

James A. Coleman

Profiling the advanced language learner: the C-Test in British further and higher education

The article describes a programme of experiments, coordinated by the University of Portsmouth with the collaboration of the Universities of Duisburg and Bochum. The intention is to use the C-Test across a range of languages (English, French, German, Spanish, Russian) and a range of levels of learner competence, to investigate both properties of the C-Test and aspects of language education in the United Kingdom. A pilot test in French has confirmed the C-Test's reliability, concurrent validity, discrimination and economy, and has indicated rates of progress for groups of advanced learners in formal education. A series of parallel investigations to be carried out in October 1993, linking a C-Test to a questionnaire for several thousand subjects, is expected to yield further information on the profile of UK language undergraduates, to identify subjects for longitudinal study, and to highlight correlations between language competence and other factors, including personality, attitude and motivation. Future experiments will look at these topics in greater depth, as well as at aspects of C-Test research.

1. Why use the C-Test?

The C-Test, developed by Ulrich Raatz and Christine Klein-Braley at the University of Duisburg, is a test of reduced redundancy similar to the cloze test, but in which the second half of every second word is deleted, from the second sentence on. It works on the principle that the better your knowledge of the language, the less linguistic information you actually need to be able to construct the meaning of an utterance.

The C-Test has been very thoroughly researched – indeed it has perhaps been used more often as a research topic than as a research tool (see summaries and bibliographies in Grotjahn, 1991, 1992; Klein-Braley, 1994a, 1994b). As a test, it offers a number of advantages:

- testees must apparently call on their entire language processing competence to restore the missing endings;
- it relies (though not exclusively) on authentic texts;
- it works with native-language, second-language and foreign-language speakers;

- it is economical in time, staff-hours and money since it is easy to construct, administer and score;
- it has excellent reliability and concurrent validity;
- it is objective, allowing exact scoring (cloze tests may be ambiguous, with a range of "acceptable" answers);
- it includes a range of items, typically 100 to 125 gaps in five superitems, thus not favouring a particular topic and providing items which are, taken together, syntactically representative;
- it is integrative;
- it provides a rapid overview of overall proficiency, and is therefore ideal for placement or as a research tool;
- it can be automated;
- it is norm-referenced, with good discrimination, especially where texts are pre-tested; it discriminates better than cloze for advanced homogeneous groups;
- it is likely that a formula to predict difficulty of texts for C-Test use may soon be developed for several languages (Bolten, 1992; Klein-Braley, 1993; Anckaert & de Valck, in press).

On the other hand,

- the task is inauthentic, with little face validity and arguable content validity;
- there is no oral/aural or interactive element;
- it provides little information on specific areas, so is less useful for aptitude, diagnostic, or progress testing, and has little teaching value;
- there is a danger of unhelpful backwash if widely used;
- it is unclear exactly what procedures learners are using (but this is true of all integrative tasks; and current research through a combination of error-analysis, thinking-aloud, introspective and retrospective methods may clarify the issue);
- it is unclear exactly what is being tested.

In summary, the C-Test, unlike many other test types, is easy to construct, quick to administer, easy and unambiguous to score, and highly reliable. Though it may resemble a narrower test of reading (cf. Baur &

Meder, 1989; Grotjahn & Tönshoff, 1992) or lexis or morphology, the C-Test is unrivalled in providing a snapshot of a learner's general competence in a foreign language.

2. The experimental context and the research questions

C-Test research, whether into the test itself or into aspects of language acquisition, may have universal significance. In the present case, there is also a purely local objective: to focus attention on language teaching and assessment in British Universities. Several factors combine to make such an evaluation an urgent need.

2.1 National university standards

British universities, apart from the six pre-Renaissance foundations (Oxford, Cambridge, four in Scotland) have been created in three waves. Most of the colleges born under industrial patronage in the nineteenth century achieved university status by the end of the Second World War, or shortly afterwards; in the 1960's, Colleges of Advanced Technology became universities and ten brand new ones were created; and in the 1990's Polytechnics were rechristened, bringing the total number of universities to nearly one hundred. But British universities have unequal prestige, and select new students on the basis of grades obtained in the specialised school-leaving examinations known as "A-levels" (Advanced Levels). Universities with high reputations can demand higher entry qualifications and thus recruit "better" students, in languages as in other subjects; but levels of linguistic competence at graduation are generally thought to be standardised nationally through the system of external examiners. The first question is therefore whether students starting from different baselines do achieve comparable levels of competence four years later. If so, differential rates of progress should be measurable, and factors influencing progress must be worth researching.

