omfattende skønlitteratur og sagprosa. Herudover opgives ikke-skrevet materiale, for eksempel billeder, lyd og video. Stof om flere engelsktalende lande skal være repræsenteret i opgivelserne.

Stk. 3. Prøveoplægget består af skrevet og talt engelsk, eventuelt kombineret med billedmateriale. Det skrevne oplæg skal have en længde af ca. 1 normalside og det talte oplæg en varighed af 2¹ indholdsmæssig sammenhæng med det opgivne stof og afspejle sider af kultur- og samfundsforhold i engelsktalende lande. Oplægget skal være ukendt for eleven.

Stk. 4. Eleven får 20 minutters forberedelsestid og må anvende ordbøger og tage notater.

Stk. 5. Prøven foregår på engelsk. Den består i elevens redegørelse for det ukendte oplæg samt en samtale om indholdet, hvori også relevante dele af det opgivne stof inddrages. Der prøves i kommunikative færdigheder, herunder lytte- og læseforståelse og mundtlig sprogfærdighed. Der lægges vægt på, at eleven kan udtrykke sig forholdsvis præcist og nuanceret. I vurderingen indgår desuden elevens kendskab til kultur- og samfundsforhold i engelsktalende lande. Der gives en karakter.

FSA tysk/fransk

§ 31/§32. Prøven er mundtlig.

Stk. 2. Der opgives et alsidigt sammensat tekstudvalg på 30-40 normalsider af passende sværhedsgrad, omfattende skønlitteratur og sagprosa. Herudover opgives ikke-skrevet materiale, for eksempel billeder, lyd og video.

Stk. 3. Prøveoplægget består af skrevet og talt tysk/fransk, eventuelt kombineret med billedmateriale. Det skrevne oplæg skal have en længde af ca. 1 normalside og det talte oplæg en varighed af 2¹ indholdsmæssig sammenhæng med det opgivne stof og være ukendt for eleven.

Stk. 4. Eleven får 20 minutters forberedelsestid og må anvende ordbøger og tage notater.

Stk. 5. Prøven foregår på tysk/fransk. Den består i elevens redegørelse for det ukendte oplæg samt en samtale om indholdet, hvori også relevante dele af det opgivne stof inddrages. Der prøves i kommunikative færdigheder, herunder lytte- og læseforståelse samt mundtlig sprogfærdighed. Der lægges vægt på, at eleven kan udtrykke sig forholdsvis klart i et enkelt sprog. Der gives en karakter.

10. klasseprøven - engelsk

§ 40. Prøven er skriftlig og mundtlig, jf. dog § 5, stk. 4.

Stk. 2. Den skriftlige del af prøven omfatter en opgave i skriftlig fremstilling. Til besvarelsen gives der 3 timer. Eleven skal kunne disponere stoffet og udtrykke sig hensigtsmæssigt og sammenhængende i overensstemmelse med opgaven. Der stilles krav om færdighed i at udtrykke sig forholdsvis præcist og nuanceret. Der gives en karakter.

Stk. 3. Ved den skriftlige del af prøven må der anvendes ordbøger og grammatiske oversigter.

Stk. 4. Til den mundtlige del af prøven opgives der et alsidigt sammensat tekstudvalg på 50-60 normalsider af passende sværhedsgrad, omfattende skønlitteratur og sagprosa. Herudover opgives ikke-skrevet materiale, for eksempel billeder, lyd og video. Stof om flere engelsktalende lande skal være repræsenteret i opgivelserne.

Stk. 5. Prøveoplægget består af skrevet og talt engelsk, eventuelt kombineret med billedmateriale. Det skrevne oplæg skal have en længde af ca. 1 normalside og det talte oplæg en varighed af 2 ¹ indholdsmæssig sammenhæng med det opgivne stof og afspejle sider af kultur- og samfundsforhold i engelsktalende lande. Oplægget skal være ukendt for eleven.

Stk. 6. Eleven får 20 minutters forberedelsestid og må anvende ordbøger og tage notater.

Stk. 7. Prøven foregår på engelsk. Den består i elevens redegørelse for det ukendte oplæg samt en samtale om indholdet, herunder de kultur- og samfundsforhold, som oplægget berører. Der prøves i kommunikative færdigheder, herunder lytte- og læseforståelse samt mundtlig sprogfærdighed. Der lægges vægt på, at eleven kan udtrykke sig præcist og nuanceret. I vurderingen indgår elevens kendskab til kultur- og samfundsforhold i engelsktalende lande. Der gives en karakter.

10. klasseprøven tysk/fransk

§ 41/§42. Prøven er skriftlig og mundtlig, jf. dog § 5, stk. 4. Stk. 2. Den skriftlige del af prøven omfatter en opgave i skriftlig fremstilling. Til besvarelsen gives der 3 timer. Eleven skal i overensstemmelse med opgaven kunne udtrykke sig forståeligt og hensigtsmæssigt

i en sammenhængende skriftlig fremstilling, der i det væsentlige er sprogligt korrekt. Der gives en karakter.

Stk. 3. Ved den skriftlige del af prøven må der anvendes ordbøger og grammatiske oversigter.

Stk. 4. Til den mundtlige del af prøven opgives der et alsidigt sammensat tekstudvalg på 50-60 normalsider af passende sværhedsgrad, omfattende skønlitteratur og sagprosa. Herudover opgives ikke-skrevet materiale, for eksempel billeder, lyd og video.

Stk. 5. Prøveoplægget består af skrevet og talt tysk/fransk , eventuelt kombineret med billedmateriale. Det skrevne oplæg skal have en længde af ca. 1 normalside og det talte oplæg en varighed af 2 ¹ indholdsmæssig sammenhæng med det opgivne stof og være ukendt for eleven.

Stk. 6. Eleven får 20 minutters forberedelsestid og må anvende ordbøger og tage notater.

Stk. 7. Prøven foregår på tysk/fransk. Den består i elevens redegørelse for det ukendte oplæg samt en samtale om indholdet, hvori også relevante dele af det opgivne stof inddrages. Der prøves i kommunikative færdigheder, herunder lytte- og læseforståelse samt mundtlig sprogfærdighed. Der lægges vægt på, at eleven kan udtrykke sig klart, herunder i det væsentlige sprogligt korrekt. I vurderingen indgår elevens kendskab til kultur- og samfundsforhold i tysk/fransktalende lande. Der gives en karakter.

Tema: Evaluering - hvad skal vi kunne?



Prøveformer med IT i fransk og spansk

Jens Peder Weibrecht

Lektor, Nørresundby Gymnasium og HF-kursus.

Pædagogisk konsulent, Amtscentret for Undervisning i Nordjylland.

jp@weibrecht.dk · <http://www.weibrecht.dk>.

Mange lærere har opdaget, at internettet er en udtømmelig ressource af aktuelt stof. Der er allerede fri adgang til snesevis af aviser med dagligt aktuelt stof, i flere tilfælde med både tekst, billede og lyd. For sprogfag som fransk og spansk er denne mulighed for aktualitet af stor betydning fremover. Man kan beskæftige sig med stof, der kun sjældent når frem i den danske presse, man kan høre/se det på originalsproget, og man kan trække ting ind i undervisningen, som ikke tidligere har været umiddelbart tilgængelige.

Ud over aktualitet er autenticitet en væsentlig ny mulighed i forbindelse med IT. Det vil være oplagt at udvikle en mere autentisk sprogundervisning, der så hurtigt som muligt inddrager autentiske kommunikationssituationer mellem elever i Danmark og f.eks. andre europæiske lande.

Det er særdeles vigtigt, at man skaber mulighed for kommunikation fra elever til elever, fra lærere til lærere og elever og lærere indbyrdes. Som vi kender den nu, er kommunikationen på nettet skriftlig, men mon ikke vi om ti år vil have muligheder for mundtlig kommunikation?

Disse nye muligheder i sprogundervisningen må nødvendigvis følges af ændrede undervisningsformer i retning af mere projektorienteret arbejde og nye prøveformer, hvor IT indgår. Vi har gennem inddragelsen af internettet givet eleverne nogle nye kompetencer, som de selvfølgelig skal have lov til at vise ved eksamen og årsprøver. I det følgende vil jeg omtale et par forsøg med alternative prøveformer med IT ved årsprøver og studentereksamen.

Alternativ årsprøve: Eksempel 1

I maj 97 kom nogle elever til en alternativ årsprøve i fransk efter 2g. Ideen bag prøven var, at eleverne foruden den sædvanlige prøvetekst med oversættelse og spørgsmål, blev udstyret med noget materiale nedlogget fra internettet, som skulle indgå i besvarelsen. Dette materiale havde et omfang, der krævede, at eleverne kunne overskue det, men uden at kunne nå at læse alt, og de skulle så vise, at de var i stand til at fokusere på det væsentlige med et kritisk blik.

Emnet for opgaven var situationen for unge piger, der bliver gravide i en alder af 15-16 år. Hovedteksten handlede om Arlette, der skjuler sin graviditet for forældrene og dræber sit barn ved fødselen. Webstedet, der hed *Bébé-Bonheur*, indeholdt en udskrift fra en film af samme navn, hvor fem unge piger fortæller om deres situation, forhold til barnets far, forældrene, skolen etc.

Eleverne fik 5 timer til besvarelse af opgaven og lavede gode besvarelser, hvor de brugte både tekst og web-siderne på udmærket vis. Elevernes egen vurdering fremgår af følgende udsagn:

- "Jeg synes, at teksten var rimelig let at læse."
- "Stoffet på internet var med til at gøre selve opgaven lidt mere alsidig, og jeg synes desuden, det var godt, vi ikke selv skulle surfe på nettet."
- "Det dér med hyper-teksterne synes jeg også var en god udfordring - og en god måde at få flettet interaktivitet (*liens*) og sådan noget ind."
- "Det var spændende at skulle navigere rundt mellem de forskellige tekster, og jeg tror, det er godt at vi er vant til at bruge internet, for ellers tror jeg, det ville have været meget svært." [note 1](#)

Alternativ årsprøve: Eksempel 2

Som led i forberedelsen af en rejse til Frankrig/Spanien var 1d i 1999 til en 2-dages alternativ årsprøve i fransk (12 elever) og spansk (10 elever).

Vi ønskede med årsprøven at afprøve og styrke elevernes sproglige og grammatiske kunnen, men ville på samme tid gerne via arbejdet med IT udvide deres kundskaber med henblik på fortsat arbejde med telekommunikation og internet i undervisningen. Som forberedelse af studieturen skulle der indgå emner af kulturel og samfundsmæssig karakter, der ville styrke deres viden om Frankrig og Spanien. Arbejdsformerne og opgaverne skulle stille krav til elevernes selvstændighed og give dem mulighed for selv at søge og bearbejde information, så de kunne formidle deres resultater til klassen. Lærernes rolle skulle dels være at fungere som *coach* og konsulent, dels som eksaminator og censor.

Forberedelse til årsprøven: repetition af hidtil lært grammatik, d.v.s. basisgrammatikken i de to sprog. De to prøvedage blev afviklet i edb-lokaler, hvor eleverne havde rådighed over hver sin computer.

Første dag trak eleverne et emne, som de skulle søge oplysninger om på nettet. Emnerne havde tilknytning til den kommende rejse, og der var givet en række links på forhånd. Ved dagens afslutning skulle de aflevere et par sider som svar på opgaven. Besvarelsen skulle indeholde noget fransk/spansk tekst, men en del af den kunne også være på dansk. Denne besvarelse skulle danne udgangspunkt for en eksamen på andendagen.

På den anden dag kom alle til en ca. ti minutters individuel mundtlig eksamen, som bestod af oplæsning fra opgavebesvarelsen fra første dag, samtale på fransk/spansk ud fra samme og endelig et par grammatikspørgsmål. Nogle elever havde arbejdet parvis om en opgave, og de kom sammen til eksamensbordet. I løbet af dagen udarbejdede eleverne desuden enkeltvis e-mails. Der skulle være én personlig mail til den kommende vært i Spanien og én, der handlede om det emne, som de havde arbejdet med den første dag. Disse e-mails skulle være på fransk/spansk.

Vi havde håbet på at få en *real-time-chat* stablet på benene på et tidspunkt i løbet af de to dage, men det lykkedes kun ganske kortvarigt på grund af tekniske problemer.

Den alternative årsprøve blev afsluttet med en samlet karaktergivning for hele forløbet og den enkeltes indsats og en evaluering af projektet sammen med eleverne.

Eleverne skulle orientere sig i en stor stofmængde og uddrage centrale steder. Den store brug af IT betød, at de fik repeteret og trænet tekstbehandling, de søgte oplysninger på nettet via browser, de indsatte tekst og billeder i dokumenter og endelig brugte de FirstClass-kommunikationssystemet.

Eleverne arbejdede aktivt og med en høj grad af selvstændighed og der var enighed om, at klassen på de to dage lærte meget, både færdighedsmæssigt og mht. viden om det, de skulle ned og opleve. Nogle mente, de havde tilbragt lige lovlig lang tid ved maskinerne. [note 2](#)

Studentereksamen

Ved studentereksamen i 1998 for den elektroniske klasse og ved eksamen i 2001 for en fortsætter højniveauklasse har eleverne haft alternative opgaver, der foruden en tekst indeholdt en række websider. Disse alternative eksaminer er et tilbud til alle, som ønsker det, og det gælder også for eksamen 2002.

