

DOCUMENT RESUME

ED 275 163

FL 016 067

AUTHOR Alderson, J. C.; Urquhart, A. H.
TITLE ESP Tests: The Problem of Student Background Discipline.
PUB DATE 84
NOTE 15p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 065.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English for Special Purposes; Foreign Countries; Foreign Students; Higher Education; *Language Tests; *Prior Learning; *Reading Comprehension; *Testing Problems
IDENTIFIERS England

ABSTRACT

This study examines the effects of foreign students' knowledge of a particular academic area on their comprehension of written texts in English. The students' university majors or areas of study included administration and finance, engineering, math and/or physics, and liberal arts. In two studies, the students performed tasks related to texts in engineering, administration and finance, and general material. In the third study, students took different versions of a standardized English language study skills test. In the first two studies, consistent differences in comprehension were found between groups and among texts. Some slight discrepancies with expected results were found in the third study, but overall the three studies confirm the importance of prior knowledge in reading comprehension in English as a second language. (MSE)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED275163

ESP TESTS: THE PROBLEM OF
STUDENT BACKGROUND DISCIPLINE

J.C. ALDERSON (University of Lancaster)
A.H. Urquhart (University of Aston)

FL 016 067

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

P. Culhane

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

ESP TESTS: THE PROBLEM OF STUDENT BACKGROUND DISCIPLINE

J C Alderson (University of Lancaster) and A H Urquhart (University of Aston)

This paper reports on three studies designed to investigate the effect of students' background discipline - i.e. knowledge of a particular academic area - on their comprehension of written texts. One of the principal motivations underlying the studies was the wish to collect empirical data on the necessity or desirability of ESP proficiency tests for tertiary level students coming to study in an English-speaking country.

The studies were as follows:

- a) Study 1: the pilot study; at Aston University; five short texts gap-filling task.
- b) Study 2: at Aston and Lancaster Universities: same five texts; gap-filling and questions/short answers.
- c) Study 3: at Aston and Lancaster Universities; different versions of the British Council ELTS Study Skills Module (MI).

Study 1: An account of this study has already appeared in print (Alderson and Urquhart 1983). The description given here will therefore concentrate on those parts of the study which also relate to Study 2.

Subjects: The students taking part were studying on an English for Study Skills Course at the University of Aston. They fell into four groups depending on the academic area in which they were going to pursue their post-graduate studies. The groups were 1) Development Administration/Finance (15 students); 2) Engineering (11); 3) Maths and/ or Physics (6); 4) Liberal Arts (5).

According to an internal placement test, the first three groups were of very similar linguistic proficiency; the fourth group was more proficient.

Texts: There were five texts in all, two Engineering, two Development Adm'n/Finance, and one 'general' text. The Engineering texts (titles have been supplied for reference purposes) were 'Turbines', from an article in an Engineering periodical (Hulme, 1981) and 'Electrolytes', from an academic monograph (Gregory 1972). The Development/Finance texts were 'Malaysia' and 'Polanyi', both from a university text-book (Latham 1978). The 'general' text was 'Quixote', one of the 'Rate-builder' texts from the SRA 3B kit (Science Research Associates 1963). According to the publishers, this kit is designed for high or junior-high school L1 students.

Each text was between 250 and 280 words long. 'Electrolytes' included a diagram with caption, and 'Malaysia' a table. The Fog index for each text was between 17.5 and 18.5.

Tasks: A number of lexical items (between 15 and 22) were deleted from each text. In the judgement of the authors, all were restorable in terms of the text itself, without, that is, appeal to knowledge outside the text. An effort was made to choose items whose restoration depended on an overall grasp of the text, rather than on nearby cotextual items. In the case of 'Electrolytes' and 'Malaysia', certain items could be restored only after consulting the diagram or table respectively.

Given that the object of the experiment was to test the effect of background knowledge, it might appear paradoxical to insist that all items should be restorable on the basis of the text. Anything else, however, would have been trivial. The ability to restore the missing item in the text, "The battle of Hastings took place in the year ..." clearly depends on background knowledge, but in an uninteresting way. It would seem that one of the assumptions underlying traditional 'general' tests is that the information required to be extracted from the text is equally available to all readers. Since it is one of the major hypotheses of this experiment that in spite of such availability, certain groups will still be advantaged by a particular text, it is clearly essential that in the experiment, all the required information should at least in theory be equally accessible to all subjects.