2.2 Student profile, and quality of teaching and testing

In spring 1992, there were at least 26,000 students of French in the UK (Thomas, 1993, p. 15), out of a total of over 53,000 (Ibid., 10). This is

a substantial number of students, whose successful learning is of national importance, yet little research has been carried out on why they are there, what they hope to gain from the experience, what their linguistic purposes and needs are, and whether they are learning efficiently and being reliably assessed.

Data collected in 1986 from a representative sample of 586 students for the Nuffield Modern Languages Inquiry was analysed by Dr Paul Meara but remains unpublished. It paints a disturbing picture. Despite the widespread criticism of the grammar-translation method of language teaching in the literature, most students in 1986 spent about 20% of their weekly study time on translation, with three quarters spending at least two hours a week on each of L1>L2 and L2>L1 translation. Even in 1993, an unpublished survey of final language exams in traditional universities by George Evans showed that all but one of the thirty respondents include a prose translation in final assessment, and in half of them students' written skills are evaluated only by means of a prose and general essay. If indeed institutions are relying on a few narrow, subjective, non-unidimensional forms of assessment; and since, in the name of academic independence, each university devises, administers, scores and interprets its own assessments, questions might legitimately be asked about the validity, the reliability and the economy of traditional university methods of language testing.

Surveys of the year abroad, motivated in part by the fear that if residence abroad were not shown to be a significant learning experience, funding for it might be withdrawn, have shown, not unexpectedly, that it does have significant value (Dyson, 1988), especially where students are in a work placement rather than in a university or a school assistantship (Willis et al, 1977). But the fact remains that although generalisations are often made about university language learners in the UK, they are based more on anecdote than on evidence.

Recent government policy has had the effect of promoting, within modern language departments, a split between research and teaching, between teachers of "content" (literature or area studies) and teachers of language (for more detail, see Coleman & Parker, 1992, pp. 5-16, Coleman & Rouxville, 1993, pp. 5-20). While the former have been rewarded with research funding, the activity of the latter is being increasingly deprofessionalised: it is frequently carried out either by subject specialists who have neither the training nor the desire to teach language, or by part-time staff who make

up over 50% of teachers in many universities, and 19% nationally (Thomas, 1993, p. 36).

Added to concern about the commitment and skills of university language teachers, there is plentiful anecdotal evidence that university-level language teaching may not be achieving its objectives. Students themselves voice the widespread perception that they know less French/German/Spanish at the end of second year than they did at A-level. And there exists a caricatural portrait of language teaching in British universities, by which, in past decades, universities relied on schools to teach the formal grammar and written skills of the target language, and on the year abroad to teach the oral/aural skills, including intonation and fluency, and the sociolinguistic competence. According to this picture, three years of university language teaching, relying heavily on the grammar-translation approach, in fact had virtually no effect on students' competence.

There is also anecdotal evidence, which needs further investigation, of ignorance among practising university language teachers concerning research findings in second language acquisition and language testing. It could be argued that language teachers do not need a specialised research knowledge on testing, but in the UK students are nearly always assessed by their teachers. There is no external objective feedback except from the external examiner, who is always an acquaintance, usually a friend, of the head of department and is partly chosen because of a shared general outlook: in any event, it is most unusual for an external to look at other than Finals examinations.

An informal, pre-pilot study (January 1993) gave predictable but nonetheless disturbing results: of thirty university language teachers, only one could name the obvious, eponymous journal on language testing, and only one could suggest a (more or less) relevant book. Half knew of the cloze test, about a third knew or guessed correctly about basic theoretical notions (test reliability, norm-referencing), about a quarter knew content validity, and only two were able to prove they could accurately define face validity or the fundamental statistical term "standard deviation". Without these concepts, I would argue that an evaluative reading of literature on language testing research is simply not possible.

On acronyms, the British Association for Applied Linguistics – the major British association for those interested in the theory and practice of language learning – is known to only one higher education language teacher

in six. The English-Speaking Union, whose definitions of levels of language competence are widely referred to, is known to only one in ten. Only one in thirty identified the Multiple Choice Question from its initials.

It is hoped that the C-Test research, which already involves one in eight British universities and a similar proportion of students of French, will in due course throw light on learner progress and on the effects of instruction, as well as on the profile of the learners and the factors influencing their learning. At the same time, it is hoped that the research will sensitise language teachers and testers in British Universities to some of the issues involved, and thus improve practice in the longer term.