Det overordnede emne for opgaven i 1998 var *L'exil - entre deux vies*, hvor teksten fortæller om en algiersk kvinde bosiddende i Schweiz og hendes problemer med tilpasning og savnet af hjemlandet.

Webteksterne til emnet blev fordelt til fire opgaver med emner om den algierske kvinde, om borgerkrigen i Alger, om immigration og endelig om liv og traditioner i Algeriet. [note 3](#)

Evaluerings af studentereksamensopgaverne

Vi valgte fortrinsvis korte tekster og bragte også et helt fotogalleri, som kunne inddrages i besvarelsen, og da opgavesættet blev fremsendt på cd-rom, kan man nok i fremtiden forestille sig, at der også vil være andre kilder end tekst og billeder, der kan bruges. Hvorfor ikke sang, filmklip, interviews eller i det hele taget alt, hvad man kan hente fra internettet? En opgave skal altså ikke kun kunne udleveres som fotokopier.

Erfaringerne viser også, at man ikke skal begrænse omfanget af elevernes besvarelse. Hvis de har lyst til at skrive, så lad dem gøre det. Den eneste grænse for antal ord bør være en minimumsgrænse. Variationen af informationen og teksterne til opgaven betyder, at der er større mulighed for, at eleverne kan udfolde sig inden for det område, som de behersker, og de kommer ikke til at sidde tilbage med et valg af tre eller fire 'umulige' emner. Det er selvfølgelig svært at holde samme sværhedsgrad i tekstmaterialet, når det bliver så stort, men gør det noget? Skal der ikke være mulighed for differentiering?

Ved censorudtalelsen til opgaven i 1998 siges bl.a.:

"Det er klart relevant at arbejde med internettekster. Disse samtidens tekster er skabt til ekstensiv læsning, og selvfølgelig skal vi lære eleverne at sortere i informationsmængden. De skal trænes i en kvalificeret læsning af det elektroniske materiale og i en målrettet anvendelse af IT-stof.

C-opgavernes omfang og disposition: Alle eleverne udnytter, som de skal, internetmaterialet. Naturligvis med skiftende held og stringens. I en enkelt besvarelse er internetteksten næppe helt forstået, i en anden plukkes lidt tilfældigt i materialet, mens de øvrige formår at udtrække få og væsentlige pointer eller citater fra det yderst righoldige materiale."

Konklusionen må blive, at det er vigtigt at arbejde videre med IT i nye prøveformer, som giver eleverne mulighed for at afprøve og fremvise deres nye kompetencer, og hvor de får lejlighed til at bruge alle de muligheder som moderne teknik og internet nu tilbyder.

Noter

1 Se endvidere rapporten på adressen: <http://pub.uvm.dk/1999/elskole/>

2 Opgaver mm fra årsprøve kan ses på følgende adresse:

<http://www.udd.nrsbgym.dk/jwmappe/1dmappe/index.html>

3 Se opgaven: <http://www.fr.gymfag.dk/studentereksamenmaj01/Startside.htm>

Tema: Evaluering - hvad skal vi kunne?



Studieprøven i dansk som andetsprog - en ny prøve

Else Lindberg

Pædagogisk konsulent.

Ministeriet for Flygtninge, Indvandrere og Integration.

Studieprøven i dansk som andetsprog afholdes første gang ved prøveterminen maj/juni 2002 og erstatter dermed Danskprøve 2, der siden slutningen af 70'erne har været det dansksproglige adgangsgrundlag til videregående uddannelser for kandidater, som ikke har en dansk adgangsgivende eksamen.

Udviklingen af studieprøven Dansk som fremmedsprog blev igangsat af undervisningsministeriet i foråret 2000. Udgangspunktet for dette udviklingsarbejde var en behovsundersøgelse på de videregående uddannelser, som områderne for Folkeoplysning og De Videregående Uddannelser gennemførte i samarbejde. Som grundlag for undersøgelsen benyttedes de funktionelle beskrivelser af sprogfærdighed, såkaldte *Can-Do-Statements*, som er udviklet af den europæiske sammenslutning *Association of Language Testers in Europe* (ALTE). Disse funktionelle beskrivelser viste sig at være særdeles anvendelige i analysen af, hvilke sproglige krav, der må stilles ved optagelse på en række videregående uddannelser. Det viste sig bl.a., at de receptive færdigheder i væsentlig grad udgør en barriere på studierne.

Ministeriet nedsatte efterfølgende en arbejdsgruppe af fagkyndige, der bistod med at udarbejde specifikationer for en færdighedsprøve, hvor prøvens formål, fokus og format blev nærmere fastlagt. På grundlag af disse specifikationer blev der udarbejdet prototyper på opgaver, som blev afprøvet, evalueret og justeret.

Studieprøven svarer til den hidtidige Danskprøve 2, dvs. Europarådets angivne C1-niveau (*Higher Level*). [note 1](#) I mange europæiske lande er dette niveau adgangsgrundlaget for videregående uddannelser (f.eks. *Test in Swedish for University Studies* (TISUS) og *Certificate of Advanced English* (CAE)).

Og hvad karakteriserer så i øvrigt den nye prøve? Som udgangspunkt er der tale om en kriterierelateret færdighedsprøve, som omfatter de fire kardinalfærdigheder i kommunikativ kompetence: læseforståelse, skriftlig fremstilling, lytteforståelse og mundtlig kommunikation. Der er således opstillet kriterier som gør det muligt at bedømme de receptive og produktive færdigheder i forhold til fastlagte og beskrevne niveauer inden for de enkelte færdigheder. Denne bedømmelsesmåde står i modsætning til en normrelateret bedømmelse, hvor man bedømmer besvarelsene i forhold til hinanden. De receptive færdigheder har fået en fremtrædende plads, og mundtlig kommunikation er gjort til en væsentlig og selvstændig del af prøven.

Præstationerne vurderes som en profil af de enkelte færdigheder, hvilket gør prøveresultaterne informative for

brugerne og giver kandidaterne mulighed for selektivt og målrettet at forbedre profilen efter behov. Der er ingen begrænsning på det antal gange, man kan indstille sig til Studieprøven, og en bestået mundtlig eller skriftlig del skal ikke tages om. Har man f.eks. bestået den skriftlige del, kan man ved en senere prøvetermin indstille sig til den mundtlige del og omvendt.

Da der er tale om en højrisiko-prøve har der været mange overvejelser mht. de krav, der må stilles til prøvens kvalitet. Det har f.eks. haft betydning for valget af læse- og lytteformål i opgaverne og den mere tekniske afprøvning af de receptive færdigheder, som er lagt ind i 'opgavebatteriet' for at styrke validiteten og reliabiliteten. Opgaverne er gjort indholdsmæssigt brede, således at flere tematiske domæner dækkes i teksterne, uden at disse dog er tilpasset bestemte studieretninger.

I skriftlig fremstilling er der sat fokus på funktionelle sproghandlinger som f.eks. at kunne videregive faktuel information, tage stilling til udsagn og underbygge med relevant og sammenhængende argumentation. Dette gælder også for mundtlig kommunikation, hvor prøvegrundlaget i øvrigt er en række emner, som kandidaterne forbereder ved at søge informationer og lægge det fokus på emnet, som de selv finder interessant og relevant for samtalen under prøven.

De bindende regler for prøven fremgår af bekendtgørelse nr. 76 af 12. februar 2002 om prøver inden for undervisning i dansk som andetsprog for voksne udlændinge m.fl., som er udstedt af Ministeriet for Flygtninge, Indvandrere og Integration. De endelige prototyper indgår i en vejledning om Studieprøven, som udsendes af ministeriet i løbet af april måned 2002.

Note

1 Se Common European Framework of Reference for Languages: Learning, teaching, assessment, Cambridge University Press, 2001.

Tema: Evaluering - hvad skal vi kunne?

Skriftlig eksamen i engelsk i læreruddannelsen



Else Brendholdt
Lektor, cand.mag. Skive
Seminarium.
Medlem af opgaveudvalget
(engelsk) 1999-2002.

Tage Øe
Seminarielektor, cand.pæd.
Nørre Nisum
Seminarium. Formand for
opgaveudvalget.



Anders Hove
Seminarielektor, cand.mag.
Ribe Statsseminarium/CVU
Vest. Medlem af
opgaveudvalget.

I 1992 skete der noget bemærkelsesværdigt i dansk læreruddannelsestradition. Der blev ikke længere udarbejdet fælles skriftlige eksamensopgaver. Der var stadigvæk et krav om skriftlig eksamen, men fremover skulle hvert enkelt seminarium selv sørge for sine egne opgaver. I engelsk tog den faglige forening, SELF, udfordringen op og forsøgte på faglige seminarer at nå frem til fælles retningslinier, men det viste sig hurtigt at meningene var meget forskellige. Derfor endte vi med at have 18 forskellige opgaver. Nogle havde fællestræk, men der var også stor variation. Nogle var meget traditionelt udformet, mens andre viste spændende nytænkning. SELF gjorde en dyd af nødvendigheden og betragtede situationen som en slags forsøgsordning, og tilskyndede derfor til fortsat debat og inspiration, blandt andet ved i hvert nummer af foreningens tidsskrift at trykke et eksamenssæt fra forskellige seminarier, suppleret med det pågældende seminariums begrundelser og erfaringer. Hertil kom den faglige drøftelse der følger af udveksling af censorer.

Én fælles

Hele decentraliseringen af læreruddannelsen, og herunder også de skriftlige prøver, begyndte imidlertid at give politisk bekymring. Det gjaldt såvel forskellighederne som det faglige niveau. Hvis man fra politisk side havde været parat til at gennemføre en egentlig evaluering, kunne 'forsøgsperioden' måske have ført til virkelig gode og måske endda forholdsvis ensartede lokale opgaver. Men i stedet blev de centralt stillede opgaver genindført med læreruddannelsesloven af 1997. Endvidere skulle der i modsætning til tidligere gives en selvstændig karakter for besvarelsen. De ministerielt udpegede opgaveudvalg fik meget frie hænder til udformningen af opgaverne. Forud for det første rigtige eksamenssæt skulle der dog udarbejdes to modelsæt, der skulle drøftes i de faglige kredse, herunder også i censorkorpset der rummer censorer uden for seminarierne.

Hvad skal engelsklæreren kunne?

I opgaveudvalget for engelsk hilste vi den store frihed velkommen, men måtte jo samtidig erkende at den lagde et stort ansvar på os. Det var nødvendigt med nogle styringsredskaber, og vi brugte derfor en del tid på at opstille de kriterier vi ville lægge til grund for opgaverne. De ville også være en stor hjælp ved drøftelserne af modelsættene. Vi tog naturligvis helt overordnet afsæt i bekendtgørelsen og i liniefagets CKF, altså ministeriets angivelser af de centrale kundskabs- og færdighedsområder. Ligeledes måtte vi se på hvad man som engelsklærer i folkeskolen i dag skal kunne; det vil sige folkeskolens CKF og skriftlige prøver, ministeriets årlige evalueringer m.v. Når vi tager dette - egentlig helt selvfølgelige - udgangspunkt, står vi med det problem at tid og ressourcer, og måske de studerendes forudsætninger, gør det svært fuldt ud at leve op til kravene. Men det løser vi ikke ved blot at indskrænke kriterierne, og slet ikke i en tid hvor der stilles krav om større faglighed.

Selv om der desværre aldrig blev foretaget en egentlig evaluering af de lokalt stillede opgaver, brugte vi naturligvis de erfaringer der trods alt var blevet opsamlet. Et af de træk der generelt syntes at have været en succes, var en tilknytning til emner og tekster man lokalt havde arbejdet med. Dette syntes at være mere befordrende og inspirerende end emner uden en sådan sammenhæng. Der var også udsagn om at tilknytningen gav større seriøsitet i arbejdet med litteratur, samfund og kultur. Disse erfaringer passer jo dårligt med centralt stillede opgaver, men for ikke at tabe dem helt på gulvet kom opgavesættene til at bestå af en større tekstsamling der udleveres 24 timer før den egentlige prøve. Det har efterfølgende vist sig at de øvrige fag har valgt en helt tilsvarende todeling af prøven.

Kriterierne

Som sagt opstillede vi en række kriterier. De blev sammen med de to modelsæt sendt ud til kollegial drøftelse. De blev ligeledes lagt på SELF's hjemmeside og på Skolekom. I forlængelse af de kollegiale drøftelser har vi foretaget nogle ændringer, men grundlæggende er kriterierne blevet godt modtaget. Her er de som de gælder i dag:

CKF (undtagen 1.3/2.3/2.4) danner grundlaget

Prøven er todelte:

- 24 timers forberedelsestid med en udleveret tekstsamling omkring et tema
- 6 timers prøvetid

Vedrørende 24-timers forberedelsestiden:

Det forventes at de studerende anvender forberedelsestiden til grundig læsning af de forelagte tekster, refleksion, sproglig forberedelse samt inddragelse af evt. tidligere bearbejdet stof. Endvidere kan informationssøgning være relevant. Den studerende kan forberede sig sammen med en eller flere medstuderende, men må ikke have kontakt med læreren.