Results: Detailed results can be found in Alderson and Urquhart (op.cit). They can be summarized here as follows:

- 1) on the two Engineering texts, the Engineering group did better than the Development Admin/Finance group.
- 2) similarly, the Development Admin/Finance group outperformed the Engineers on the two Development texts.
- 3) there was a marked similarity between the performance of the Engineers and Physics/Maths students, on the one hand, and the Development and Liberal Arts groups on the other. The latter two groups outperformed the Engineering and Physics/Maths group on 'Quixote', the 'general' text, suggesting that Engineering and Science students may be disadvantaged by such texts.
- 4) there was a marked text effect. Overall, 'Turbines' proved to be the easiest text, and 'Polanyi' and 'Quixote' the most difficult. There was no reason to predict this: the only measure of linguistic complexity used, the Fog Index, had indicated that all five texts were all at the same level. At this stage, one should perhaps refer, not to text difficulty, but to text x method. Indeed, without a satisfactory correlative of linguistic complexity, this is perhaps all that can ever be claimed. The text effect did not affect the hypothesis; it may, however, be relevant to the interpretation of the results of Study 2, as discussed below.

Study 2: This fell into two parts, namely:

- 1) an exact replication of Study 1
- 2) a partial replication using the same texts but a different test task, namely Short Answers.

The reason for replicating the Pilot Study is obvious - the numbers involved there were very small, too small, in fact, to justify calculating significance values. The reason for extending the Pilot Study by adding another test task is also obvious: without such an extension, the study is vulnerable to method effect.

Subjects: students on ESS courses at both Aston and Lancaster University took part. The groups were very similar in academic discipline to those participating in Study 1, namely:

- 1) Development Admin/Finance (Aston) and Economics (Lancaster); from now on referred to as DAFE.
- 2) Engineers (ENG)
- 3) Science and Maths (S&M)
- 4) Liberal Arts (L/A)

Numbers of subjects doing (a) gap-filling (b) Short Question/Answer, and (c) both are shown in table 1.

Table 1

	Gap-filling			Short Answer			Both	
	Aston	Lancaster	T	Aston	Lancaster	T	Both	
DAFE	18	8	26	24	9	33	16/8	24
ENG.	17	4	21	16	4	20	16/4	20
S M	32	6	38	28	6	34	24/6	30
L/A	13	9	22	10	16	26	10/9	19

The Gap-filling test was administered first; the Short Answer test about a week later.

Results: The Gap-filling test was marked three times, once for exact restoration, once for exact restoration, once for synonyms, and once for any response judged acceptable in the general context. Results are presented in Table 2.

TABLE 2

A. Gap-filling (Variable N)

	Economics						Engineering									
	Polanyi			Malaysia			Turbines			Electrolytes			Quixote			
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
DAFE	X	9	31	43	17	33	39	33	47	56	19	34	40	10	28	33
	SD	9	23	22	8	16	22	14	20	21	12	18	16	9	12	20
ENG	X	3	14	27	10	23	24	32	49	52	21	40	47	5	18	25
	SD	4	12	18	9	17	19	17	19	19	12	16	16	7	12	14
S&M	X	6	17	27	20	31	33	30	42	47	28	47	54	7	20	22
	SD	6	13	17	7	11	14	15	21	20	11	13	15	6	11	14
L/A	X	13	30	54	21	33	34	43	61	63	22	37	44	16	38	44
	SD	12	19	24	14	23	21	13	17	23	16	21	21	11	14	14

a= exact
b= synonym
c= acceptable

Table 3

Gap-FIT (Eng: significant differences between means (t-test) a = exact b = synonym c = acceptable

	ENG			S&M			L/A		
	a	b	c	a	b	c	a	b	c
1) Polanyi									
DAFE	NS	pe.05	pe.05	NS	pe.05	pe.01	NS	NS	NS
ENG				NS	NS		pe.05	pe.05	pe.01
S&M							pe.05	pe.05	pe.001
2) Malaysia									
	a	b	c	a	b	c	a	b	c
DAFE	NS	NS	pe.05	NS	NS	NS	NS	NS	NS
ENG				pe.01	NS	NS	NS	NS	NS
S&M							NS	NS	NS
3) Turbines									
	a	b	c	a	b	c	a	b	c
DAFE	NS	NS	NS	NS	NS	NS	NS	NS	NS
ENG				NS	NS	NS	NS	NS	NS
S&M							pe.05	pe.05	pe.05
4) Electrolytes									
	a	b	c	a	b	c	a	b	c
DAFE	NS	NS	NS	NS	pe.05	pe.01	NS	NS	NS
ENG				NS	NS	NS	NS	NS	NS
S&M							NS	NS	NS
5) Quixote									
	a	b	c	a	b	c	a	b	c
DAFE	NS	NS	NS	NS	NS	pe.05	NS	NS	NS
ENG				NS	NS	NS	pe.01	pe.01	pe.01
S&M							pe.01	pe.01	pe.001

Significant differences between groups for all three methods of scoring are presented in Table 3.