3. The February 1993 pilot test

The February 1993 pilot test is the first stage of a planned programme of experimentation whose purposes are (1) to further define the usefulness (reliability, validity, practicality, difficulty prediction) of C-Tests in a number of languages; and (2) to use the C-Test to support investigations into university language learning and testing.

A selection was made of five C-Test texts (see Appendix) kindly provided by Dr Rüdiger Grotjahn of Bochum University. These were administered between 1 and 12 February 1993 to 427 testees, made up of

- pupils from local Further Education Colleges and independent schools in their final two years, i.e. preparing for A-level or AS-level examinations after five years' schooling in French;
- students at the University of Portsmouth studying French within the context of a degree course in French Studies, German and French Studies, Literary Studies with French, or Hispanic and French Studies; subjects were taken from the first year, intermediate year and final year of undergraduate courses; Hispanic-French students spend their second year abroad, half of it in France, while other students spend the third year of study abroad, in a French-speaking country for those on French or Literary Studies, in two locations for those studying French and German;
- students specialising in other disciplines and studying French via the Institution-Wide Language Programme (IWLP). Level 3 learners would be expected to be able to embark on an A-level course; while level 5 and 6 are both post-A-level, although students in the groups tested receive

only 2 hours' tuition a week and their French, unlike that of language specialists, receives no regular reinforcement in other classes;

- students on the MA in Contemporary French Studies;
- native French speakers, all exchange students preparing for the University of Portsmouth Diploma in English Studies, or the final year of the BA in French Studies.

The objectives of the experiment were to gain some experience in C-Test research, to look at the reliability of the test materials, and to obtain some idea of areas of research where the C-Test would provide a useful tool. All the scoring and analysis was done by the present author; there was no anonymity of testees and no double-marking. Data analysis was delayed by the unavailability of SPSS. It was always intended that this should be a pilot study, and the sample sizes in some cases were too small to be scientifically significant. The results were, however, of considerable interest.

They confirmed the C-Test, and the particular texts used, as being highly reliable (using Cronbach's alpha, $r_{xx} = .95$, $N = 427$) and providing excellent discrimination across a range of abilities, including the most advanced. For the whole experimental sample, the mean percentage score was 63.68 (median 64.00, standard deviation 18.88, range 77, minimum 23, maximum 100). The test proved very economical to administer and to score (90 seconds per C-Test once experience was gained).

In scoring, although, for objectivity's sake, only answers figuring in the original text were regarded as acceptable, it emerged that there does remain an element of subjective judgment, e.g. where the presence or inclination of an accent was uncertain. Students are of course past masters at producing deliberately ambiguous spellings when they themselves are unsure. The computerisation of C-Tests will obviate this problem.

In addition to the findings concerning the C-Test itself, the resulting statistics pointed to certain tentative conclusions concerning UK further and higher education.

They suggested, first, that students did progress from one year of study to the next, from the twelfth year of compulsory schooling (often called the "Lower Sixth Form") through to final degree level and even postgraduate study (see Table 1).

The second finding was that the level of competence of language graduates does not approximate to that of a native speaker (see Table 1, rows

Table 1
Mean scores on French C-Test

Group	Texts					Total	N
	1	2	3	4	5		
Native Speakers	19.28	18.25	19.74	19.56	19.03	95.84	31
University:							
MA students	15.50	17.00	18.25	16.25	18.00	85.00	4
Year 4	16.23	15.03	17.77	15.90	17.57	82.50	30
Year 2/3	14.76	13.89	16.31	14.87	15.93	75.76	45
Year 1	13.40	13.08	15.00	14.13	15.15	70.75	40
Secondary:							
Year 13	11.42	11.11	13.75	11.54	13.11	60.98	108
Year 12	10.17	8.76	12.08	9.73	11.76	52.51	108

Table 2
Comparison of native speakers with final-year students:
Range (R) and Standard Deviation (SD)

Group	Texts										Total	
	1		2		3		4		5		R	SD
	R	SD	R	SD	R	SD	R	SD	R	SD		
Native Speakers	17-20	.96	17-20	.92	18-20	.51	18-20	.62	17-20	1.12	90-100	2.31
University Year 4	13-20	2.08	12-18	1.61	14-20	1.65	12-20	1.88	13-20	1.78	69-94	6.31

1 and 3): there was very little overlap indeed between the performances of the final-year students and those of the native speakers (see Table 2). All the native speakers scored 90 or above; only 3 finalists scored over 90.