Stikord vedrørende den udleverede tekstsamling:

- Tekster på engelsk fra de engelsksprogede områder
- Tekster på engelsk som lingua franca
- Kan indeholde tekster på dansk
- Fiktion såvel som non-fiktion
- Lyttetekst (evt med udskrift)
- Billeder

Vedrørende 6-timers prøven:

Der stilles tre opgaver der alle skal besvares:

- En ræsonnerende/argumenterende tekst. Opgaven kan fokusere på kultur- og samfundsforhold, herunder litteratur, den kan være didaktisk, og den kan være kombinationer af disse elementer.
- En kreativ tekst.
- En tekst der afprøver sproglig præcision i form af transformation eller kort oversættelse. Transformationer vil være på grundlag af (dele af) udleveret dansk eller engelsk tekst. Oversættelser skal have forbindelse med temaet, enten som et udsnit af allerede udleveret dansk tekst, eller en tekst der først udleveres ved 6-timers prøvens begyndelse.

Der sættes en begrænsning i de enkelte produkters længde. Til gengæld kræves gennemarbejdning af produktet hvad angår indhold, form og sprog.

Hjælpemidler til 6 timers prøven:

- Alle hjælpemidler tilladt, herunder notater fra de 24 timers forberedelse.
- Regler for brug af PC afgøres på det enkelte seminarium.
- Internetopkobling og netkommunikation ikke tilladt under prøven.

Bedømmelseskriterier:

- Dokumentation for forståelse af de forelagte teksters centrale holdninger, meninger og følelser
- Tilpasning til målgruppen
- Bevidst brug af genrer
- Fokusering og vægtning, herunder klar fremlæggelse af synspunkter og argumentation
- Kohærens og kohæsion
- Semantisk præcision mht. register, idiomatik og kollokationer
- Korrekt grammatik, stavning og tegnsætning

Emner og tekster

Emnerne har hidtil været:

- *Child Labour* (modelsæt 1: 16 tekster, 46 sider)
- *The Environment* (modelsæt 2: 11 tekster, 29 sider) *Body*
- *Culture* (sommer 2000: 12 tekster, 28 sider) *Partnership*
- *with Parents* (januar 2001: 13 tekster, 32 sider) *Beginning*
- *Teacher* (sommer 2001: 14 tekster, 45 sider)

Alle teksterne i et sæt drejer sig om et bestemt emne. Tekster defineres bredt og indeholder blandt andet en hel del billeder og i øvrigt noveller, uddrag af romaner eller drama, digte, non-fiction, herunder reklamer og websider. Disse sidste volder lidt problemer, idet de ikke er omfattet af hverken CopyDan-aftale eller lignende. Hidtil har kun modelsæt 2 rummet en lydtekst. Vi vil gerne have lyttetekster med, men har indtil videre måttet bøje os for nogle tekniske hindringer. Vores vision er at hele opgavesættet udsendes på en CD-ROM. Det vil åbne for såvel lyd som film og naturligvis billeder og tekst i en integreret helhed. Problemet er at det forudsætter fuld adgang til computere for alle, både i forberedelsesfasen og ved selve prøven.

Som det fremgår af emneoversigten ovenfor, har omfanget hidtil varieret temmelig meget. Noget af forskellen skyldes at der ikke er tale om normalsider, men blot trykte sider med stor variation i typografi og layout. I de kollegiale drøftelser har det været fremført at opgaverne ikke i tilstrækkelig grad krævede inddragelse af teksterne. Det er klart at man ikke kan inddrage 10 til 16 forskellige tekster i sin besvarelse, men vi har dog ikke planer om at reducere antallet væsentligt ud fra den betragtning at en del af teksterne er til inspiration både hvad angår indhold, form og sprogbrug. Men i øvrigt har vi taget kritikken til os og stiller nu krav om at nogle tekster - enten bestemte eller efter eget valg - tydeligt inddrages.

6-timers prøven

Ud fra den betragtning at formål og modtager betyder meget for teksternes udformning, og at det skal være naturligt at skrive på engelsk, har vi for hvert opgavesæt angivet en kontekst. Vi tilstræber at undgå kunstige eller indviklede kontekster. Det er måske nok nogenlunde lykkedes i modelsæt 2 herunder, men nogle af de studerende der har brugt sættet som prøvekaniner, har dog kritiseret os for i B-opgaven at lægge op til én bestemt 'politisk korrekt' besvarelse, når vi bad dem identificere sig med en faktisk eksisterende miljøaktivist med meget radikale synspunkter som de absolut ikke delte.

Modelsæt 2:

A. Reflection and argumentation.

As a teacher of English in the Danish school system, write an article for a website where European teachers promote the teaching of environmental matters across the curriculum in their school systems.

*WHY should the teaching of English also be the teaching of environmental matters?
HOW can the teaching of English also be the teaching of environmental matters?*

Your use of the materials package must be clearly seen in the article.

Use quotes very sparingly.

The length should not exceed 2 pages (5600 characters).

B. Creative Writing.

Imagine that Julia Butterfly (pages 2-11) celebrated her second anniversary as a tree sitter delivering a speech to Luna in which she showed her concern, commitment and determination. Write her speech.

The length should be approximately 1 page (2800 characters).

C. Translation.

As a teacher of English you are involved in planning an exchange visit from a group of foreign pupils. The visit will include field work about environmental matters for both the foreign pupils and their Danish counterparts, and to this end you are asked to send an English translation of the following text to your foreign colleagues:

NATURVEJLEDNING

Formålet med naturvejledning er blandt andet at formidle, at naturen er en helhed af

landskab, hav, klima, planter, dyr, historie og kultur, som mennesket indgår i, er afhængig af og påvirker. Et af formålene er også at formidle om naturen i byen og sammenhængen mellem bylivet og miljøet. Naturvejledningen koster omkring 45 millioner kroner om året. De 17 millioner kommer fra centrale fonde, blandt andet naturforvaltningsmidler, mens de resterende 28 millioner kroner kommer fra forskellige offentlige myndigheders, institutioners og foreningers driftsmidler, hvor naturvejlederne er ansat. (Søren Olsen, Morsø Folkeblad, fredag den 1. oktober 1999)

Valgmuligheder?

Nogle af de mest debatterede spørgsmål har uden tvivl været opgavernes karakter og genrekrav, samt hvorvidt der bør gives valgmuligheder. Det har været vigtigt for udvalget at holde fast i at når eleverne i folkeskolen skal kunne skrive i forskellige genrer eller teksttyper, ja, så skal læreren kunne det endnu bedre, og derfor skal de ikke kunne fravælge bestemte opgaver og dermed teksttyper. Det har især været den 'kreative' B-opgave der har vakt bekymring, men også den gamle strid om oversættelse kontra anden form for sproglig præcisionsopgave. Som det kan ses i januar 2001-sættet, har vi forsøgt et kompromis med valgmuligheder inden for B- og C-opgaverne.

Januar 2001:

Context:

Imagine that you participated in the conference in Paris organised by the European Parents' Association (EPA) under the title "Training of teachers on partnership relations with parents" (see Text 2).

One of the purposes of the conference was to "launch a project focusing on the creation of a module to be included in the initial and in-service training of teachers about partnership with parents."

Working groups with participation of parents and teachers were established in order to discuss the content and structure of such a module. As a teacher representative from Denmark you became a member of one of these working groups.

Your working group decided to base your discussion on materials from both real life and fiction, selected by the whole group. (see Texts 1 - 12)

Besides, each group member has selected one text from his own country. (see Text 13)

As a last task each group member was asked to write an individual contribution to be read by the other members in the group.

Examination tasks:

You have **three** writing tasks:

A Partnership with parents (maximum length 2 pages)

Discuss the question "How much should the teacher know about families in order to create a real and constructive partnership with parents?" Your discussion must be based on an analysis of one or two of the family situations in the materials.

B Memoirs of childhood (maximum length 1 page)

You are required

either

to write your own personal memoirs of a decisive incident in your childhood,

or

to write Melinda's memoirs of one of the incidents in "Approximations" (Text 10).

Imagine that Melinda is now a grandmother.

C. The Danish contribution

In order to communicate "Børneliv år 2000" (Text 13) to the non-Danish readers, you are required

either

to select a key passage of about 15 lines of the winning mother's entry and translate into English,

or

to write an English summary of the winning mother's entry (maximum length 1/2 page)

Bedømmelseskriterierne

Genrebegrebet har givet anledning til nogen diskussion. Det er blevet fremført at uanset hvad vi beder om, så er genren i sidste ende 'eksamensbesvarelse'. Det er måske rigtigt nok, men det berettiger alligevel efter vores mening ikke til at droppe kriteriet om 'bevidst brug af genrer', med mindre man da generelt afviser *make believe*-situationer og -opgaver i uddannelsessammenhæng. Kriteriet er i øvrigt ændret fra "brug af tekstgenrens konventioner", i erkendelse af at genrer - ud over det helt banale - ikke er veldefinerede størrelser, men hænger sammen med hvad teksten skal bruges til og hvem den er rettet imod.

Det er måske lidt overraskende at vi angiver maximumslængde for hver opgave, men vi ønsker gennemarbejdede besvarelser og ikke lange, dårligt gennemtænkte og ustrukturerede besvarelser. Det fremgår af bedømmelseskriterierne at vi ønsker at stille lige så store krav til indhold, struktur, genre og semantik som vi traditionelt gør til korrekthed. Vi er overbeviste om at denne vægtning vil påvirke det daglige arbejde, blandt andet i form af bevidst arbejde med skriftsprogets virkemidler og seriøst arbejde med processkrivning. Kravet om gennemarbejdning kan i øvrigt lettere imødekommes ved brug af PC, og vi håber at dette snart bliver en selvfølgelig del af skriftlig eksamen ved alle seminarier. Da vi har en situation med besvarelser både med og uden brug af PC, ville det være interessant at undersøge om der er kvalitative forskelle.

I det hele taget er der brug for et egentligt evaluerings- og udviklingsprojekt omkring det skriftlige arbejde og eksamen, så fremtidige ændringer i såvel arbejdsformer som eksamen kan baseres på et solidt grundlag. Vi havde i opgaveudvalget gerne set at en begrænset del af censorkorpset tog sig af den skriftlige censur. Det ville medvirke til at bedømmelseskriterierne virkelig slog igennem så der kom en nødvendig ensartethed i bedømmelserne. På en faglig dag i efteråret 2001 skulle omkring 25 seminarielærere bedømme den samme besvarelse. Nogle mente at den var klart bestået, mens andre mente den var lige så klart ikke-bestået. De første begrundede det med at der stort set ikke var fejl i grammatik, stavning og tegnsætning, mens de sidste gjorde opmærksom på at der var alvorlige problemer med alle de øvrige bedømmelseskriterier. Det er en uholdbar situation!

Det forekommer yderst relevant at runde af med at citere folkeskolens fagkonsulent der i ministeriets 'Prøver - evaluering - undervisning 2000' skriver:

"Nogle elever overfører desuden en del uvaner med at udtrykke sig i ultrakorte sætninger og anvender næsten ingen konjunktioner. Her ligger der et stykke arbejde for læreren med at få eleverne til at opfatte forskellen på de forskellige genrer, og at man alt efter form og modtager må variere stillaget. () Mange censorer påpeger, at eleverne bør bruge mere tid på at disponere deres besvarelse og sørge for afsnitsinddeling. Hvis eleverne i dansk

har arbejdet med strukturering af tekst, skinner det også igennem i deres engelske besvarelser. Det er påfaldende, at der er flere veldisponerede besvarelser med indledning og afslutning, når det er håndskrevne tekster."

Dette er en opfordring til ikke blot dansklæreren, men så sandelig også engelsklæreren. Og bedst bliver det ved et godt samarbejde. Dette er en vigtig del af alle de kompetencer vi skal give de kommende engelsklærere med i rygsækken.

CKF (undtagen 1.3/2.3/2.4) danner grundlaget

Prøven er todelt:

- 24 timers forberedelsestid med en udleveret tekstsamling omkring et tema
- 6 timers prøvetid

Vedrørende 24-timers forberedelsestiden:

Det forventes at de studerende anvender forberedelsestiden til grundig læsning af de forelagte tekster, refleksion, sproglig forberedelse samt inddragelse af evt. tidligere bearbejdet stof. Endvidere kan informationssøgning være relevant. Den studerende kan forberede sig sammen med en eller flere medstuderende, men må ikke have kontakt med læreren.

Stikord vedrørende den udleverede tekstsamling:

- Tekster på engelsk fra de engelsksprogede områder
- Tekster på engelsk som lingua franca
- Kan indeholde tekster på dansk
- Fiktion såvel som non-fiktion
- Lyttetekst (evt med udskrift)
- Billeder

Vedrørende 6-timers prøven:

Der stilles tre opgaver der alle skal besvares:

- En ræsonnerende/argumenterende tekst. Opgaven kan fokusere på kultur- og samfundsforhold, herunder litteratur, den kan være didaktisk, og den kan være kombinationer af disse elementer.
- En kreativ tekst.
- En tekst der afprøver sproglig præcision i form af transformation eller kort oversættelse. Transformationer vil være på grundlag af (dele af) udleveret dansk eller engelsk tekst. Oversættelser skal have forbindelse med temaet, enten som et udsnit af allerede udleveret dansk tekst, eller en tekst der først udleveres ved 6-timers prøvens begyndelse.

