DOCUMENT RESUME

ED 373 556 FL 022 356

AUTHOR	Lynch, Tony
TITLE	The University of Edinburgh Test of English at Matriculation: Validation Report.
PUB DATE	94
NOTE	14p.; For serial publication in which this paper appears, see FL 022 351.
PUB TYPE	Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT	Edinburgh Working Papers in Applied Linguistics; n5 p66-77 1994
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; College Students; Comparative Analysis; *English (Second Language); Foreign Countries; Higher Education; *Language Proficiency; *Language Tests; Outcomes of Education; *Predictive Validity; *Test Validity
IDENTIFIERS	International English Language Testing System; Test Equivalence; *Test of English at Matriculation; University of Edinburgh (Scotland)

ABSTRACT

This paper reports on a two-part evaluation of the Test of English at Matriculation (TEAM) in use at the University of Edinburgh. TEAM has been used since 1987 to identify entering non-native speakers of English who are likely to be at risk linguistically and who should receive English language support. Separate samples of candidates' scores were used to assess: (1) TEAM's concurrent validity with other measures of English language proficiency, such as the English Language Proficiency Test Battery (EPTB) and the International English Language Testing Service (IELTS); and (2) TEAM's predictive validity in relation to academic outcome. The results indicate strong correlations between TEAM and existing proficiency tests, particularly with EPTB. The findings also suggest that TEAM performs predictively as well as other measures, with scores on the TEAM listening subtest being especially indicative. (MDM)

The University of Edinburgh Test of English at Matriculation: Validation Report

Tony Lynch (IALS)

THE UNIVERSITY OF EDINBURGH TEST OF ENGLISH AT MATRICULATION: VALIDATION REPORT

Tony Lynch (IALS)

Abstract

This paper reports on a two-part evaluation of the Test of English at Matriculation (TEAM) in use at the University of Edinburgh. Separate samples of candidates' scores were used to assess (1) TEAM's concurrent validity with other measures of English language proficiency and (2) its predictive validity in relation to academic outcome. These statistical comparisons established strong correlations with existing tests, particularly the English Proficiency Test Battery, and suggest that TEAM performs predictively as well as other measures, scores on the TEAM listening subtest being especially indicative.

1. Background

Since the early 1970s the University of Edinburgh's policy has been to provide in-session English tuition for non-native students who have fulfilled the linguistic entry requirement but are thought likely to gain, in terms of improved course performance, from further language support. The entry requirements vary among the faculties at Edinburgh, but most currently take IELTS 6.0, TOEFL 550 or English Proficiency Test Battery (EPTB, Version D) 40.0 as the minimum for acceptance.

TEAM is the most recent of three matriculation tests that have been used by the University at matriculation to identify students who are likely to be at risk linguistically and who should receive English language support. The first was the English Language Battery (ELBA), which was used until 1982; the second was the British Council/UCLES ELTS test, taken at matriculation in the period 1982-86, while the ELTS Validation Project was under way at the University of Edinburgh. As the project approached its end, a decision was taken by the University's English Language Testing and Tuition committee to replace ELBA (a multiple-choice test of grammar, vocabulary and reading) with a test that would also sample students' listening and writing.

TEAM was introduced for the academic session 1987-88 and piloted over two years in tandem with ELBA. It consists of four parts: a vocabulary test, a dictation test, a reading comprehension test and a writing test. In deciding whether or not to refer students for the in-session courses, their overall average score is interpreted as follows: less than 50% - at least 50 hours' tuition required; 50-59% - tuition strongly recommended; 60% and above - tuition may be recommended, depending on subtest scores. In comparing TEAM with ELBA it was therefore of particular importance to

compare the distribution pattern among the score bands used as the basis for referral (see Table 1).