Thirdly, it emerged from closer analysis of figures for the intermediate year (for whom more data was available) that those studying French alone outperformed those studying French with another subject – German, Literary Studies, or Spanish. The last finding is surprising since, unlike French Studies students, the Hispanic-French students had spent six months in France in the previous year; samples, however, were too small and the standard deviation too large to draw any significant conclusions.

Fourthly, for the intermediate-year students, the grades at A-level (on which their admission to the course had depended) were very poor predictors of their language competence 18 months into the course, with the exception of the high-fliers who had scored an A grade, though numbers were very small (see Table 3).

Table 3
A-Level Grades and Mean C-Test Total
of Intermediate-Year Students

Group	Mean	N
A	87.50	4
B	76.13	8
C	75.00	22
D	74.11	9

In this context, it should be noted that (a) the A-level syllabus contains, as well as language tests, work on literature and/or area studies, so that the grade represents a composite of performance in different domains; (b) A-level grades are known to be poor predictors of success in university studies – though paradoxically they remain the major factor determining admission to such studies; (c) students with lower grades and lesser motivation may have been excluded at the end of Year 1.

Fifthly, it appeared that the French tuition given to IWLP levels 5 and 6 was inadequate even to maintain their competence level, which had slipped behind that of many secondary pupils, although the IWLP sample was too small to draw firm conclusions. In any event, contact hours are being increased to 3 per week in accordance with common practice elsewhere.

Sixthly, there was clearly a difference in profile between the four secondary institutions (see Table 4), although our study did not look into the complex of factors which might be responsible. In the absence of such analysis, of course, crude league tables have little value, but, in Britain at least, there is a political angle to C-Test research.

Seventhly, it was noted that, although not designed as a diagnostic test, the C-Test would provide valuable data for research into issues such as order of acquisition which are based on error analysis, for example in the areas of infinitive constructions and plural markers. There is also ample evidence to support the conclusion of Little & Singleton (1992, p. 188, cf. Singleton

Table 4
Mean C-Test Scores in Secondary Institutions

College	Year	Texts					Total
		1	2	3	4	5	
1	13	12.89	12.54	15.35	14.19	14.52	69.92
2	13	11.95	11.00	13.67	11.24	13.43	61.29
1	12	11.36	9.23	13.23	10.86	13.55	58.23
3	13	10.90	10.75	13.32	10.70	12.42	58.09
3	12	10.33	9.39	12.37	10.28	11.76	54.07
2	12	9.45	7.79	11.31	8.46	11.04	48.00
4	13	8.71	9.00	11.57	9.00	10.60	48.00
4	12	9.23	7.92	10.85	8.75	10.17	46.58

& Little, 1991) that "subjects tended to give priority to a ready lexical solution over morpho-syntactic and more general semantic issues". To give some examples:¹

Mais ne perdez pas courage et oubliez votre timidité (Text 1)

was frequently rendered by the syntactically impossible

**Mais ne personne pas ...*

while for

Comment ne pas se laisser écraser par le stress (Text 2)

by far the most popular answer was

**se laisser écrire*

a clear victory for familiar lexis over semantic feasibility.

There were also instances where subjects had apparently been misled by the spelling into postulating the wrong vowel sound (e.g. *Mais, sans, arranger* in Texts 1, 3, 4). Such short-circuiting of normal language processing, triggered by the form of the stimulus, is expected to be more frequent among lower-level learners: our data has not yet been analysed to establish whether this is so.

Finally, and more flippantly, no test will be widely adopted for use in British higher education unless it provides the marker with a level of

¹ Bold type indicates the deleted portion; the erroneous response is marked by an asterisk (*). The C-Test texts are listed in the Appendix.

amusing wrong answers similar to that provided by existing test formats, particularly translation and general essay. I suspect that such unreliable and hybrid tests are retained in no small measure for their entertainment potential. In this domain too, however, the C-Test performed well.

Text 2:

*Pour réussir, il faut d'abord être en forme (en *force, en *foi).*

*Faut-il se lancer dans le jogging, commencer une cure de vitamines (*comme une cuisson de vitalité, *comme une curriculum de vitae), ou partir se détendre (*se détenter, *se déteindre, *se déterminer, *se déteriorer, *se déterer, *se détester) ...*

*... les fatigues de l'année scolaire (*de l'anglais scolaire) ...*

Text 3:

*C'est le système éducatif qu'il faut changer (*chanter). De l'argent (*de l'argot), ça ne suffit pas! ... Il y a trop (trois, treize, trente, trois cents) d'élèves par classe.*

Text 4:

*Peut-on faire un pas sur notre bonne vieille planète Terre sans rencontrer l'homme? Non. Impossible (*Imposez, *Impovisation, *Impolite, *Impotence)! Toujours on se trouve face à une de ses innombrables réalisations (*réalisateurs).*

Text 5:

*Les hommes y affrontent une nature hostile: le brouillard, le froid (*le français) mais surtout le vent qui peut atteindre 200 km/h (*le verre, *le verglas, *le vegetation qui peut atteindre 200 km/h).*

In a parallel but smaller pilot study, Maria-Teresa Rondon tested 21 first-year students and 12 final-year students of Spanish, together with 9 native speakers: in this case, too, significant progress was found to have been made by the Portsmouth students, but the relatively low scores of the native speakers give cause for concern (see Table 5).