Der sættes en begrænsning i de enkelte produkters længde. Til gengæld kræves gennemarbejdning af produktet hvad angår indhold, form og sprog.

Hjælpemidler til 6 timers prøven:

- Alle hjælpemidler tilladt, herunder notater fra de 24 timers forberedelse.
- Regler for brug af PC afgøres på det enkelte seminarium.
- Internetopkobling og netkommunikation ikke tilladt under prøven.

Bedømmelseskriterier:

- Dokumentation for forståelse af de forelagte teksters centrale holdninger, meninger og følelser
- Tilpasning til målgruppen
- Bevidst brug af genrer
- Fokusering og vægtning, herunder klar fremlæggelse af synspunkter og argumentation
- Kohærens og kohæsion
- Semantisk præcision mht. register, idiomatik og kollokationer
- Korrekt grammatik, stavning og tegnsætning

Modelsæt 2:**A. Reflection and argumentation.**

As a teacher of English in the Danish school system, write an article for a website where European teachers promote the teaching of environmental matters across the curriculum in their school systems.

WHY should the teaching of English also be the teaching of environmental matters?

HOW can the teaching of English also be the teaching of environmental matters?

Your use of the materials package must be clearly seen in the article.

Use quotes very sparingly.

The length should not exceed 2 pages (5600 characters).

B. Creative Writing.

Imagine that Julia Butterfly (pages 2-11) celebrated her second anniversary as a tree sitter delivering a speech to Luna in which she showed her concern, commitment and determination. Write her speech.

The length should be approximately 1 page (2800 characters).

C. Translation.

As a teacher of English you are involved in planning an exchange visit from a group of foreign pupils. The visit will include field work about environmental matters for both the foreign pupils and their Danish counterparts, and to this end you are asked to send an English translation of the following text to your foreign colleagues:

NATURVEJLEDNING

Formålet med naturvejledning er blandt andet at formidle, at naturen er en helhed af landskab, hav, klima, planter, dyr, historie og kultur, som mennesket indgår i, er afhængig af og påvirker. Et af formålene er også at formidle om naturen i byen og sammenhængen mellem bylivet og miljøet.

Naturvejledningen koster omkring 45 millioner kroner om året. De 17 millioner kommer fra centrale fonde, blandt andet naturforvaltningsmidler, mens de resterende 28 millioner kroner kommer fra forskellige offentlige myndigheders, institutioners og foreningers driftsmidler, hvor naturvejlederne er ansat.

(Søren Olsen, Morsø Folkeblad, fredag den 1. oktober 1999)

Januar 2001:**Context:**

Imagine that you participated in the conference in Paris organised by the European Parents' Association (EPA) under the title "Training of teachers on partnership relations with parents" (see Text 2).

One of the purposes of the conference was to "launch a project focusing on the creation of a module to be included in the initial and in-service training of teachers about partnership with parents."

Working groups with participation of parents and teachers were established in order to discuss the content and structure of such a module. As a teacher representative from Denmark you became a member of one of these working groups.

Your working group decided to base your discussion on materials from both real life and fiction, selected by the whole group. (see Texts 1 - 12)

Besides, each group member has selected one text from his own country. (see Text 13)

As a last task each group member was asked to write an individual contribution to be read by the other members in the group.

Examination tasks:

You have **three** writing tasks:

A Partnership with parents (maximum length 2 pages)

Discuss the question "How much should the teacher know about families in order to create a real and constructive partnership with parents?" Your discussion must be based on an analysis of one or two of the family situations in the materials.

B Memoirs of childhood (maximum length 1 page)

You are required

either

to write your own personal memoirs of a decisive incident in your childhood,

or

to write Melinda's memoirs of one of the incidents in "Approximations" (Text 10). Imagine that Melinda is now a grandmother.

C. The Danish contribution

In order to communicate "Børneliv år 2000" (Text 13) to the non-Danish readers, you are required

either

to select a key passage of about 15 lines of the winning mother's entry and translate into English,

or

to write an English summary of the winning mother's entry (maximum length 1/2 page)

Tema: Evaluering - hvad skal vi kunne?

In English
Please!



Kvalitet og retfærdighed ALTE's Code of Practice

Nick Saville
University of Cambridge.
Local Examinations Syndicate, UK.

ALTE - *Association of Language Testers in Europe* (Foreningen af Sprogtestere i Europa) - blev dannet i 1990 da repræsentanter for otte institutioner deltog i et stiftende møde i Barcelona. Alle stiftende Medlemmer var udbydere af eksaminer og arbejdede med evaluering af deres egne sprog som fremmedsprog. På dette første møde blev det klart at alle havde en fælles interesse i høje standarder for sprogevaluering og for retfærdig behandling af de kandidater der tog deres eksaminer. Denne fælles interesse har været en samlende faktor, og har skabt et fokus for samarbejde i det forløbne årti. Så tidligt som i 1991 blev det besluttet at foreningens hovedmål skulle være:

"at skabe professionelle standarder for alle trin i sprogtest-processen",

med det formål at give sprogtilegnere adgang til eksaminer af høj kvalitet, som kunne hjælpe dem med at nå deres mål i livet, og som kunne have en positiv effekt på deres læringsoplevelser.

I de senere år er ALTE's arbejde blevet kendt i Europa, og medlemstallet er steget støt. Der er nu 27 institutionsmedlemmer og -tilknyttede, som repræsenterer 24 europæiske sprog.

Medlemmerne omfatter nu: Undervisningsministerier og andre ministerielle afdelinger, eksamenskommissioner, kulturinstanser, universitetsafdelinger og sprogundervisningsinstitutioner. I nogle tilfælde er der blevet dannet sprogtest-konsortier inden for rammerne af ALTE, f.eks. i Danmark, hvor Undervisningsministeriet samarbejder med Københavns Universitet og sprogcentre.

Inden for ALTE besluttede de stiftende Medlemmer at det var vigtigt for både eksamensudviklere og eksamensbrugere at følge en fastlagt *code of practice* (kodeks for praksis) som kunne sikre at evalueringsprocedurerne var af en høj kvalitet, og at alle interessenter blev retfærdigt behandlet. En sådan *code of practice* tager udgangspunkt i sunde principper for god evalueringspraksis, som gør det muligt at opnå høje standarder for kvalitet og retfærdighed. Diskussionen om hvad god praksis inden for ALTE er, viser at man er optaget af at der skal kunne redegøres for alle områder af den evaluering som ALTE's Medlemmer udfører. I denne henseende lægger foreningen vægt på betydningen af validering og den rolle som forskning og udvikling

spiller for evalueringsprocesser.

Med publiceringen af sin første *Code of Practice* i 1994 opstillede ALTE de standarder som Medlemmerne havde sat som mål for deres sprogtests. Den trak på *The Code of Fair Testing Practices in Education* (Kodeks for Retfærdig Testpraksis inden for Uddannelser) udviklet af *Washington D.C. Joint Committee on Testing Practices* (1988), og hensigten var at der skulle være en bred formulering af hvad brugerne af ALTE-eksaminer kunne forvente, og hvilke roller og hvilket ansvar interessenterne havde i forsøgene på at opnå retfærdighed.

Brugergrupper

Code of Practice identificerer tre hovedgrupper af interessenter i test-processen:

- Opgavestillerne, (dvs. eksamenskomitéer og andre institutioner som er Medlem af ALTE)
- De primære brugere - de kandidater der tager eksamen efter eget valg, gennem henvisning, eller af nødvendighed
- De sekundære brugere, som er kandidaternes sponsorer, eller andre som har en interesse i at kandidaterne tager eksamen (f.eks. forældre, lærere, arbejdsgivere, ansvarlige for optagelse osv.).

Desuden udstikker *Code of Practice* disse interessentgruppers ansvar inden for fire brede områder:

- Udvikling af eksaminer
- Tolkning af eksamensresultater
- Bestræbelser på retfærdighed
- Informering af eksaminander

Det er et vigtigt træk at *Code of Practice* understreger interessenternes fælles ansvar og fokuserer på eksamensbrugernes såvel som eksamensudviklernes forpligtelser i bestræbelserne på at opnå retfærdighed.

I Nordamerika har der inden for psykologiske og uddannelsesmæssige evalueringer været en lang tradition for at opstille standarder, således som det afspejles i 1999-udgaven af *Standards for Educational and Psychological Testing* (Standarder for uddannelsesmæssig og psykologisk testning (aera/apa/ncme)). Tidligere udgaver af disse standarder har eksisteret siden 1950'erne, og har haft indflydelse på udvikling og brug af sprogprøver, især i USA. Der bliver jævnligt henvist til dem i litteraturen om sprogprøver (se f.eks. Bachman 1990 og Bachman & Palmer 1996).

Især inden for sprogevalueringområdet har der været en stigende interesse for at fastlægge standarder, herunder for udviklingen af *codes of practice*, og en interesse for etiske spørgsmål. ILTA - *International Language Testing Association* - gennemførte en undersøgelse af internationale prøvestandarder (1995) og udgav i 2000 sin *Code of Ethics* (2000). Dette dokument præsenterer en række principper med kommentarer som "trækker på moralfilosofi og tjener som en ledetråd for god professionel adfærd."

Retfærdighed?

I 2000 blev en publikation med oplæg, redigeret af Antony Kunnan, udgivet i serien *Cambridge Studies in Language Testing Series* baseret på den *19th Language Testing Research Colloquium* (Itrc Orlando 1997). Denne publikation med titlen *Fairness and validation in language assessment* fokuserer på retfærdighed inden for to områder: *Section One - Concept and Context* og *Section Two - Standards, Criteria and Bias*. Kunnan selv har udviklet en 'retfærdighedsmodel' som søger at integrere de traditionelle begreber validitet og reliabilitet med nyere begreber som fravær af skævvægtning (*absence of bias*), retfærdig adgang (*fair access*) og retfærdighed i administreringen (*fair administration*).

Mens ALTE's *Code of Practice* opridser sine generelle principper i 18 erklæringer^{[note 1](#)}, giver den kun ringe praktisk vejledning til dem der skal bruge dem, om hvordan disse principper skal implementeres, eller hvordan der kan opstilles standarder og garanteres for dem. I et forsøg på at tage fat på dette problem udviklede ALTE

Medlemmerne et supplerende dokument med titlen *Principles of Good Practice for ALTE examinations* (Principper for god praksis ved ALTE eksaminationer). Meningen med dette dokument var at det skulle give en mere detaljeret beskrivelse af hvordan principperne kunne omsættes til sprogprøvepraksisser som ALTE. Medlemmerne skulle indføre for at nå deres mål om høje professionelle standarder.

Inspirationen til hvordan man skulle opnå god praksis kom fra forskellige kilder, bl.a. Bachmans arbejde med *usefulness*, Messickas arbejde med validitet samt tidligere udgaver af *APA Standards* (f.eks. 1985). Selv om dokumentet ikke blev udgivet i sin helhed, blev dele af det senere inkluderet i *User's Guide for Examiners* (1997), udgivet af ALTE på Europarådets vegne (dette dokument har fået navneændring til *Language Examining and Test Development* og er under udgivelse af Europarådet som tillæg til *Common European Framework*, der skulle markere Det Europæiske Sprogår 2001).

I slutningen af 1990'erne havde en række andre projekter høj prioritet for ALTE, herunder arbejdet med ALTEs *Can Do Project*, der har udgjort det empiriske grundlag for tilknytningen af ALTE niveauerne til *Common European Framework*. Men i 1999/2000 rettede ALTE igen sin opmærksomhed mod *Code of Practice* og genetablerede en arbejdsgruppe der skulle bringe projektet videre.

I 2001 mødtes Arbejdsgruppen tre gange. Indtil nu er der kommet to vægtige resultater ud af det, som blev diskuteret af den samlede Medlemskreds på mødet i Budapest i november:

- En revideret udgave af *Principles of Good Practice for ALTE Examinations* (2001) er blevet udviklet og har cirkuleret i Medlemskredsen,
- *Code of Practice* er blevet omformet og udvidet som en tjekliste der bliver anvendt af Medlemskredsen som en del af et *Quality Management System* (QMS).

Den reviderede version af *Principles of Good Practice for ALTE Examinations* (2001) er baseret på tidligere udgaver, men er blevet ajourført og bearbejdet på flere punkter. Den beskæftiger sig mere detaljeret med de centrale spørgsmål om validitet og reliabilitet og ser på disse spørgsmål i relation til den påvirkning (*impact*) eksaminationer har på den enkelte og samfundet. Denne version trækker ligesom de tidligere udkast på de reviderede *era/apa/ncme Standards* (1999) - især i afsnittene om validitet og reliabilitet - såvel som på Bachman og Palmers arbejde, 1996.

Kvalitetsvurdering

Den nye dimension der er blevet indført i arbejdet på dette trin, er koncepterne *Total Quality Management* (TQM) og *Quality Management Systems* (QMS).