The following points emerge:

1) The DAFE group did significantly better on the Development texts than did the Engineers, although on 'Malaysia', there was no difference between DAFE and S&M.

2) However, on the Engineering texts, the Engineers did not out-perform the DAFE group, although S&M did significantly better on 'Electrolytes'.

3) L/A did significantly better than either S&M or ENG on 'Quixote'.

Results for the Short Answer test are presented in Table 4.

Table 4

8. Short Answers

		Polanyi	Malaysia	Turbine	Quixote	Electrolytes
DAFE	X	47.1	56.1	53.8	40.5	48.5
	SD	25.1	22.3	17.1	18.6	21.9
ENG	X	24.5	42.0	58.1	46.6	49.5
	SD	22.7	23.7	24.1	18.6	21.9
S&M	X	34.9	46.1	58.0	46.6	40.8
	SD	22.3	20.0	16.4	18.6	21.3
L/A	X	55.7	54.4	60.3	46.6	70.8
	SD	22.8	20.0	20.2	19.4	14.3

Significant differences between groups are shown in Table 5. (See over).

The pattern of results is similar to that shown by the scores on Gap-filling. On the two Development texts, DAFE performed significantly better than ENG. However, there was no significant difference between DAFE and ENG on the two Engineering texts. The L/A group out-performed DAFE and ENG on 'Polanyi' and all other groups on 'Quixote'.

If one compares results under the two test methods, the following points emerge:

1) there was a strong method effect. Although the overall correlation between scores on both tests is quite high (.78), correlations between scores on particular tests vary widely. Another sign of method effect is that the rank order of text difficulty is not the same under both methods. On Gap-filling, 'Electrolytes' is a comparatively easy text, ranked second overall; 'Quixote' is difficult, ranked fourth. On Short Answers, however, these ranks are reversed.

Table 5

Significant differences between means (t-test)

B. Short answers

1) Polanyi	ENG	S&M	L/A
DAFE	p .01	NS	NS
ENG		NS	p .001
S&M			p .01
2) Malaysia	ENG	S&M	L/A
DAFE	.05	NS	NS
ENG		NS	NS
S&M			NS
3) Turbines	ENG	S&M	L/A
DAFE	NS	NS	NS
ENG		NS	NS
S&M			NS
4) Electrolytes	ENG	S&M	L/A
DAFE	NS	NS	NS
ENG		NS	NS
S&M			NS
5) Quixote	ENG	S&M	L/A
DAFE	NS	p .05	p .01
ENG		NS	p .001
S&M			p .001

2) In spite of this, the text effect found in Study 1 remains strong over both methods, with 'Turbines' being the easiest text, and 'Polanyi' the most difficult.

3) As in Study 1, the ENG and S&M groups performed very similarly (though not identically). This was also true of DAFE and L/A. This finding, of course, may have an important bearing on the number of ESP tests required.

4) The main point that emerges under both methods is that the hypothesis was only partially confirmed: the DAFE group outperformed the Engineers on Development texts, and the L/A group did better on 'Quixote' than any other group. However, the Engineers did not appear to be advantaged by reading Engineering texts.

In part, this may be due to differing degrees of linguistic proficiency. Pre-testing of students at both universities revealed that the L/A group was the most proficient. At Aston, the three other groups were at the same level, but the Lancaster Economists were superior both to ENG and SAH.

Linguistic proficiency may go part-way to explaining the results. For the Lancaster students, correlations were available between scores on different texts, and scores on ELBA and ELTS tests. These correlations are set out in Table 6.

Table 6

Test correlations with ELBA and ELTS

A. Short Answer	N=35			N=20		
	ELBA 1	ELBA 2	ELBA Total	ELTS G1	ELTS G2	ELTS M1
Malaysia	.45	.38	.43	.45	NS	.38
Polanyi	.52	.52	.56	.68	.69	.69
Turbines	.42	.50	.50	.55	.41	.49
Electrolytes	NS	NS	NS	.44	NS	NS
Don Quixote	NS	.45	.40	.61	NS	.49

B. Gap-filling (Acceptable)	N 21/25			ELTS N = 11/13		
	ELBA 1	ELBA 2	ELBA Total	ELTS G1	ELTS G2	ELTS M1
Malaysia	NS	.34	NS	.55	NS	.54
Polanyi	.53	.61	.63	.67	.51	.71
Turbines	.57	.66	.66	.73	NS	NS
Electrolytes	NS	NS	NS	NS	NS	NS
Don Quixote	.50	.46	.51	NS	NS	NS

The highest correlations are between scores on the language tests and on 'Polanyi' and 'Turbines'. Under both methods, the L/A group, linguistically the most competent, did better than the other 3 groups. One might, then, envisage scores as being made up of (a) a linguistic proficiency component; (b) a background knowledge component. On Engineering texts, the Engineers' proficiency component was smaller than the DAFE group, but their background knowledge component was greater, thus levelling scores. On Development texts, however, the DAFE group was superior on both components, hence their higher scores.