Table 5
Rates of Progress in Spanish (Mean C-Test-Totals)

Groups	Mean
Year 1 students	57.38
Year 4 students	67.42
Native speakers	72.33

3.1 Concurrent validity

There have been a number of studies seeking to establish the concurrent validity of C-Tests (as opposed to its construct validity, see Stemmer, 1992, p. 97). One secondary institution provided me with a listing of testees' results in A-level examinations. The Board chosen by the College² was the University of Cambridge Local Examinations Syndicate (UCLES), whose A-level examination has five components: Oral (25%), Reading and Writing (25%), Listening (20%), Essay (10%) and Coursework (20%). Written accuracy is assessed in components 2 and 4, and comprehension in 2 and 3 (and 1 in a different context). Individual papers are marked on a scale of 8, and the weighted average over 5 components converted into a grade. 1 and 2 correspond to the top grade A, 3 to B, 4 to C, 5 to D, 6 to E, 7 to N (near-miss) and 8 to U (Ungraded). 56 students took both the C-Test and, three to four months later, the A-levels.

The A-level components themselves are reliable ($\alpha = .87$). In the first calculation of a Pearson product-moment correlation coefficient between A-level grades and C-Test totals, raw marks (themselves quite normally distributed) were used: although both scores 1 and 2 correspond to grade A, the correlation coefficient is very highly significant at $r = -.74$ ($P = .000$). The correlation is negative since C-Test marks are percentages while for A-levels the best performances are represented by the lowest figures). Two alternative approaches to calculating correlations gave similar results: if $A = 7$, $B = 6$, $C = 5$, $D = 4$, $E = 3$, $N = 2$ and $U = 1$, $r = 0.74$; if learners are ranked by total score (weighted as in the calculation of A-level grades), the Spearman rank-order correlation coefficient $\rho = .70$.

² Whilst A-levels are national examinations, schools can choose from those offered by several Examination Boards, although standards are deemed to be equivalent.

The Pearson correlations between C-Test totals and individual A-level components are also very highly significant:

Component 1: $r = -.48$

Component 2: $r = -.78$

Component 3: $r = -.76$

Component 4: $r = -.47$

Component 5: $r = -.48$

The highest correlation is thus with Reading and Writing and with Listening. (Comparable figures for A-level grade with the five components are .71, .90, .85, .74, and .78.)

The fact that students are graded differently on different linguistic skills emphasises, if further emphasis were needed, that the Unitary Competence Hypothesis is a chimaera. Nonetheless, the C-Test, in this small-scale study, does seem to provide similar results to the overall A-level grades, and is of course far more economical. To suggest replacing A-level exams with a C-Test would naturally be impossible, not least because of the likely harmful backwash and the absence of oral-aural assessment. Nonetheless, it is interesting to note that the papers which can be heavily prepared are 1, 4 and 5; papers 2 and 3 are the most unpredictable, the ones which bring out, in one teacher's words, "the general intellectual skills" of "the real A-level student" and which best identify "university material". These are precisely the papers with which the A-level grade, and the C-Test, correlate best.

Another secondary institution provided A-level grades for the 25 students who had earlier completed the C-Test. Using the same conversion ($A = 7$, $B = 6$, etc.), the Pearson correlation coefficient was rather lower than the first institution at .63 ($P = .001$); this is probably because A-level grades at the second institution were grouped at the high end, with only one grade lower than C.

Marks for a group of intermediate-year students at Portsmouth University were compared to their C-Test totals. Unsurprisingly, there was no correlation with certain components, such as the video project for which students are awarded a group mark on the basis of criteria which are largely non-linguistic. There was, however, a significant correlation with the written skills coursework component ($r_{tc} = .50$, $P = .001$, $df = 41$), and with the written-grammar coursework component ($r_{tc} = .50$, $P = .001$, $df = 42$),

as well as with the written exam ($r_{tc} = .54$, $P = .000$, $df = 36$) and the oral exam ($r_{tc} = .43$, $P = .008$, $df = 36$). But the two highest correlation coefficients were with the overall coursework mark, a composite of six exercises including the group video project ($r_{tc} = .60$, $P = .000$, $df = 33$), and with the global mark for all coursework and both exams ($r_{tc} = .73$, $P = .000$, $df = 33$). These results reinforce the judgment that the C-Test is an economical way of estimating overall language competence.