QM-systemer er i stort omfang blevet taget i brug både inden for produktions- og serviceområdet med henblik på at forbedre kvaliteten så 'kundens krav' kan imødekommes. Den samme tilgang er på det seneste blevet taget i brug i mange uddannelsesmæssige sammenhænge over hele Europa, herunder inden for nationale uddannelsessystemer og af internationale NGO'er. *Customer focus* er et fælles tema både i forretnings- og uddannelsesmæssige sammenhænge, og det samme gælder ALTE som eksamensudbyder. ALTE Medlemsskarens største 'kunder' er de kandidater der tager eksaminer (de primære brugere) og som kræver evalueringssystemer af høj kvalitet for at få en retfærdig behandling. Ved at anvende qm-tilgangen håber man at målsætningen om 'fortsat forbedring' vil bidrage til at sikre at standarderne fortsat højnes, og at 'kundernes' krav til eksaminerne fortsat imødekommes.

Anvendelsen af en QM-tilgang har ført til en bearbejdning af *Code of Practice* med henblik på at lade den afspejle de praktiske aspekter af evalueringsarbejdet inden for ALTE's Medlemskreds. Målet er at fastsætte brugbare procedurer og programmer for forbedring som i sidste ende vil kunne anvendes til at garantere minimumskvalitetsstandarder baseret på principperne i *Code of Practice*.

Anvendelsen af *Quality Management* er baseret på følgende nøglebegreber, som er taget fra litteraturen om QM:

organisationen

- selvevaluering og gensidig evaluering blandt elever/studerende på samme niveau (*Peer Monitoring*)

I denne tilgang er det vigtigt at identificere de roller og det ansvar som nøgle-interesserterne i ALTE-institutionerne har, og at anvende systemet på en fleksibel måde i overensstemmelse med hver organisations særtræk (dvs. de forskellige ALTE-Medlemmer og deres interessent-grupper). I projektets nuværende fase har Arbejdsgruppen fokuseret på de forskellige organisatoriske faktorer inden for ALTE-Medlemskredsen og på den række af forskelligheder der er. I bestræbelserne på at fastsætte standarder er det ikke målet at få alle ALTE-Medlemmer til at tilpasse sig den samme evalueringsmodel for alle de 24 sprog der er repræsenteret, og det er vigtigt at anerkende de forskellige sproglige, uddannelsesmæssige og kulturelle kontekster inden for hvilke eksaminerne bliver udviklet og anvendt. Det er nødvendigt med en passende balance mellem behovet for at garantere professionelle standarder over for brugerne og behovet for at tage højde for ALTE-organisationernes forskelligartede organisatoriske træk, og de sammenhænge hvori deres eksaminer anvendes.

Medlemmernes selvevaluering

Som det næste skridt anbefaler Arbejdsgruppen at alle Medlemmer identificerer deres nuværende stærke sider og de områder inden for deres egen organisation der trænger til øjeblikkelige forbedringer. På denne baggrund vil det så være muligt at opnå de ønskede resultater af kort- og langsigtede udviklinger. Målet bør være:

- at fastsætte et minimum for acceptable standarder
- at etablere *best practice* modeller
- at have fortløbende udvikling som mål (at bevæge sig mod *best practice*)

Der er sandsynligvis ikke noget ALTE-Medlem som vil opnå *best practice* inden for alle områder, og det vil altid være muligt at lave forbedringer. Målet for alle Medlemmer bør være at fortsætte med at udveksle ekspertise og gradvist hæve standarden over tid, dvs. at stræbe efter *best practice* modeller gennem en løbende udviklingsproces.

I et QM-system af denne slags bliver standarder ikke pålagt 'udefra', men iværksat gennem systemets måder af fungere på, og de procedurer der anvendes til at holde øje med standarder, er fra begyndelsen af baseret på bevidsthedsskabelse og selv-evaluering. Overvågning fra eksterne ligestillede kommer ind på et senere trin for at bekræfte at man lever op til minimumsstandarderne. I sin nuværende form er *Code of Practice* blevet bearbejdet så den kan fungere som et bevidsthedsskabende redskab i denne fase i projektet. Det bearbejdede format afspejler nu de fire aspekter i test-udviklingscyklussen (*test development cycle*), som alle ALTE-Medlemmer er bekendt med:

- Eksamensudvikling
- Administration af eksaminerne (herunder retning, karaktergivning og information om resultaterne)
- Analyse og efter-eksaminationsopsamling

Som opfølgning på mødet i Budapest i november 2001 blev det besluttet at alle Medlemmer skulle færdiggøre de procedurer der skulle indgå i de reviderede *Code of Practice* tjeklister, og sende dem til Arbejdsgruppen. I 2002 vil Arbejdsgruppen lave yderligere analyser og rapportere tilbage til Medlemskredsen i forbindelse med deres møder i april og november.

Note

1 ALTE's *Code of Practice* kan ses på: www.alte.org/about/code/index.cfm

Litteratur

Bachman L.F.: Fundamental Considerations in Language Testing. Oxford: OUP, 1990.

Bachman, L. F. and Palmer, A.: Language Testing in Practice. Oxford: OUP, 1996.

Code of Ethics. ILTA, 2000.

The **Code** of Fair Testing Practices in Education. Washington D.C. Joint Committee on Testing Practices, 1988.

Code of Practice. ALTE, 1994.

Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Council of Europe /CUP, 2001.

Handbook of Language Examinations and Examination Systems. ALTE, 1998.

Kunnan, A.: Fairness and Validation in Language Assessment. Cambridge: UCLES/CUP, 2000.

Linn, R.L. ed.: Educational Measurement. 3rd edition. New York: ACE/Macmillan, 1989.

Messick, S.A.: Test validity and the ethics of assessment. i: American Psychologist, 1980, 35.

Messick, S.A.: Validity, i: Linn, op.cit.

Milanovic, M. and Saville, N.: Principles of Good Practice for ALTE Examinations. 1991, 1993.

Standards for Educational and Psychological Testing. Washington: AERA/APA/NCME, 1985/99.

Task Force on Testing Standards. ILTA, 1995.

Users Guide for Examiners. ALTE/Council of Europe, 1997.

Oversat af Michael Svendsen Pedersen

Theme: Assessment - what should we know?



Quality and Fairness: the ALTE Code of Practice and Quality Management Systems

Nick Saville

**University of Cambridge. Local Examination Syndicate
(UCLES), UK.**

ALTE - the Association of Language Testers in Europe - was formed in 1990 when representatives of eight institutions attended an inaugural meeting in Barcelona. Each of these founder Members was an examination provider involved in the assessment of their own language as a foreign language. At that first meeting it became clear that they all shared a common concern for high standards in language assessment and for fair treatment of the candidates who take their exams. These shared concerns have been a unifying factor and have provided a focus for working together in the past decade. As early as 1991 it was agreed that a key objective of the Association should be:

" to establish professional standards for all stages of the language-testing process,"
the aim being to provide language learners with access to high quality examinations which would help them to meet their life goals and which would have a positive impact on their learning experience.

Since 1990 the Members of ALTE have been meeting regularly (twice a year) and have collaborated on a range of projects (some funded by the EU's Lingua Bureau). In recent years the work of ALTE has become well known in Europe and the membership has grown steadily. There are now 27 institutional Members and associates, representing 24 European languages (including the widely spoken languages, such as English, French, German and Spanish, as well as many less-widely spoken languages, including those of Scandinavia and central Europe). The Members now include: Ministries of Education and other government departments, examination boards, cultural bodies, university departments and language teaching organisations. In some cases, language testing consortia have been formed within ALTE, as for example, in Denmark for the testing of Danish, where the Ministry of Education now collaborates with Copenhagen University and the Studeskolerne. Some ALTE Members have been involved in language assessment for many years, while others are relative newcomers, or in some cases, are just starting out as examination providers. Despite the wide diversity which now exists within the membership, concern for quality and fairness continues to be a priority, as well as a focus on professional standards.

Without reference to technical jargon, most people have some idea of what fairness might mean in language examinations; for example, it is important for all candidates to be treated equally and for the results they achieve to be a fair reflection of what has been taught and learned. Many would also agree that the outcome, in terms of

a certificate awarded, should be useful as a qualification and have a recognised value for study, work or other purposes, (preferably both at home and in other countries).

While fairness has always been of concern in educational contexts, in recent years, the concept of fairness has also received particular attention by testing professionals, alongside familiar psychometric concepts such as validity and reliability. Varying views of what fairness means have emerged, but typically these reflect the layman's concern for the just treatment of groups and individuals in society and relate to issues such as access, equity, lack of bias, trustworthiness, and dependability.

Within ALTE, the founder Members agreed that it was important for both examination developers and examination users to follow an established code of practice which would ensure that the assessment procedures are of high quality and that all stakeholders are treated fairly. A code of practice of this kind had to be based on sound principles of good practice in assessment which allow high standards of quality and fairness to be achieved. The discussion of what constitutes good practice within ALTE has continued since then and reflects a concern for accountability in all areas of assessment which are undertaken by the ALTE Members. In this respect it recognizes the importance of validation and the role of research and development in examination processes. Over the years ALTE has provided a forum for this work to be carried on.

In publishing its first Code of Practice in 1994, ALTE set out the standards that Members of the association aimed to meet in producing their language exams. It drew on The Code of Fair Testing Practices in Education produced by the Washington D.C. Joint Committee on Testing Practices (1988) and was intended to be a broad statement of what the users of the ALTE examinations should expect and the roles and responsibilities of stakeholders in striving for fairness.

The Code of Practice identifies three major groups of stakeholders in the testing process:

- the examination developers, (i.e. examination boards and other institutions which are the Members of ALTE)
- the primary users - the candidates, who take the examination by choice, direction or necessity
- the secondary users, who are the sponsors of the candidates or who require the examination for some decision-making or other purpose (e.g. parents, teachers, employers, admissions officers etc.)

In addition the Code of Practice lays down responsibilities of these stakeholder groups in four broad areas:

- developing examinations
- interpreting examination results
- striving for fairness
- informing examination takers

An important feature is that it emphasises the joint responsibility of the stakeholders and focuses on the responsibilities of examination users as well as the examination developers in striving for fairness.

In North America, in the fields of psychological and educational assessment, there has been a long tradition of setting standards, as reflected in the 1999 edition of the Standards for educational and psychological testing (AERA/APA/NCME). Previous editions of these Standards have been around since the 1950s and have had an influence on language test developers and test users, especially in the USA. They are regularly referred to in the language testing literature (see for example Bachman 1990 and Bachman and Palmer 1996).

The latest edition includes several new sections which were not in the earlier version (1985), including Part II of the volume which is entitled Fairness in Testing. The sub-sections in this part cover: fairness in testing and test use, the rights (and responsibilities) of test takers, and the testing of individuals from different backgrounds or with disabilities.

More specifically in the field of language assessment there has been a growing interest in this topic, including the development of codes of practice and an interest in ethical concerns. ILTA - the International Language

Testing Association conducted a review of international testing standards (1995) and in 2000 published its Code of Ethics (2000); this document presents a set of nine principles with annotations which "draws upon moral philosophy and serves to guide good professional conduct."

Justice?

In 2000, a volume of papers edited by Antony Kunnan was published in the Cambridge Studies in Language Testing Series based on the 19th Language Testing Research Colloquium (LTRC Orlando 1997). This volume, entitled Fairness and validation in language assessment, focuses on fairness in two sections: Section One - Concept and Context, and Section Two - Standards, Criteria and Bias. Kunnan himself has been developing a fairness model" which seeks to integrate the traditional concepts of validity and reliability with newer concepts such as absence of bias, fair access and fair administration. His recent work in this area was presented at the ALTE Conference in Budapest (November 2001) and formed part of a debate within ALTE on fairness and the ALTE Code of Practice.

While the ALTE Code of Practice outlines its broad principles in 18 general statements, it provides little practical guidance to the practitioner on the implementation of these principles or how standards can be set and guaranteed. In attempting to address this issue, the ALTE Members produced a supplementary document entitled Principles of Good Practice for ALTE Examinations (drafted by Milanovic & Saville) which was discussed at ALTE meetings in 1992 and 1993 (Alcalà de Henares, Paris and Munich). This document was intended to set out in more detail what the principles might mean in terms of language testing practices which ALTE Members should adopt in order to achieve their goal of high professional standards.

The approach to achieving good practice was influenced by a number of sources from within the ALTE Membership (e.g. test development work being carried out by UCLES in Cambridge) and from the field of assessment at large. This included the work of Bachman on "usefulness", Messick on validity, and earlier versions of the APA Standards (e.g. 1985). ALTE Members sought feedback on the document from external experts in the field (Jan-Mar 1994) and it was discussed again in Arnhem 1994. While it was not published in its entirety, parts of the document were later incorporated into the Users Guide for Examiners (1997) produced by ALTE on behalf of the Council of Europe (this document has been renamed Language Examining and Test Development and is being published by the Council of Europe to accompany the Common European Framework, marking the European Year of Languages 2001.)

In the late 1990s a number of other projects were a high priority for ALTE, including the work on the ALTE Can Do Project which has provided an empirical basis for linking the ALTE levels to the Common European Framework. But in 1999/2000, ALTE turned its attention back to the question, re-establishing a Code of Practice Working group to take this project forward. ALTE Members were keen that the Working Group should represent the different perspectives of the expanded ALTE membership and the varying historical, cultural and linguistic influences. To this end, members of the Working Group were selected from ALTE Member organisations providing representation of a range of types of institution, including two examination boards, a university department, a ministry of education and a cultural body. They also represent widely and less-widely spoken languages and have varying degrees of experience in conducting language assessment on an international scale.