Table 1. Student distribution (%) by score band: ELBA and TEAM 1987-89

ave.	ELBA	TEAM
<50	40	29
50-59	24	29
60-69	16	19
<69	20	23
	100	100

The key score bands, i.e. those interpreted as indicating that English in-session tuition is 'required' and 'strongly recommended', show a broadly similar distribution of students on the two tests (64% on ELBA and 58% on TEAM). Concurrent performances on the two tests by matriculating students during the two-year trial (n=95) showed a Spearman correlation of .81 ($p < 0.01$). The pilot study report (IALS 1989) concluded that TEAM was in general terms an adequate replacement for ELBA, yielding a similar picture of students' English proficiency.

TEAM has been in independent use as the University's matriculation test of English since the 1989/90 academic session. When advising students and staff of results, we may be asked about the relationship between TEAM and other measures, particularly the test that students have taken in their home country, and about how TEAM scores relate to academic success. A two-part study was therefore undertaken to investigate these two issues - TEAM's concurrent and predictive validity.

2. Concurrent validity

2.1 Method

Data for the study of concurrent validity was available in IALS archives in the form of the test scores of students attending our pre-sessional EAP courses over the period 1982-92 who had been required to take a test at the end of the pre-sessional for acceptance onto their subject courses (n=358). These records allowed comparison of individuals' performances on at least two tests: ELTS or EPTB (taken in Scotland to achieve acceptance onto the subject course), and ELBA or TEAM (taken at matriculation). In addition, approximately a quarter of the sample (n=80) had taken an IALS cloze reading test for EAP placement purposes.

Although all these tests were taken in September of the relevant year, it should be emphasised that this first part of our validation project cannot claim to assess strict concurrent validity, since the test data it investigated was not gathered under controlled conditions. With the exception of a cohort of students who were included in a three-way comparative study of ELTS/ELBA/EPTB for the ELTS Validation Project in 1982, the test candidates in the IALS pre-sessional sample did not take their tests on the same day. The interval between test sessions ranged from one to two weeks in the case of the EPTB, ELBA, TEAM and ELTS, and up to three weeks in the case of the Cloze test. However, as TEAM scores are interpreted in an approximate way (firstly as the individual student's average over the four subtests,

and secondly through the use of decile score bands) it was considered reasonable to aim for a broad-brush comparison with other tests. Table 2 shows the breakdown of the pre-sessional sample into inter-test comparisons.

Table 2. Inter-test comparisons in the pre-sessional sample 1982-92 (n=358)

ELTS x EPTB x ELBA	24
ELTS x EPTB	45
EPTB x TEAM	194
ELTS x TEAM	36
ELTS x Cloze	30
ELBA x Cloze	26
TEAM x Cloze	80

It will be noted that comparison figures exceed the subject total of 358, since a number of students took more than three tests. Although this pre-sessional sample contained no direct comparison of ELBA and TEAM, figures were available on students (n=95) taking both tests concurrently at matriculation in 1987 and 1988 for the TEAM pilot study (IALS 1989).

2.2 Results and discussion

Table 3. Means, standard deviations, minimum and maximum scores (1982-92)

	ELTS	EPTB	ELBA	Cloze	TEAM
mean	5.78	39.12	51.70	66.82	50.69
s.d.	0.75	7.56	14.41	20.22	11.85
min.	3.50	23.00	18.00	8.00	25.00
max.	7.00	59.00	84.00	120.00	86.00
poss.	9.00	65.00	100.00	147.00	100.00

These mean scores indicate broad similarity with the standard interpretation scale in use at British universities to compare EPTB with ELTS for acceptance on a university course, in which ELTS 6.0 is regarded as equivalent to EPTB 40.0 (and TOEFL 550). It also confirms that, taken over the five academic sessions since its initial trialling in 1987, TEAM has achieved reasonable similarity with its predecessor, ELBA.