While linguistic proficiency would definitely seem to have been a factor, however, the explanation outlined above will not stand. It is true that in the DAFE group, the Lancaster Economists seem to have been more proficient; this was not true, however, of the Aston Development Admin/Finance group, who formed the larger part of the DAFE group as a whole. A calculation based on the Aston group alone indicated that there was still no difference between them and the Engineers on the Engineering texts, while they did better than the Engineers on the Development texts. Straight linguistic proficiency is not enough.

A rather more complex explanation involves linguistic proficiency, background knowledge, and the rank order of text difficulty. On the Gap-filling test, in both Study 1 and 2, 'Turbines' was the easiest text overall, with 'Electrolytes' second. These are, of course, the two Engineering texts. One could argue that on linguistically easy texts, groups at the same level of proficiency would score about the same. As the linguistic difficulty of the texts increased, scores based on proficiency would tend to drop away, and be replaced by scores achieved through background knowledge, if that was available. Thus on an 'easy' text, like 'Turbines', groups at the same level of linguistic proficiency, e.g. ENG, DAFE and S&M, would score roughly the same (the L/A group would score higher). On linguistically more difficult texts, e.g. 'Polanyi', ENG and S&M would be left behind. This does seem to explain some results. On 'Turbines', exact scoring, the scores were 30 (S&M), 32 (ENG), 33 (DAFE), and 43 (L/A).

However, neat though such an explanation may be, it cannot cover all the data. On Gap-filling, 'Electrolytes' was an easy text; on Short Answers, however, it was comparatively difficult. Moreover, if it was linguistic simplicity that made 'Electrolytes' easy on Gap-filling, one would expect a high correlation between scores on that text and on BA and ELTS. Examination of Table 6 shows that this was not the case. On both Short Answer and Gap-filling, correlations for this particular text were lower than for any other.

Conclusion: The evidence is that background knowledge has a significant effect. On the evidence of Study 2 scores, however, it does not seem to operate symmetrically; Engineers were disadvantaged by being tested outside their speciality - the DAFE group were not. No satisfactory explanation has been found for this asymmetry.

Study 3: The test-tasks used in Study 1 and 2 had certain advantages. Both Gap-filling and Short Answer tests are easy and quick to construct and administer, and reasonably easy to score. However, they are both open to charges of subjectivity. Pre-tests with suitable native-speakers, including subject specialists, and item-analyses of their responses, would help to make the tests less subjective. However, such procedures would, in effect, complicate tests whose major justification is their simplicity. Moreover, they would not help to improve the construct or face validity of such tasks used as proficiency tests.

We therefore decided to conduct a study using groups of students similar to those who took part in Studies 1 and 2, but replacing the

original test tasks with a module (M1) of the British Council ELTS test (1). Using part of the ELTS test had the following advantages:

1) the ELTS test is a sophisticated test, based on a complex, if somewhat non-empirical needs analysis (Munby, 1978).

2) it is the proficiency test used now by the British Council for assessing the language proficiency of potential scholarship students, and hence sat by a large proportion of tertiary-level foreign students wishing to study in Britain.

3) the study-skills module (M1) contains reading and study-skills material appropriate for the groups taking part in Studies 1 and 2.

Method: The tests used were (a) 3 subject-specific alternatives of the ELTS M1 Study Skills module, namely 'Social Science', 'Technology' and 'General Academic'; (b) the G1 (Reading) and G2 (Listening) modules of the ELTS test. The order of presentation was as follows:

1) G1, G2 and M1 (Social Science): at the start of the ESS course; used for placement purposes.

2) M1 (Technology): in week 6 of the 8 week course, week 2 of the 4 week.

3) M1 (General Academic): 2 weeks later; used as exit test.

Subjects: students belonged to the same general areas as before. Partly because of the constraints imposed by the experimental design, and the way the design was adapted to the administrative requirements of the ESS courses, the Engineers and the Science/Maths groups were combined on this occasion. This was justified by the evidence from Studies 1 and 2 that they are closely related groups. Only subjects who sat all three M1 tests were counted. Of these, 41 were DAFE, 35 S&M and E, and 41 L/A.