In 2001 the Working Group met three times in January, May and October and reported on progress at the full ALTE meetings in Perugia (May 2001) and Budapest (November 2001). So far two substantive outcomes have been achieved and these were discussed by the full Membership at the Budapest meeting:

- a revised version of Principles of Good Practice for ALTE Examinations (2001) has been produced and circulated to Members;
- the Code of Practice itself has been re-designed and expanded as a checklist which is being used by Members as part of a Quality Management System (QMS).

The revised version of Principles of Good Practice for ALTE Examinations (2001) is based on the earlier version but has been updated and reworked in many parts. It addresses in more detail the central issues of validity and

reliability and looks at the related issues surrounding the impact of examinations on individuals and on society. This version, like the earlier drafts, draws on the revised AERA/APA/NCME Standards (1999) - especially in the sections on validity and reliability - as well as the work of Bachman and Palmer, 1996.

The new dimension which has been introduced into the work at this stage is the concept of Total Quality Management (TQM) and Quality Management Systems QMS.

QM systems have been widely adopted in both manufacturing and service industries with the aim of improving quality in order to "meet customer requirements". More recently the same approach has been adopted in many educational contexts across Europe, including within national education systems and by international, non-governmental organisations. "Customer focus" is a common theme in both the business and educational contexts and this is the same for ALTE as examination providers. The main "customers" of the ALTE Members are the candidates who take the examinations (primary users) and who require high quality assessment systems in order to be treated fairly. By adopting the QM approach, it is hoped that the goal of "continuous improvement" will help to ensure that standards continue to rise and the requirements of the "customers" for the exams can always be met.

The adoption of a QM approach has led to the reworking of the Code of Practice in order to reflect the practical aspects of assessment work within the ALTE membership. The aim is to establish workable procedures and programmes of improvement which ultimately will be able to guarantee minimum quality standards based on the principles in the Code of Practice.

The approach to Quality Management which is being implemented is based on the following key concepts taken from the literature on QM:

- the organisation
- self-assessment and peer monitoring

In this approach, it is important to identify the roles and responsibilities of key stakeholders in the ALTE institutions and to apply the system with flexibility according to the specific features of each organisation (i.e. the different ALTE Members and their stakeholder groups). In the current phase of the project, the Working Group has been focusing on the different organisational factors within the ALTE Membership and on the range of diversity that exists. In seeking to establish standards it is not the aim to make all ALTE Members conform to the same models of assessment for all 24 languages represented, and it is important to recognise the varied linguistic, educational and cultural contexts within which the examinations are being developed and used. An appropriate balance is required between the need to guarantee professional standards to users, and the need to take into account the differing organisational features of the ALTE institutions and the contexts in which their exams are used.

As a next step, the Working Group recommended that all Members should attempt to identify their current strengths and the areas in need of immediate improvement within their own organisation. On this basis, it will then be possible to establish the desired outcomes for both short- and long-term developments. The aim should be:

- to set minimum acceptable standards
- to establish "best practice" models
- to aim at continuous improvement (move towards best practice)

It is likely that no ALTE member will meet best practice in all areas and improvements will always be possible. The aim for all Members should be to continue to share expertise and gradually to raise standards over time, i.e. to aim at the best practice models through an on-going process of development.

In a QM system of this kind, standards are not imposed from outside" but are established through the mechanism of the system itself and the procedures to monitor standards are based on awareness raising and self-assessment in the first instance. External (peer) monitoring is introduced at a later stage to confirm that the

minimum standards are being met. In its current form the Code of Practice has been reworked to function as an awareness raising tool at this stage of the project. The re-designed format now reflects the four aspects of the test development cycle with which all ALTE Members are familiar:

- Examination development
- Administration of the examinations
- Processing of the examinations (including the marking, grading and issue of results)
- Analysis and post-examination review

Following the meeting in Budapest in November 2001, it was agreed that all Members would complete the procedures to fill in the revised Code of Practice checklists and will return these to the Working Group. In 2002 the Working Group will carry out further analysis and will report back to the membership at their meetings in April and November.

References

- AERA/APA/NCME** (1985/1999). Standards for educational and psychological testing. Washington: AERA.
- ALTE** (1994) Code of Practice
- ALTE/Council of Europe** 1997) Users Guide for Examiners
- ALTE** (1998) Handbook of Language Examinations and Examination Systems
- Bachman L.F.** (1990). Fundamental considerations in language testing. Oxford: OUP.
- Bachman, L. F.** and Palmer, A. (1996). Language Testing in Practice. Oxford: OUP.
- Council of Europe** (2001) Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: CUP
- ILTA** (1995) Task Force on Testing Standards
- ILTA** (2000) Code of Ethics
- Kunnan, A.** (2000) Fairness and validation in language assessment. Cambridge: UCLES/CUP
- Linn, R.L.** ed. (1989) Educational Measurement. Third edition. New York: ACE/Macmillan
- Messick, S.A.** (1980) Test validity and the ethics of assessment. American Psychologist, 35.
- Messick, S.A.** (1989) Validity, in Linn 1989
- Milanovic, M.** and Saville, N. (1991, 1993), Principles of Good Practice for ALTE Examinations
- Washington D.C.** Joint Committee on Testing Practices (1988) The Code of Fair Testing Practices in Education

Tema: Evaluering - hvad skal vi kunne?



Evaluering og prøver i projektorganiserede sprogstudier

Karen Sonne Jakobsen
Lektor, cand.mag. (tysk).
Roskilde Universitetscenter.

Evaluering spiller en central rolle i projektorganiserede studier, både i deres selvforståelse og i praksis, som den har udviklet sig ved Roskilde Universitetscenter og Aalborg Universitet. Denne artikel skal på baggrund af erfaringer fra RUC handle om projektevalueringer og argumentere for at principper fra denne evalueringsform med fordel kan anvendes også ved evaluering af sprogfagenes færdighedsside.

Evaluering er på RUC et begreb der dækker både over den *interne evaluering* og over *prøver*. De to evalueringsformer har hver deres formål: Formålet med prøverne er en bedømmelse af om den studerendes kvalifikationer svarer til de mål og krav der er fastlagt for den pågældende uddannelse eller uddannelsesdel. Der er typisk tale om en ekstern prøve, dvs. med censur, og bedømmelsens resultat udtrykkes gennem anvendelse af 13-skalaen. Prøverne er reguleret af eksamensbekendtgørelsen for de videregående uddannelser. Formålet med de interne evalueringer er at de studerende lærer *selv* at bedømme deres arbejdsresultater og arbejdsproces. Den interne evaluering har altså primært et studiemæssigt sigte, idet den skal bidrage til at opfylde den del af formålet der handler om at udvikle gode arbejdsformer, selvstændighed og kritisk sans. De interne evalueringer er reguleret af studieordninger og holdes i hævd af en evalueringstradition der især er forankret i RUCs basisuddannelser.

Interne evalueringer

Interne evalueringer optager en betydelig del af studie- og vejledningstiden på den humanistiske basisuddannelse, som går forud for de fagspecifikke studier på overbygningen. Det har den begrundelse at det er i de første studieår studiekompetencen skal udvikles, og erfaringen er at det kan betale sig. Når de studerende kommer på overbygningen, har de lært at gennemføre et projektarbejde, og vejlederen kan her i praksis koncentrere sig om den faglige side af vejledningen.

Den interne evaluering er organiseret i 3 faser:

Den første er *problemformuleringsseminaret*, hvor de enkelte projektgrupper fremlægger og diskuterer deres problemformuleringer. Er den problemstilling der skal behandles, klar nok? Hvilke afgrænsninger er relevante? Hvilke teorier og hvilket empirisk materiale skal der arbejdes med? Er projektet gennemførligt på den givne tid?

- det er typisk sådanne spørgsmål der diskuteres.

På *midtvejsseminaret* fremlægges en plan over projektet som helhed, en projektstatus og evt. udvalgte arbejdspapirer.

Størst vægt har den *afsluttende interne evaluering* der foregår på basis af den færdige projektrapport. I denne evaluering deltager projektgruppen, en opponentgruppe, vejlederen og en opponentvejleder. Andre, også udefra, kan inviteres til at deltage, f.eks. hvis gruppen har været i kontakt med eksterne parter. Diskussionerne drejer sig ikke kun om det faglige indhold, men også om arbejdsprocessen og de vanskeligheder der har været undervejs, og om det lykkes gruppen at tilrettelægge og bruge evalueringen konstruktivt. Konklusionen skal være fremadrettet, nemlig hvad gruppen har lært med henblik på senere projekter.

Den interne evaluering omfatter også projektarbejdets sproglige og kommunikative sider, i princippet uanset om arbejdssproget er dansk eller et andet sprog. En vurdering af rapportens sproglige og tekstlige kvalitet og af de studerendes evne til at strukturere situationen, til mundtligt at fremlægge deres resultater og synspunkter og til at indgå i en dialog, bør være en naturlig del af den interne evaluering.

Evalueringer og prøver i sprogfagene

Selv om intern evaluering og prøvning tjener forskellige formål, er der alligevel en sammenhæng. I formen ligner de hinanden, og gennem de interne evalueringer lærer de studerende at omgås prøver på en fagligt produktiv måde.

En formel beskrivelse af prøveformen som her er hentet fra studieordningen for faget tysk, lyder:

Mundtlig prøve med udgangspunkt i projektrapport. Eksaminationen tager udgangspunkt i den udarbejdede projektrapport. Den enkelte studerende redegør først for dele af projektet han eller hun særligt har beskæftiget sig med. Redegørelsen danner udgangspunkt for en samtale mellem den studerende, eksaminator(er) og censor. Der kan stilles spørgsmål i hele modulets indhold.

Med udgangspunkt i denne formulering vil jeg komme ind på nogle af projektprøvens kvaliteter, problemer og potentialer i et sprogfag.

Samtalesituationen

Ved projektprøven har eksamensspørgsmål gennemgående en lidt anden karakter end ved den traditionelle mundtlige eksamen. Det hænger sammen med at de studerende fremlægger et stykke fagligt arbejde i form af en rapport, som de står til ansvar for, omtrent som ved den interne evaluering. Eksamensspørgsmål stilles ikke til et opgivet pensum, men til de studerende og deres arbejde: Hvorfor de har valgt den givne problemstilling, om dens faglige og praktiske relevans, valg af teorier og metoder, af empirisk materiale, analyseresultater, konklusioner etc. Som man kan se af den sidste sætning i prøvebestemmelsen, er der også mulighed for at stille de klassiske eksamensspørgsmål. Men jeg vil tro at hovedparten af de spørgsmål der stilles ved projektprøven, er 'ægte' spørgsmål i den forstand at spørgeren ikke kender svaret i forvejen. Det er dog stadig eksamensspørgsmål, idet formålet med samtalen ikke er at blive klogere - det er i så fald en sidegevinst! - men en vurdering af de studerendes standpunkt.

Måske kan man karakterisere disse spørgsmål som 'ikke-trivielle eksamensspørgsmål'. En nøjere analyse af samtaler i forskellige mundtlige prøveformer kunne give nyttig viden om hvordan forskellige prøveformer fungerer i praksis, hvilke prøveformer der resulterer i de mest varierede samtaler, og dermed giver den bredeste basis for en prøvning af de studerendes kvalifikationer, herunder af deres kommunikative kompetencer.

Gruppen og den enkelte

Karakteristisk for projektprøven er at den har flere aktive deltagere, nemlig projektgruppen, eksaminator og

censor. Som det fremgår af bestemmelserne forventes censor at deltage aktivt. Flere deltagere har den effekt at prøven undgår at få karakter af en overhøring med censor på sidelinjen, og mere ligner en ægte faglig samtale og diskussion. Censors bidrag er væsentligt idet det ofte er hende der kommer med nye vinkler på projektet.

En projektprøve er altså en gruppeprøve, men også ved gruppeprøver skal der ifølge reglerne ske en individuel prøvning af den enkelte studerende. Det betyder at de studerende har en dobbelt opgave under prøven: De kan - lige som ved de interne evalueringer - ikke nøjes med at koncentrere sig om deres egne bidrag, de skal også sørge for at gruppen fungerer optimalt, herunder at alle kommer til orde i passende omfang og dermed leverer grundlag for en individuel bedømmelse. Ansvar for at alle kan bedømmes, ligger dog hos eksaminator og censor, som om nødvendigt må gribe ind og stille direkte spørgsmål til gruppemedlemmer der ikke er aktive nok.

Spørgsmålet om individuel bedømmelse er denne prøveforms måske vanskeligste punkt og med mellemrum genstand for kritik. I en nylig evaluering af basisuddannelserne gennemført ved Evalueringsinstituttet anbefales det at gruppeprøverne nedtones, og at der indføres flere 'individuelle elementer' i prøveformerne. Med et moderat antal gruppemedlemmer, med den rigelige tid der er til rådighed, og med erfarne eksaminatorer og censorer er det dog i praksis ikke noget problem at vurdere den enkelte studerende. Resultatet af vurderingen er som oftest at gruppen får den samme karakter, dvs. at de dygtigste i gruppen får en lidt lavere, de svageste en lidt højere karakter end de ville have fået individuelt - måske! Denne vurderingspraksis tilgodeser gruppearbejdet, idet en højere grad af individuel differentiering utvivlsomt ville have en uheldig *wash back-effekt* på samarbejdet i projektgruppen.