Table 4. Pearson correlation matrix for the five tests

	ELTS	ELBA	Cloze	TEAM
EPTB	.74	.83	.84	.94
ELTS	-	72	.70	.72
ELBA		-	.93	.81*
Cloze			-	.77

($p < 0.01$ in all cases)
*source: IALS (1989)

A number of points may be made about the correlation values shown in Table 4. Firstly, although we have already drawn attention to the restricted sample size in

some inter-test comparisons, even the smallest subsample (n=24), for ELBA and EPTB shows a correlation (.83) very close to the .85 reported for a much larger sample (n=430) in Criper and Davies (1988). So these IALS pre-sessional students may be regarded as typical of the wider population of international students entering universities in Britain.

Secondly, the test that achieved the lowest correlation *vis-a-vis* the other four tests was ELTS, with figures ranging from .70 with Cloze to .74 with EPTB. One possible reason is that ELTS is the only test of the five to examine oral proficiency, through interview; it may be that performance on speaking varies among candidates in ways not reflected by their patterns of scores on the other ELTS subtests. This would in fact be the converse of the case of the two pairs of tests in Table 4 that are most similar in focus, if not format: TEAM and EPTB (testing listening, reading and writing) and Cloze and ELBA (testing grammar, vocabulary and reading); these pairs have high correlations - .94 for EPTB/TEAM, and .93 for ELBA/Cloze. Further possible weakening influences on correlations with ELTS are the low reliability of the interview module, commented on in the ELTS Validation Report (Criper and Davies 1988), and potential inconsistencies between performances on the original five-module ELTS and the revised four-module IELTS, introduced in 1989.

Cross-tabulation of scores allows us to confirm the existing interpretation scale of EPTB and ELTS, and to extend it to include TEAM and the Cloze, as shown in Table 5.

Table 5. Comparison across test score bands

TEAM	ELTS	EPTB	Cloze
80%	7.5	55.0	110
70%	7.0	50.0	100
60%	6.5	44.0	85
50%	6.0	40.0	70
40%	5.5	38.0	60
30%	5.0	36.0	50
20%	4.5	34.0	40

Two caveats are in order here, since there is a risk that the score interpretation in Table 5 will be seen as in some sense the 'principal result' of this investigation of concurrent validity. Firstly, we have already emphasised the restricted sample size available for some inter-test comparisons, even though we know that results from the smallest do bear comparison with those of the larger ELTS Validation Project sample. Secondly, the reader / user of the interpretative table should bear in mind when converting one test into another that, with the exception of the Cloze, the result of all the tests in this study takes the form of an overall score combining marks on a number of subtests; this inevitably conceals what may be markedly different patterns of achievement on the subtests, which need to be taken into account in assessing a student's ability to carry out the various academic tasks that postgraduate courses demand.

However, since the purpose of TEAM is diagnostic, to evaluate likely need for in-session language support, and not to act as a pass/fail criterion for acceptance onto a course, these results suggest that TEAM stands up well to detailed comparison with

other measures of international students' English. In particular, its high correlation with EPTB (.94) indicates a firm basis for direct comparison of performances on those two measures.

3. Predictive validity

3.1 Establishing criteria

Having discussed the extent to which TEAM scores reflect achievement on other language tests, we now turn to the issue of predictive validity. In so doing, we seek an answer to the other question we are sometimes asked by academic staff, which might be paraphrased as 'What do TEAM scores tell us about how well this student will do on our course ?' Before considering the details of this second part of our study, it is worth briefly reviewing some of the main problems in establishing predictive validity.

The first is the question of what criterion to select as a basis for measuring academic success. One might make a simple two-way distinction of Pass or Fail, but this would blur the gradations of academic performance that are an established part of the British system of percentage marking. It would also inevitably conceal differences between the student who achieves Distinction and one who scrapes a borderline pass.