3.2 Conclusions on the February tests

Nothing in the February tests brings new insights to the literature on C-Tests. But they proved very reliable, highly discriminatory even for advanced learners, and have good concurrent validity. They provide an excellent tool for looking at levels of general proficiency of advanced learners. Questions concerning their construct validity or what it is they actually test are underlined by our evidence of possible correlation with general intellectual skills, and by the solution strategies which apparently depend sometimes on graphical and lexical rather than morpho-syntactic and semantic clues (cf. Anckaert & Beckmans, 1992; Little & Singleton, 1992; Stemmer, 1992).

4. Research questions for stage 2: the Oktobertest

If, as Ellis (1987) and others have argued, there is a need for research above all into the effects of instruction, such research must depend on reliable, valid and economical quantitative measures of language competence. We believe the C-Test may meet some of these needs, hence the need we perceive for both further research into the reliability and concurrent validity of the C-Test, and for research into optimising its administration (rubrics, scoring procedures, cf. Grotjahn, 1987; Little & Singleton, 1992). Other C-Test research questions to be addressed concern its teachability, and the instructions to testees: informing them that about half (as opposed to part) of the word has been deleted seems to limit the range of errors produced and to lead to processing that can be more mathematical than linguistic (cf. Grotjahn, 1987, p. 227).

At the same time, we want to know more about the make-up of the UK population of language undergraduates. The integrative/instrumental opposition which characterised research into motivation two decades ago

(Gardner & Lambert, 1972) may well need updating in the face of the increasing homogenisation of global society, and of the fact that the world status of certain languages, particularly English, makes the target population hard to define, while increased travel and media exposure mean that the shock of the foreign is no longer as intense as in the early 1970's. Nonetheless, such an approach is a good starting point for examining motivation, allied to a series of questions in the pilot questionnaire designed to elicit personality variables, in particular willingness to take risks and to use the target language. Parallel questions seek to ascertain whether willingness to use language in the target country is matched by the learner's normal behaviour in their own language community. One member of the research team, Jane Freeland, is particularly interested in the problems encountered by UK-based students broadly classified as "bilinguals". Most classes will contain one or two students whose family background includes exposure to the target language, but as a group they have been little studied, and the Oktobertest questionnaire seeks to identify them. A broad longitudinal study is planned: indeed for students at Portsmouth University and nearby secondary schools, the longitudinal study has already begun.

It is also hoped that the evolving programme of research, which will develop in response to the personal interests of the research team as well as to funding possibilities, may look at the frequencies of particular errors in order to throw light both on the processes involved in filling the gaps in C-Tests (cf. Little & Singleton, 1992; Klein-Braley, 1994a, b; Stemmer, 1991, 1992) and on the order in which certain lexical and structural items are internalised (transfer of declarative to procedural knowledge), and whether teaching them influences the rate or the route of acquisition, as well as on which errors occur at a particular level or in relation to a particular L1. A longitudinal study might be expected to yield data on the effects of instruction and on factors influencing the progress of advanced learners, but while the shortness of the texts is a major factor in favour of the C-Test's economy, the same shortness means that on their own, they simply do not contain enough examples, and supplementary methods of elicitation will be needed. Nevertheless, since data collection is already taking place on a very large scale, it would be a pity not to exploit all the data derived from the survey, perhaps in comparison with tests of a more specifically diagnostic nature.

The international dimension of this research has been important from the start. Language testing is now an international issue (hence the formation of ALTE, the Association of Language Testers in Europe, which brings together language testing institutions from 9 European countries). The initial involvement of three countries in our C-Test experiments (Great Britain, Germany and Austria) will, it is hoped, extend to other European countries with the next stage of testing. International comparisons – of levels of learner competence in different languages, of progress rates, of motivation, etc. will thus be possible.

At this stage, we do not envisage using introspective or retrospective analyses of their responses by subjects themselves. Though undoubtedly productive, this method is very time-consuming and is open to the charge that subjects' attempts to rationalise what may be unconscious processes are unreliable.