Man kunne tydeliggøre og legitimere denne praksis ved at lade gruppens samarbejdsevne, som bl. a. ytrer sig gennem deres indbyrdes kommunikation, tælle som en del af karakteren. Det er værd at huske at det langt fra er givet at den blændende dygtige studerende havde været i stand til at gennemføre projektet alene, og at de svagere studerende meget vel kan have en stor andel i et vellykket projekt. Mellem individuel udfoldelse og samarbejde i gruppen er der ingen modsætning, men et komplekst samspil.

Prøvning af kommunikative kompetencer

Projektprøven er altså en kompleks og ganske krævende kommunikationsform. Det er nærliggende at udnytte den til prøvning af kommunikative kompetencer.

I praksis har det vist sig at *mundtlig sprogfærdighed* med fordel kan integreres i projektprøven. Projektrapporter i sprogfagene skrives i stigende grad på fremmedsproget og danner grundlag for en prøve på fremmedsproget både af indhold og af den mundtlige sprogfærdighed. De studerende oplever det som en mere tilfredsstillende prøveform at skulle bruge sproget til faglige formål, frem for den besynderlige konversation om en tilfældig tekst som de ellers udsættes for. Der er næppe tvivl om at der med en projektbaseret prøve stilles mere relevante og varierede sproglige krav, og at der dermed skabes et bredere og bedre bedømmelsesgrundlag. De studerendes evne til at tage og give ordet, til at strukturere deres samtalebidrag, til at udtrykke uenigheder på en høflig måde osv. kan indgå i bedømmelsen. Vurderet som en prøve i kommunikativ kompetence har den den klare styrke at kommunikationssituationen ikke er fiktiv - hvad den ellers ofte er i de kommunikative prøveformer - men i høj grad real. Det giver et mere realistisk og pålideligt grundlag for bedømmelsen af den kommunikative kompetence.

Prøveformen for *skriftlig sprogfærdighed* har derimod ikke ændret sig så meget. Prøven består stadig i at de studerende skal producere en tekst på en begrænset tid under overvågning og under anvendelse af godkendte hjælpemidler. Denne prøveform er imidlertid under afvikling af informationsteknologien. Det er i stigende grad besværligt og mindre legitimt at afskære folk fra information, også i prøvesituationer. På RUC har vi valgt det, sikkert midlertidige, kompromis at alle hjælpemidler er tilladt, men ikke kommunikation via internet eller på anden vis. Kontrollen indskrænkes altså til at sikre at opgaven besvares af den person der har meldt sig til prøven.

Der er flere grunde til at overveje alternativer - ud over de informationsteknologiske udfordringer også det faktum at denne prøveform er meget lidt elsket af de studerende. Det er et godt gæt at de skriftlige prøver er

upopulære fordi de studerende føler at den ikke giver dem tilstrækkelige chancer for at vise hvad de faktisk kan. Prøvesituationen opleves som akavet og som urimeligt svær at opnå gode karakterer i. Karakteren i skriftlig sprogfærdighed ligger som regel lavt sammenholdt med den studerendes niveau i øvrigt. Det tyder på at der er noget galt med formen. Men især er der grund til at tage prøveformen op til revision fordi den har en begrænsende effekt på undervisningen. De senere års udvikling i skrivepædagogikken bliver ikke understøttet af denne prøveform, tværtimod. Det kan ikke undgås at undervisningen tenderer til at begrænse sig til at indøve de genrer der optræder under de givne prøvebetingelser.

Spørgsmålet er om ikke hovedprincippet fra projektprøven - at de studerende fremlægger et skriftligt produkt med henblik på bedømmelse - også kunne bruges til prøvning af skriftlig sprogfærdighed. Eller med andre ord: En *portfoliobaseret* prøve fortjener at blive afprøvet som alternativ til de traditionelle skriftlige prøver. Bekymringen for om kontrollen er tilstrækkelig effektiv, er nok det der indtil videre har forhindret en fornyelse af prøveformen. Erfaringerne fra projektprøven siger at det ikke er vanskeligt at konstatere om de studerende faktisk er ophav til den fremlagte rapport. Jeg deler Tim Cauderys og John Irons opfattelse (i Sprogforum nr. 11) at en kombination af portfolio med mundtlig prøve kunne blive en god fornyelse af prøven i skriftlig sprogfærdighed.

Ved portfolioprøven fremlægger de studerende et eller flere skriftlige produkter som grundlag for en samtale med eksaminator og censor. Det ville give en langt større genremæssig frihed end den der hersker nu. Der kunne principielt være mulighed for fremlæggelse af både individuelle og af større kollektive skriftlige produkter. Prøveformen vil kræve af studerende, eksaminator og censor at de går ind i en evaluerende samtale om de skriftlige produkter og dermed at viden om tekster og tekstproduktion indgår i bedømmelsen, og at vurderingskriterier ekspliciteres og begrundes. Dette ville berige prøven i skriftlig sprogfærdighed med et nyt indhold som giver mening, set ud fra den studerendes - tekstproducentens - perspektiv.

Litteratur

Caudery, Tim: Portfoliobedømmelse - en mulighed i Danmark ? i: [Sprogforum 11, 1998.](#)

Irons, John: "Det var så deiligt ude på Landet.." i: [Sprogforum 11, 1998.](#)

Studieordning for den humanistiske basisuddannelse, Roskilde Universitetscenter 1996.

Studieordning for tysk, Roskilde Universitetscenter 1996.

Basisuddannelserne ved Aalborg Universitet og Roskilde Universitetscenter, Evalueringsinstituttet 2001.

Tema: Evaluering - hvad skal vi kunne?



Genren hedder eksamensopgave Om den skriftlige prøve i engelsk til lærereksamen

Vibeke Toft Jørgensen
Adjunkt, cand.mag.
Hjørring Seminarium.

Når vi bedømmer en skriftlig eksamen i engelsk, er et af kriterierne genrebevidsthed. Men hvor genrebevidste er vi selv?

Det er min opfattelse, at vi som lærere kan have en tendens til at glemme, at de studerende ikke først og fremmest skriver breve til udenlandske studerende, taler til et amerikansk publikum, eller tager referater af danske tekster til engelsksprogede personer. De skriver først og fremmest til en lærer og en censor, og genren hedder eksamensopgave.

Som lærere kan vi lade som om, det ikke er eksamensopgaver, vi har med at gøre. Vi kan være kreative og forestille os en 'virkelig' situation, som man skal 'skrive sig ind i', og vi kan undre os over, hvor lidt genrebevidste og modtagerorienterede nogle af vores studerende er. Samtidig omtaler vi naturligvis brevene, talerne og referaterne som 'opgaver', vi skal 'have rettet'. Dagligsproget taler her sit tydelige sprog om, hvad det (også) er for en genre, vi har for os, og hvilke interesser vi som læsere har: Vi læser disse tekster for at rette og give karakter.

Modsat lærerne vil de studerende sjældent kunne abstrahere fra eksamenssituationen. Som studerende skal man rette blikket to steder hen: Man skal tænke på den fiktive modtagergruppe, og man skal tænke på den faktiske modtagergruppe (lærer og censor).

Det er min påstand, at nogle studerende bliver skeløjede i denne situation, og artiklen handler derfor om genrebevidsthed og de 'skeløjede' tekster.

Jegpræsenterer kort den nye skriftlige lærereksamen i engelsk, påpeger problemer og foreslår mulige løsninger. Endelig diskuterer jeg kreativitetens vilkår i et eksamenssystem.

Ny skriftlig eksamen i engelsk

Den nye skriftlige lærereksamen i engelsk blev indført med 1997-loven og afløste de decentralt stillede skriftlige

opgaver. Den skriftlige karakter fik selvstændig status i modsætning til før, hvor den skriftlige og den mundtlige eksamen i engelsk udløste én karakter.

Det er tydeligt, at kriterierne afspejler mange års arbejde med ny skrivepædagogik, og det er glimrende, at kriterierne lægges åbent frem. I praksis kan bedømmelseskriterierne imidlertid opleves som indbyrdes modstridende.

Sammenstød mellem gamle og nye eksamenskrav

De følgende eksempler stammer fra mine egne studerendes besvarelser til eksamen 2000, hvor tekster og spørgsmål handlede om kropskultur.

Prøven består af tre typer opgaver:

- Opgave A: en ræsonnerende/argumenterende tekst.
- Opgave B: en kreativ tekst.
- Opgave C: en transformation/oversættelse.

I opgave A skal der skrives om *Body Culture. The world outside the classroom and its implications for us as teachers* til en gruppe amerikanske studerende og med inddragelse af tekstkompendiet. En studerende fandt det tilsyneladende svært at kommentere de amerikanske og engelske tekster. Altså omtalte hun kort tekstsamlingen og skrev derefter om danske forhold, bl.a. den alternative modebølge med forsøg på at lancere tøj i store størrelser under engelske navne som *2big*. Egentlig ikke dårligt tænkt i forhold til modtagergruppe 1, amerikanerne, som måske ville kende kompendiets tekster og baggrunden for dem bedre end den studerende. Men det var forkert i forhold til modtagergruppe 2 (censor og undertegnede). Vi ville have tekstsamlingen kommenteret og fokus på problematikken: 'verden i og uden for skolen' foruden en klar bevidsthed om, hvem teksten var skrevet til. Vi fik ikke helt, hvad vi ville have, måske fordi kravet om tekstkommentar (et traditionelt eksamenskrav) og kravet om modtagerrelevans (et nyt eksamenskrav) var svære at forene.

Opgave B tager udgangspunkt i et uddrag af Douglas Couplands roman: *Miss Wyoming*. Man skal skrive en tale med titlen *The Quality of Life*, som man kunne forestille sig Susan Colgate holde, efter hun har sagt farvel til livet som skønhedsdronning. En studerende skrev Miss Wyomings tale i formfuldendt sprog, med brug af gennemtænkt billedsprog. Lærer og censor følte sig imidlertid ikke overbevist om, at Couplands person ville udtrykke sig i smukt sammenhængende billedsprog og korrekt brug af infinitte sætninger. Denne studerendes opgave eksemplificerer dilemmaet mellem det traditionelle eksamenskrav om korrekt sprog og det nye krav om sprog tilpasset situation og person.

Jeg vil ikke gå udførligt ind i flere eksempler, men nøjes med at konkludere, at man i de studerendes tekster kan se problemer, som i hvert fald delvis kan forklares ud fra eksamenssituationen og opgaveformuleringerne. Problemet gentog sig i år, hvor en af mulighederne i B-opgaven var at skrive en logbogsside. En logbog er pr. definition en genre, der ikke skal rettes, så den kan være svær at skrive som eksamensopgave (og svær at bedømme).

Hvad jeg har kaldt skeløjede tekster produceres, fordi vi gerne vil have det hele. De studerende skal helst kunne opfylde både traditionelle krav om tekstkommentar og korrekthed og nye krav om modtagerrelevans og situationsafhængigt sprog. Disse krav er vanskeligt forenelige og kan opleves som uforenelige.

Kan kravene til opgaverne gøres tydeligere?

I opgave A skal de studerende skrive en to-siders sagprosatext på grundlag af et stort tekstkompendium. Bindningen til de mange tekster gør det svært. Skal man kommentere så mange af teksterne som muligt, med fare for at blive opremsende, eller skal man koncentrere sig om nogle få, med fare for ikke at dække stoffet? Problemet kunne løses med en mindre tekstsamling eller ved en klar angivelse af, hvor lidt eller hvor meget i kompendiet, man skal forholde sig til.

Bindingen til en bestemt modtagergruppe kræver en situation med en troværdig informationskløft. Ellers bliver der tale om en kunstig binding, der gør det sværere at skrive.

Hvis man vælger at udelade modtagergruppen, kan man slet og ret bede de studerende skrive en artikel eller et essay om kropskultur, om samarbejdet mellem lærere og forældre, om at være ny lærer eller hvad temaet nu er, med udgangspunkt i to eller flere af teksterne. Vi kan så bedømme, om den studerende kan finde et fokus, strukturere sin tekst, tage kritisk stilling til teksterne og argumentere på godt engelsk.

Opgavekravet ville så være ret traditionelt. Det nye består i arbejdet bag kulisserne. I undervisningen har de studerende arbejdet intensivt med skriveprocessen, så mange studerende kan skrive gode sagprosaetekster med et klart fokus, især hvis de får lov til at arbejde inden for ikke alt for snævre rammer.

B-opgaven er en kreativ opgave, som typisk lægger op til meddigtning. En meddigtende opgave giver mulighed for at bedømme, om de studerende har læst den skønlitterære tekst og forstået den, og om de kan sætte sig ind i og gengive hovedpersonens tanker. Opgaven kan især i eksamenssituationer have den svaghed, at der bliver en konflikt mellem kreativitet og korrekthed, der kan gøre den svær at vurdere. Dette kan opgavekommissionen have med i overvejelserne, når teksterne vælges, eller man kan præcisere, i hvor høj grad der i bedømmelsen indgår valg af stilleje.