More specifically, where a postgraduate course has three possible outcomes, as is the case with most courses at Edinburgh, of Pass at Master's level, Pass at Diploma level and Fail, there arises the issue of how to categorise the Diploma Pass. Should we regard it as a form of failure and take the Master's Pass as the only real success? Or should one accept the arguments of the departmental staff who regard a Diploma Pass on their course as a mark of solid achievement and a Master's Pass as a bonus? Our experience is that staff attitudes to the status of the Diploma Pass varies among (and also within) departments.

Thirdly, any comparison of language test scores with outcomes in a range of academic fields involves the assumption that all the departments in an institution are working to the same academic standards. Our purpose here is to assess the predictive validity of an English language test, rather than to attempt an academic audit, and we will therefore assume that a Diploma pass in one academic subject is the same as one in another. If this is a fiction, it seems to us a necessary one.

3.2 Method

The data for analysis comprised the TEAM scores of students matriculating at Edinburgh in the three sessions 1989-90, 1990-91 and 1991-92 for one-year taught postgraduate courses, primarily Diploma/M.Sc courses (n=291). There were two main reasons for our decision to focus on these students, rather than on those beginning research degrees. The first was related to the diagnostic aim of TEAM; the University of Edinburgh has always assumed that students on 12-month courses run a greater risk of failure than those taking research degrees, which involve a different and perhaps less intensive pattern of study, and certainly a longer period in which to remedy any language weaknesses. The second reason was a practical one: at the time of our study, data on Diploma./M.Sc. outcome was available for the three annual

intakes after 1989, whereas very few of the research students first matriculating in 1989 would have had time to complete their research.

In order to gather data on outcome, a questionnaire was sent to Faculty officers dealing with postgraduate students. The form comprised a simple checklist for each academic session, listing TEAM candidates from the relevant Faculty; staff were asked to indicate one of four outcomes - Master's Pass, Diploma Pass, Fail, or research; a final column provided space for 'other comments'. Table 6 summarises the distribution among the three taught-course outcomes.

Table 6. Overall M.Sc. success / failure rates of TEAM candidates 1989-92

M.Sc. pass	Diploma pass	failure	TOTAL
230 (79%)	34 (12%)	27 (9%)	291

The 9% failure rate may appear high and it is important to make clear precisely what we have included under that heading. The Faculty responses to our questionnaire provided a variety of comments on non-completion as opposed to a Fail: e.g. 'withdrew before resits', 'returned home because of family problems', 'discontinued', 'withdrawn during study'. We are also aware of cases where students started an M.Sc. course but experienced such difficulties with English that they left the University after the first few weeks of the Autumn Term; officially there was 'no record' of their participation in the course.

Failure is a sensitive issue in any area of life and there are obvious pressures on departments not to fail students: technically, a student who withdraws (or is withdrawn) from a course has not 'failed', but withdrawal can be taken as an indication that an individual would have failed. As Criper and Davies (1988) point out, even when medical or family reasons for non-completion are cited, it may well be in order to save embarrassment, either personal or institutional. Given the inevitable uncertainties of explicit and implicit failure and the possible hidden influence of language problems on non-completion, we decided to adopt a broad definition of 'failure' in this study, and to include in that category both outright Fails and non-completions. Although there might be objections that this has exaggerated the failure rate, it is clear from Table 7 that our categorisation has in fact resulted in an overall distribution almost identical with that found in the ELTS Validation Report:

Table 7
Overall success / failure rates on Master's courses:
ELTS validation sample (n=502)

M.Sc. pass	Diploma pass	failure
81%	12%	7%

We can assume, then, that the decision to combine 'Fail' and 'non-completion' has not skewed the pattern relative to ELTS; this will allow us to compare the predictive validity of the two tests with some confidence.