Results: Provisional results, based on band scores, are presented in TABLE 7. Since scores are still being processed, no significance values are included.

Study 3

Table 7

ELTS Scores

		G1	G2	M1 (Soc Sc)	M1 (Tech)	M1 (Gen Ac)	N
DAFE	X	5.02	5.74	4.62	5.12	5.82	41
	SD	1.049	1.215	1.254	1.654	1.105	
S/M&E	X	4.8	5.5	4.49	5.96	5.73	35
	SD	1.476	1.388	1.300	1.77	1.314	
L/A	X	5.88	6.35	5.22	5.45	6.37	41
	SD	0.914	1.2	1.280	1.219	0.873	
ENG	X	4.64	5.11	4.36	6.21	5.64	14
	SD	1.184	0.944	1.13	1.282	1.13	
S&M	X	4.9	5.76	4.57	5.79	5.79	21
	SD	1.663	1.586	1.417	2.047	1.445	

Discussion

If the hypothesis had been confirmed in all its details, each group would have scored best on the subject alternative most appropriate to its members, i.e. DAFE would have scored highest on 'Social Science', S/M and ENG on 'Technology' and L/A on 'General Academic'. Actual performance on 'Appropriate', 'Inappropriate' and 'Neutral' tests is shown in Table 8 (for L/A, there is no 'Neutral' test, although one might predict that they would do better on 'Social Science' than on 'Technology').

Table 8

Group performance on Appropriate/Inappropriate/Neutral Tests

	Approp	Inapprop	Neutral
DAFE	3	2	1
S/M&E	1	3	2
L/A	1	2/3	

With one exception, the table confirms the hypothesis and the assumptions underlying this part of the EITS test, e.g. S/M&E did best on 'Technology', the most appropriate test, second best on 'General Academic', the 'neutral' text, and worst on 'Social Science', the 'inappropriate' test. The exception is the DAFE group, who did worst on the 'appropriate' test, i.e., 'Social Science'. Apart from this, even this group conformed to the hypothesis, doing better on the 'Neutral' test than on the 'inappropriate' one.

Two explanations are immediately available to account for this deviation from the hypothesis. The first is that the 'Social Science' module is simply more difficult than the others. Table 9 shows that all 3 groups found it hardest.

Table 9

% Rank Orders (1 = easiest)

	Soc Sc	Tech	Gen Acad
DAFE	3	2	1
S/M&E	3	1	2
L/A	3	2	1

While, however, this may be true, there is another explanation. The results may be confounded by the presence of a learning effect. With the partial exception of the S&M and E group, the subjects found the first test the most difficult, the second easier, and the third easiest of all. One may, of course, entertain some doubts about such a learning effect. Many of the students had attempted the M1 module before arriving on the course; they might have been expected to be familiar with the general form and requirements. Moreover, whatever one's faith in the efficacy of either Aston or Lancaster ESS courses, one may doubt such a learning effect on such a complex test over a period, for the 4-week students, of 2 weeks. The possibility remains, however, and precludes firm conclusions being drawn about the relative difficulty of the 'Social Science' alternative. We are left with the fact that the S&M and E group performed better than the other two groups on the 'Technology' module, and this cannot be explained by a learning effect. They did this, moreover, in spite of the fact that scores on G1, G2, and the other M1 alternatives suggest that the other two groups were in general more proficient. If one separates the Engineers from the S/M and E group (as on Table 7), their superiority on the 'Technology' alternative is even more striking.

Conclusions

The three studies reported above result in somewhat conflicting results. Study 1, without significance values, indicated that the overall hypothesis was correct in virtually all respects; Study 2

suggested that Engineers were disadvantaged on texts outside their area, whereas DAFE subjects were not; Study 3 showed the opposite.

Taken together, the studies do seem to confirm the importance of background discipline. Needless to say, further research is needed to resolve the contradictions.

Notes

- 1 The authors would like to thank the British Council for making these tests available for the experiment.

Bibliography

- Alderson, J. C. and A. H. Urquhart. "The effect of student background discipline on comprehension: a pilot study". In Hughes, A. and O. Porter (eds). Current developments in language testing. London: Academic Press.
- Gregory, O.P. 1972. Fuel Cells. London: Mills and Boon.
- Hulme, B.G. "Development of off-shore turbine packages for power generation and mechanical drive." GEC Journal of Science and Technology 47.
- Latham, A.J.H. 1978. The International economy and the under-developed world, 1864-1914. London: Croom Helm.
- Munby, J. 1978. Communicative syllabus design. Cambridge: Cambridge University Press.