Lorraine Pickett, a member of the Portsmouth team, is undertaking doctoral research into the culture-fairness (or otherwise) of C-Tests, comparing learners of English from five very different linguistic and cultural backgrounds.

4.1 The Oktobertest

The shorthand title, derived from the Germano-British cooperation which underlies the research programme, refers to an experiment to be carried out during October 1993. Subjects will complete both a C-Test and a purpose-designed questionnaire focussing on age, sex, institution, course, language background and qualifications, motivation, attitudes, expectations, personality (in particular linguistic risk-taking in what we have called L1land and L2land³) and learner's own evaluation of their linguistic competence and metalinguistic knowledge. The primary function of the questionnaire is to define areas for future, more detailed investigation.

The final details of the Oktobertest remain to be determined at the time of writing, but it is likely that French learners will be tested in five

³ We have developed a convenient shorthand reference: on the model of such fictional countries as Never-Never-Land and Munchkinland, and rather than refer repeatedly to "the country where the learner's native language is spoken" and "the country where the language the learner is being tested on is spoken", we talk of L1land and L2land. This is particularly useful in an international research project.

secondary schools and in twelve universities in Britain: the latter include institutions raised in status from college to university in each of the three major expansions to the British university system (pre-1950, 1960s, 1990s). The German universities of Duisburg and Bochum will also participate. It is expected that the total testee population will be around 2000.

Since the C-Tests in German, English, Spanish and Russian have not yet been piloted, their use in the Oktobertest will be more restricted. German tests will be used in Portsmouth and Duisburg universities and in secondary schools/further education colleges. English C-Tests will be used in Duisburg, Bochum, Klagenfurt and Portsmouth universities. Spanish C-Tests will be used in Portsmouth and Southampton universities, and Russian tests in Portsmouth alone. Results will be published.

Acknowledgments

We acknowledge the financial support of the British Council and Deutscher Akademischer Austauschdienst under the Academic Research Collaboration programme, and of the Universities of Portsmouth and Duisburg. We are grateful for the collaboration of Havant College, Portsmouth College, Portsmouth Grammar School and South Downs College in the collection of data in February 1993. We are very grateful for the continuing collaboration of Dr Christine Klein-Braley, Dr Rüdiger Grotjahn, and Professor Ulrich Raatz.

References

- Anckaert, Philippe & de Valck, Inez. (in press). Le C-Test comme épreuve de compétence linguistique globale. Questions relatives au facteur texte. In *Proceedings of the 10th World Congress of AILA*.
- Anckaert, Philippe & Beeckmans, Renaud. (1992). Le C-Test. Difficulté intrinsèque, pouvoir discriminant et validité de contenu. In Grotjahn (1992), 145-172.
- Baur, Rupprecht S. & Meder, Gregor. (1989). Die Rolle der Muttersprache bei der schulischen Sozialisation ausländischer Kinder. *Diskussion Deutsch*, 20, 119-135.
- Bolten, Jürgen. (1992). Wie schwierig ist ein C-Test? Erfahrungen mit dem C-Test als Einstufungstest in Hochschulkursen Deutsch als Fremdsprache. In Grotjahn (1992), 193-204.

- Coleman, James A. & Parker, Gabrielle. (Eds.). (1992). *French and the enterprise path. Developing transferable and professional skills*. London: AFLS/CILT.
- Coleman, James A. & Rouxville, Annie. (Eds.). (1993). *Integrating new approaches. The teaching of French in higher education*. London: AFLS/CILT.
- Dyson, Peter. (1988). *The year abroad. The effect on linguistic competence of the year spent abroad by students studying French, German and Spanish at degree level*. Oxford: University Language Teaching Centre.
- Ellis, Rod. (1987). Does remedial instruction work? A L2 acquisition perspective. In James A. Coleman & Richard Towell (Eds.), *The advanced language learner* (pp. 55-73). London: AFLS/CILT.
- Gardner, Robert C. & Lambert, Wallace E. (1972). *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury House.
- Grotjahn, Rüdiger. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer.
- Grotjahn, Rüdiger. (1991). O Teste-C de leitura: Uma visão panorâmica. *Letras de Hoje* [Porto Alegre, Brazil], 86, 145-173.
- Grotjahn, Rüdiger. (Ed.). (1992). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 1). Bochum: Brockmeyer.
- Grotjahn, Rüdiger. (Ed.). (1994). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2). Bochum: Brockmeyer.
- Grotjahn, Rüdiger & Tönshoff, Wolfgang. (1992). Textverständnis bei der C-Test-Bearbeitung. Pilotstudien mit Französisch- und Italienischlernern. In Grotjahn (1992), 19-96.
- Klein-Braley, Christine. (1994a). *Language testing with the C-Test. A linguistic and statistical investigation into the strategies used by C-Test takers and the prediction of C-Test difficulty*. Duisburg: University of Duisburg. (to appear)
- Klein-Braley, Christine. (1994b). Towards a theory of C-Test processing. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3). Bochum: Brockmeyer. (to appear)
- Little, David & Singleton, David. (1992). The C-Test as an elicitation instrument in second language research. In Grotjahn (1992), 173-192.