3.3 Results and discussion

Table 8. TEAM: means, standard deviations, minimum and maximum scores Master's course sample 1989-92

	Vocab.	Dict.	Reading	Writing	Ave.
mean	53.38	63.81	51.68	63.76	59.62
s.d.	14.31	21.26	25.94	16.08	15.03
min.	6.00	9.00	0.00	15.00	14.00
max.	100.00	100.00	100.00	100.00	99.00

The overall TEAM average score is higher than the 50.69 figure in the concurrent validity sample (Table 3), but this can be explained by the differences between the two populations: the students whose scores are presented in Table 3 had been required to attend pre-sessional tuition and also included research students, while the figures in Table 8 are those of Master's course students attending the matriculation test of English, the majority of whom were not required to take tuition prior to subject course entry. So one would expect the students in the matriculation sample to produce higher scores overall.

When the overall average TEAM scores are banded by deciles and compared with outcome (Table 9), we find some initial evidence of a relationship between language proficiency as measured by the matriculation test and success on the departmental course.

Table 9. Distributions of TEAM Average scores and academic outcome Master's course sample 1989-92

TEAM Ave.	Master's pass		Diploma pass		failure		Total
<30%	1	(33%)	-		2	(67%)	3
30-39%	8	(50%)	3	(19%)	5	(31%)	16
40-49%	32	(68%)	9	(19%)	6	(13%)	47
50-59%	58	(75%)	11	(15%)	8	(10%)	77
60-69%	55	(81%)	8	(12%)	5	(7%)	68
70% or more	76	(95%)	3	(4%)	1	(1%)	80
overall	230	(79%)	34	(12%)	27	(9%)	291

The failure rate decreases with increasing English proficiency, falling from 67% at TEAM scores below 30% to a mere 1% of failure at TEAM scores of 70% or more. Conversely, Master's pass rates rise from 33% for those scoring below 30% on TEAM to 95% for those achieving above 69% on TEAM. The watershed of better-than-average chances of passing at Master's or Diploma level is around TEAM 60%. In considering the general pattern of the relationship between TEAM results and success or failure on the subject course, we might also look at the test/outcome findings of the ELTS Validation Study (Table 10).

Table 10. Distributions of overall ELTS scores and academic outcome: ELTS project sample (n=720)

overall band	failure
up to 4.0	57%
4.5	33%
5.0	33%
5.5	30%
6 0	19%
6.5	6%
7.0	5%
mean failure rate	22%

It is important to note that in Table 10, the apparently very high 'failure' rate was based on a definition of failure that encompassed both Fails and Diploma passes, and so in order to compare these findings with those of our own predictive study, we have to combine the relevant means in Table 9 - 12% Diploma passes and 9% failures, giving 21%. So again there is a close similarity between the ELTS findings and those for TEAM. Criper and Davies (1988: 92) concluded that ELTS 6.0 could be regarded as 'the dividing line between an acceptable and unacceptable risk of failure'. For our Master's course sample it appears that the cross-over point is in the 50-59% TEAM band and that this applies both to the chances of getting a pass at Diploma level and also to the likelihood of failure (whether outright Fail or non-completion). The evidence is, then, that the level of English proficiency below which a student stands an above-average chance of not passing the degree for which they are registered is 6.0 on ELTS and 50-59% on TEAM.

Overall, then, the evidence of Tables 9 and 10 is that the pattern of performance in the Edinburgh TEAM sample was similar to that in the larger ELTS sample: one in five non-native students ran a risk of not getting their Master's degree.

Having discussed the global pattern of TEAM average scores, we now consider performance on the four TEAM subtests. The figures in Table 11 suggest that some parts of TEAM perform better than others as predictors of outcome.

Table 11. 'Failures' by TEAM subtest bands (all figures %)

	Vocab	Dict	Read	Wri
<30	50	50	15	25
30-39	14	16	13	0
40-49	10	20	14	13
50-59	10	6	5	13
60-69	5	4	3	13
70 or more	4	5	5	4

The vocabulary test and the dictation test both produce clines of increasing scores and falling rates of failure. However, the rather flat spread of scores on the reading subtest means that it does not discriminate sufficiently at lower levels; the chances of failure are not differentiated among reading scores up to 50%. On the other hand, the 50-59% band does appear to mark a division, with a decline in failure rates with

TEAM scores above 50%. The writing test produces a level bunching of students who performed relatively well on that subtest (40-69%) but nevertheless failed or did not complete their degrees.