I opgave C skal man typisk referere en dansk tekst eller oversætte et uddrag af en dansk tekst.

Der har i prøvesæt til skriftlig eksamen også været andre opgavetyper på banen, f.eks. en transformation, hvor en tekst omskrives til en læzerskare med begrænsede sprogkundskaber. Af disse tre muligheder foretrækker jeg referatet. Oversættelsesarbejde er spændende og krævende, men bruges ikke i særlig høj grad i folkeskolen. En omskrivning af en tekst fra et svært til et lettere niveau finder jeg vanskelig at bedømme, da både de dygtige og de svage studerende ender med en forholdsvis simpel tekst.

At referere en dansk tekst til engelsksprogede læsere er en troværdig kommunikationssituation, og opgavetypen vil være relevant i mange sammenhænge. Samtidig giver opgaven - som en god eksamensopgave skal - muligheder både for den dygtige og den svage skribent, da man kan vælge sit eget niveau. Man kan udelade et krav om en bestemt modtagergruppe, da en sådan alligevel sjældent vil have afgørende indflydelse på referatets udformning.

Jeg mener ikke, man skal være bange for traditionelt formulerede opgaver, f.eks. et essay eller et referat. Man kan bruge processkrivning til at løse traditionelle opgaver kreativt - hvilket er noget andet end at arbejde med kreative opgaver i betydningen meddigtning, synsvinkelskift, genreskift eller lignende. De kreative opgaver kan til gengæld blive banale, især hvis man ikke forholder sig kritisk til den udbredte opfattelse, at alle kreative opgaver er gode. Både i folkeskolen og i læreruddannelsen har jeg set opgaver, hvor der skal skrives kreativt og meddigtende på virkelige personers historie, evt. ud fra et billede af dem, og man kan med en sådan opgavetype invitere til noget, der ligner formiddagspressens journalistik. Den slags kreative opgaver kan jeg godt undvære.

Både sagprosaen, fiktionen og den sproglige præcisionsopgave er relevante teksttyper for en kommende lærer. B-opgaven markerer den største fornyelse. At man ikke blot skal læse fiktion, men skrive med selv, afspejler udviklingen inden for de senere års litteratur- og skrivepædagogik. Det er imidlertid også den kreative opgave, der volder de største problemer, når den skal indpasses i et eksamenssystem.

Kreativitetens vilkår i et eksamenssystem

Vi får under ingen omstændigheder en fri kreativ opgave i en eksamenssituation. Vi siger måske kreativitet, men giver ordet en ny betydning, fordi kreativiteten, hvad enten vi kan lide det eller ej, optræder i en eksamenssituation.

Eksamensopgavens grundvilkår er, at den er samfundets kontrol med målbare kundskaber og færdigheder, altså en genre, der pr. definition er underlagt kontrol. At kontrollen til dels defineres og udøves af lærerne selv, gør

det måske netop sværere at få øje på dens omfang og betydning.

Inden for læreruddannelsen har vi vilkår, der kan gøre det svært at sikre en ensartet bedømmelse af de skriftlige opgaver. Der er ikke fra ministeriets side afsat ressourcer til at udarbejde vejledninger til skriftlig lærereksamen med eksempler på opgaver; der er ikke nogen central studielektor, og endelig er læreren selv med til at rette de studerendes opgaver (modsat f.eks. gymnasiet, hvor skriftlige opgaver rettes af to uvildige censorer).

Disse vilkår må sætte ikke mindst de udefrakommende censorer på en hård prøve, for hvem spørger man, og hvad henviser man til, hvis man er uenig med læreren? Nærlæser man de bedømmelseskriterier, der står på selve opgaven, er det tydeligt, at de kan tolkes ret forskelligt, ligesom man kan tolke deres indbyrdes vægtning. Derfor er det nødvendigt at have eksempler at forholde sig til.

Sidst men ikke mindst: De studerende kan opleve, at deres skriftlige arbejde ikke bedømmes ensartet, og at de bliver gidsler i et spil om traditionelle og nye eksamenskrav, der endnu ikke er blevet afgjort.

At kontrollen og kvalitetssikringen i den grad placeres hos lærere og censorer er betænkeligt. Måske får vi ligefrem en blind plet, der får os til at glemme grundlæggende forhold vedrørende eksamen. Det ville være paradoksalt, om netop vi - der er de første til at tale om genrebevidsthed - skulle overse, hvad der er af begrænsninger i den særlige genre, som hedder eksamensopgave.

Litteratur

Lærereksamen i skriftlig engelsk maj 2000

Lærereksamen i skriftlig engelsk januar 2001

Lærereksamen i skriftlig engelsk maj 2001

Tema: Evaluering - hvad skal vi kunne?

Ordliste

Michael Svendsen Pedersen

Bevis (*certificate*): Dokument som angiver, at en navngivet person har bestået en test eller en del af en test og har opnået en bestemt karakter, normalt minimum, bestået'.

Deskriptor (*descriptor*): Kort beskrivelse tilknyttet en kategori på en bedømmelsesskala, som opsummerer den færdighedsgrad eller præstationstype, som kræves for at få tilstrækkeligt med point til at havne i den pågældende kategori.

Diplom (*diploma*): Dokument som angiver at en navngivet person har bestået en test eller en del af en test og har opnået en bestemt karakter, normalt mindst et 'bestået'. Anses ofte for at dokumentere en højere grad af kvalifikation end et bevis.

Fleksibel test (*adaptive test*): Test hvor de forskellige items vælges undervejs på baggrund af deres sværhedsgrad som reaktion på en vurdering af deltagerens sprogevne. Ofte foregår testen ved hjælp af computer, men en fleksibel test kan også bestå af et interview.

Forprøve (*Pre-test*): Test som gives før starten på et kursusforløb. Resultatet af fortesten kan sammenlignes med de resultater, som opnås ved en anden test i slutningen af forløbet, hvilket muliggør evaluering af kursets effektivitet.

Item (*Item*): Element i en tekst, som bedømmes separat. For eksempel: en udfyldning i en cloze test; et multiple-choice spørgsmål med tre eller fire valgmuligheder; en sætning, som skal udsættes for grammatiske ændringer; et spørgsmål, hvortil der forventes et svar af sætningslængde.

Item-responsteori (*Item Response Theory (IRT)*): Gruppe af matematiske modeller til relatering af individuelle testpræstationer til individuelle færdighedsniveauer. Disse modeller er baseret på den grundlæggende teori, at en persons forventede besvarelser af et bestemt item, eller item, er en funktion af både det pågældende items sværhedsgrad og personens færdighedsniveau.

Konstrukt (*construct*): Hypotetisk abilitet eller mental egenskab, som ikke nødvendigvis kan konstateres eller måles direkte. I sprogtestning kan det f.eks. være lytteevne. Sprogtest forsøger ikke at måle de forskellige konstrukter, som ligger til grund for vores sprogfærdighed. Ud over

sprog-færdigheden selv, udgør motivation, indstilling og kulturel tilpasning alle relevante konstrukter.

Måling (*measurement*): Almindeligvis det at finde frem til mængden af et eller andet ved at sammenligne med en fast enhed, f.eks. at benytte en lineal til længdemåling. Inden for [socialvidenskaberne] henviser måling ofte til kvantificeringen af personers færdigheder, f.eks. sprogfærdighed.

Påvirkning (*impact*): Den effekt som skabes af en test, både med hensyn til indflydelse på de generelle uddannelsesmæssige processer og med hensyn til de personer, som påvirkes af testresultaterne.

Reliabilitet (*reliability*): Konsekvens eller stabilitet af måleresultater. Jo mere pålidelig en test er, des færre tilfældige fejl vil den indeholde. En test, som indeholder systematiske fejl, f.eks. forudindtagethed imod en bestemt gruppe, kan være pålidelige, men ikke valide.

Skævvægtning (*bias*): Test eller et item kan betragtes som skævvægtning, såfremt et bestemt segment af deltagergruppen vil få særlige fordele eller ulemper på grund af et eller andet træk ved testen eller itemet, som ikke har relevans for det, der skal måles. Skævvægtningen kan være i forhold til køn, alder, kultur etc.

Validering (*validation*): Indsamling af bevismateriale til støtte for slutninger, som er draget på basis af testresultater. Det er slutningerne vedrørende specifikke anvendelser af en test, der valideres, ikke selve testen.

Validitet (*validity*): Den udstrækning i hvilken testresultater gør det muligt at drage slutninger, som er hensigtsmæssige, meningsfulde og anvendelige ud fra testens formål. Det er muligt at skelne imellem forskellige validitetsaspekter, f.eks. indholds-, kriterierelateret- og begrebsvaliditet. Disse udgør forskellige former for beviser, som kan anvendes til bedømmelse af den generelle validitet af en test til et bestemt formål.

Ordforklaringerne er fra:

(Prepared by ALTE members) (1998): Multilingual Glossary of Language Testing Terms.
Cambridge University Press.

Tema: Evaluering - hvad skal vi kunne?

Godt Nyt Annoteret litteratur om temaet

Elsebeth Rise

Access. Issues in language test design and delivery. Ed. by Geoff Brindley & Gillian Wigglesworth. Sidney, Macquarie University, National Centre for English Language Teaching and Research, 1997. 245 s.

The Australian Assessment of Communicative Language Skills (access) er udviklet til brug for indvanderprocedurer. Den politiske baggrund beskrives, og der redegøres for studier fra et opfølgende forskningsprogram over validitet, reliabilitet og praktikabilitet med henblik på feedback til beslutningstagere og forbedring af design og administration. Om forholdet mellem lyttefærdighed beskrevet i testspecifikationerne og den aktuelle testpræstation; om betydningen af bedømmerens adfærd for de mundtlige karakterer; om antallet af opgaver og bedømmere til sikring af reliabiliteten m.v.

Alderson, J. C.: Assessing Reading. Cambridge, Cambridge University Press, 2000. 398 s.

Oversigt over igangværende og afsluttet forskning i psykologiske, sociale og kognitive faktorer i læsning; om forholdet mellem læsning og konstruktionen af sprogtests og vurderingsprocedurer, og en diskussion af hvordan vurderingen af læsning kan udvikle sig i forbindelse med computerbaseret testning.

Bachmann, Lyle F. & Adrian S. Palmer: Language Testing in Practice. Oxford, Oxford University Press, 1996. 376 s.

Den teoretiske referenceramme og testudviklingsprocessen fra planlægning til administration fremlægges, illustreret med en række testudviklingsprojekter.

Douglas, Dan: Assessing Languages for Specific Purposes.

Cambridge, Cambridge University Press, 2000. 311 s.

Teoretisk tilgang til testning på området fagsprog (f.eks. spansk til forretningsbrug, japansk for turistførere, italiensk for sproglærere, engelsk til trafikkontrol) med beskrivelse og vurdering af en række kendte tests. Om principper og procedurer for testudvikling i klasseværelser og på arbejdspladser.

Language Testing and Assessment. Ed. by C.M. Clapham & D. Corson. Kluwer, Academic, 1997. (Encyclopedia of Language and Education; 7)

Forskningsoversigter vedrørende testning af færdigheder i modersmål og fremmed/andetsprog, metoder, validitet, etik.

Little, David & Radka Perclová: The European Language Portfolio: a guide for teachers and teacher trainers. Strasbourg, Council for Cultural Co-operation, 2000. 90 s.

Introduktion til og vejledning i anvendelsen af ELP. Om trinnene i Europarådets Common European Framework som det fælles udgangspunkt; om at lære at lære, motivation og refleksion, selvvurdering, kollegadiskussion, bibliografi.

McNamara, Tim: Language Testing. Oxford, Oxford University Press, 2000. 140 s.

Oversigt over områdets hovedtræk, korte klip fra kildetekster med studiespørgsmål, referencer og glossar.

Poulsen, Erik & Kirsten Haastrup: Målfokuseret evaluering i engelsk - Principper og praksis. Kbh., Danmarks Pædagogiske Universitet, 2001. 99 s.

Et udviklingsarbejde om intern evaluering ud fra de fire Centrale Kundskabs- og Færdighedsområder (Kommunikative færdigheder; Sprogtilegnelse; Sprog- og Sprogbrug; Kultur og Samfundsforhold). Ved hjælp af logbøger, evalueringsopgaver, udviklingsportefølje, elevsamtaler vurderer elever og lærere løbende det faglige udbytte.

Shohamy, Elana: The Power of Tests. A critical perspective on the uses of language tests. Harlow, Longman, 2001. 182 s.

Om testning set i sociologisk, pædagogisk og politisk sammenhæng. Del 1: Om testning og magt; del 2: empiriske studier og cases i forbindelse med test brug; del 3: om resultater og konsekvenser for samfund og testtagere; del 4: kritiske forslag til fremme af demokrati og beskyttelse af testtagernes rettigheder.

Spolsky, Bernard: Measured Words. The development of objective language testing. Oxford, Oxford University Press, 2001. 402 s.

Om den historiske udvikling af og søgen efter objektive engelskprøver med støtte af store markedsorienterede institutioner. Om indføringen af TOEFL testen og virkningen heraf i den etablerede test-verden.