- Singleton, David & Little, David. (1991). The second language lexicon: some evidence from university level learners of French and German. *Second Language Research*, 7(1), 61-82.
- Stemmer, Brigitte. (1992). An alternative approach to C-test validation. In Grotjahn (1992), 97-144.
- Thomas, Gareth. (1993). *Survey of European languages in the United Kingdom*. London: Council for National Academic Awards.
- Willis, Frank, et al. (1977). *Residence abroad and the student of modern languages - A preliminary survey*. Bradford: University of Bradford.

Appendix: French C-Test for Oktobertest

Surname _____ Forenames _____
Date of birth _____
University, college or school _____
Class or year group _____ date / /1993

In the following texts, part of some of the words is missing. Please write in the missing letters. You have 25 minutes.

Text 1

Devenez une star en espagnol!

Dire deux ou trois mots d'espagnol, presque tout le monde sait faire. Mais téléph_____ à Madrid o_____ parler polit_____ avec u_____ Mexicain, voi_____ qui e_____ déjà u_____ peu pl_____ compliqué. Ma_____ ne per_____ pas cour_____ et oubl_____ votre timid_____ : il n'e_____ jamais tr_____ tard po_____ faire d_____ progrès. L'espa_____ est tr_____ proche d_____ français. C'est donc pour vous une langue relativement facile à apprendre.

Text 2

La clé de la réussite aux examens.

La réussite à un examen, c'est 50% de connaissances et 50% de résistance physique et nerveuse. Pour réu_____, il fa_____ d'abord êt_____ en fo_____. Mais com_____ faire? Comment n_____ pas s_____ laisser écr_____ par l_____ stress e_____ les fati_____ de l'an_____ scolaire? Faut-_____ se lan_____ dans l_____ jogging, comm_____ une cu_____ de vita_____, ou part_____ se déte_____ sur une plage dorée de Thaïlande? Du calme! Il faut réagir sagement.

Text 3

Comment améliorer le système scolaire en France? Les suggestions d'une lycéenne française.

Les horaires des cours sont trop longs, les programmes trop ambitieux. Bref, c'e_____ le sys_____ éducatif qu'_____ faut cha_____. De l'arg_____, ça n_____ suffit p_____! Même quel_____ milliards ne vo_____ pas arra_____ notre situ_____! C'est u_____ réforme tot_____ qu'il no_____ faut. I_____ y a tr_____ d'élèves p_____ classe, o_____ n'a p_____ assez d_____ profs, comment faire? Proposons alors de réduire les cours de 60 à 40 minutes.

Text 4

Antarctique: la mémoire de la terre

Dernier continent découvert par l'homme, l'Antarctique est un laboratoire fabuleux pour les scientifiques du monde entier. Da_____ ses vas_____ glaces s_____ trouve gravée l'hist_____ de la plan_____ Terre. Peut-on fa_____ un p_____ sur no_____ bonne vie_____ planète Terre sa_____ rencontrer l'ho_____? Non. Impo_____! Toujours o_____ se tro_____ face à u_____ de s_____ innombrables réalisa_____. Presque par_____ on const_____ les dég_____ qu'il cause à l'environnement. Pourtant il y a encore un vaste continent à peu près préservé: l'Antarctique.

Text 5

Chercheurs de pétrole en mer du Nord.

Frigg en mer du Nord. A 250 km des côtes norvégiennes, 200 hommes travaillent pour extraire quelques millions de mètres cubes de gaz. Les hom_____ y affrontent u_____ nature hos_____ : le broui_____, le fr_____, mais sur_____ le ve_____ qui pe_____ atteindre 200 km/h. S_____ Frigg, les condi_____ de tra_____ sont diffi_____ : douze heu_____ quotidiennes pen_____ quinze jo_____ de su_____ avant d_____ pouvoir pre_____ un con_____ de trois sema_____. L'hélicoptère, seul lien avec la terre ferme, assure le transport des hommes, des marchandises et du courrier.