Table 12. Mean TEAM subtest scores (%) by outcome

	Master's pass	Diploma pass	failure
Vocabulary	57.03	48.07	47.70
Dictation	67.80	57.65	47.70
Reading	54.01	37.71	40 44
Writing	67.54	61.47	58.52
Ave	63.38	53.29	49 59

On the evidence of the results in Table 12, the dictation subtest produces the clearest differentiation among the three outcomes, with a mean interval of some 10%. The vocabulary section of TEAM appears not to discriminate sufficiently between Diploma Pass and failure. Scores on reading are erratic and those on the writing subtest have a restricted range.

Table 13. Pearson correlations: TEAM subtest scores with outcome

Vocabulary	0.24
Dictation	0.31
Reading	0.22
Writing	0.19
Average	0.32

($p < 0.01$ in all cases)

Dictation emerges as the subtest with the closest association with students' eventual success on their course, and the correlation of .32 for the association between Average and outcome is comparable with those reported in the ELTS Validation Report of .34 between outcome and ELTS taken at home, and .35 between outcome and ELTS retaken in Britain. The extent to which each of the subtests can be said to have contributed to eventual success is shown in Table 14. The dictation score is the only statistically significant coefficient.

Table 14. Regression analysis - logistic estimates (depend. variable: 1=M.Sc./Dip. Pass; 0=failure)

	coefficient (*t* tests)
Vocabulary	.0088
	(.866)
Dictation	.0223
	(3.259)*
Reading	-.0012
	(-.224)
Writing	.0029
	(.358)

* significant at the 1% level

The fact that the dictation subtest performs best as a predictor is of particular interest. One might have expected that, since the assessment of performance on postgraduate courses is based predominantly on written assignments (essays, projects, examination and dissertation), it would be measures of text skills (reading and/or writing) that reflected subject course performance better than a test of listening comprehension. Foreign language use being complex rather than simple, it seems likely that the link between listening and outcome is an indirect one. It is evident to subject staff and language tutors alike (and to the students themselves) that individuals who, from the very beginning of the first term of a one-year taught course, have difficulty in understanding lecturers are likely to fall behind in their grasp of the factual and conceptual content of the course and may never catch up in what is a relatively short and intensive period of study.

From the wider perspective of research into second language acquisition (e.g. Faerch and Kasper 1986; Rost 1990), listening is regarded as a powerful source of input to the acquisition process, provided that the messages are comprehensible. But second language users who are unable to cope with the pace and complexity of lectures may experience a multiplier effect - losing confidence in their ability to understand spoken English, therefore becoming more anxious about lecture comprehension and note-taking and all the while appearing to lose ground to their peers who are able to follow the language and content of the lectures. More generally, the comprehension barrier can cut them off from the host culture, and this may in turn contribute to the loneliness and homesickness that can later surface as 'family' and 'medical' reasons for withdrawal from the course. Interestingly, there is North American evidence that aural comprehension ability exerts a strong influence on academic success even in the first language; Oxford (1993) cites an extensive survey by Conaway (1982), which found that poor listening comprehension was a more significant factor in academic failure than poor reading comprehension and low academic aptitude.

Our analysis of the TEAM scores suggests that, as in the L1 case, listening skills tapped by the dictation subtest may be a key element in academic success for international postgraduates on one-year courses. However, it could be that what enables students to respond well to the specific demands of a dictation is a more

general language proficiency factor and not only aural comprehension; proponents of dictation such as Oller (1976, 1979) have long argued that a dictation test is an effective probe of the learner's expectancy grammar, providing insight into general language competence.

The measurable predictive power of TEAM overall, like that of other British language tests, is relatively limited. Criper and Davies (1988) established a correlation of approximately .3 between overall ELTS scores and academic outcome, and described that as typical of similar investigations of predictive validity. It is true that a number of North American studies (reviewed in Graham 1987) have reported correlations as high as .5 between English proficiency scores (usually TOEFL) and academic performance, but the measure of the latter has tended to be the student's first-semester grade-point average (GPA), rather than performance later in their course career. It may also be significant that the US studies have tended to focus on undergraduates rather than graduates, since the demands placed on non-native users by the two types of degree are likely to be different.

However, to conclude that TEAM accounts for some 10% of the variance in academic performance across the sample as a whole does not exclude the possibility that (in)ability in English may represent much more than 10% of the difficulty that linguistically weak students encounter in following their degree course. 'It is feasible that the low correlations between language level and final outcomes mask a non-linear relationship: that the effect of language increases steeply at lower levels' (Criper and Davies 1988: 91).

4. Conclusions

On both issues investigated in this study, concurrent and predictive validity, TEAM bears comparison with established and more widely used tests. We have found reasonable grounds for confidence in the interpretation of TEAM scores in terms of its concurrent validity relative to other measures of academic English proficiency, particularly EPTB. Since EPTB is offered as an alternative to IELTS to pre-sessional students studying in Edinburgh for acceptance onto a university course, the evidence of a close relationship between EPTB and TEAM is an especially valuable finding of this study of concurrent validity.

As a predictive instrument, TEAM performs on a par with the original version of ELTS, achieving a correlation of .32 between overall TEAM average score and academic outcome. We have stressed that this is an association across the whole population, encompassing a wide range of ability; a reasonable case can be made that for students with relatively weak English - those likely to be identified as requiring language tuition - the influence of language ability (and listening in particular) will in fact have a substantially greater influence on their particular performance on a course than is apparent from the 10% global figure.

In both the validation studies reported here, we have compared TEAM's performance with the original version of ELTS. We await with interest the publication of the ongoing UCLES validation study of IELTS (Ferguson and White, in progress), which will allow us to relate TEAM more closely with the current version of the test.

Although TEAM appears to do as well as other tests, there is a need to revise some of its subtests; while the TEAM dictation score acts as a reasonable predictor of academic outcome, our analysis has demonstrated that the reading and writing subtests require adjustment in order to raise their predictive power. A revised version of TEAM has now been introduced and we intend to evaluate the effects of those revisions in a future study.

Acknowledgment

I would like to express my thanks to the staff in the Registry and Faculty Offices of the University of Edinburgh for their assistance with the data collection for the predictive validation component of this research, and to two EWPAL readers for their helpful comments on an earlier version of this paper.

References

Conaway M. 1982. 'Listening learning tool and retention agent'. In A. Algier and K. Algier (eds.) Improving Reading and Study Skills. San Francisco: Jossey-Bass.

Criper C. and A. Davies. 1988. Edinburgh ELTS Validation Project Report. British Council / University of Edinburgh.

Faerch C. and G. Kasper. 1986. 'The role of comprehension in second language learning'. Applied Linguistics 7/3: 257-74.

Ferguson G. and E. White. (in progress) A Predictive Validity Study of IELTS University of Cambridge Local Examinations Syndicate / University of Edinburgh.

Graham J. 1987. 'English language proficiency and the prediction of academic success'. TESOL Quarterly 21/3: 505-21.

IALS. 1989. English Language Testing and Tuition Report: TEAM Pilot Study. Mimeo University of Edinburgh.

Oller J. 1976. 'Evidence for a general proficiency factor: an expectancy grammar'. Die Neueren Sprachen 2: 165-174.

Oller J. 1979. Language Tests at School: a Pragmatic Approach. Longman: London.

Oxford R. 1993 'Research update on teaching L2 listening'. System 21/2: 205-11.

Rost M. 1990. Listening in Language Learning. Longman: London.