



TOEFL[®]

Monograph Series

MS - 6
NOVEMBER 1996

*A Review of Research into
Needs in English for Academic
Purposes of Relevance to the
North American Higher
Education Context*

Alan Waters

**A Review of Research into Needs in English for Academic Purposes
of Relevance to the North American Higher Education Context**

Alan Waters

**Educational Testing Service
Princeton, New Jersey
RM-96-7**



Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 1996 by Educational Testing Service. All rights reserved.

No part of this report may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Violators will be prosecuted in accordance with both U.S. and international copyright laws.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, GRE, SPEAK, TOEFL, the TOEFL logo, TSE, and TWE are registered trademarks of Educational Testing Service.

SAT is a registered trademark of the College Entrance Examination Board.

Foreword

The TOEFL® Monograph Series features commissioned papers and reports for TOEFL 2000 and other Test of English as a Foreign Language program development efforts. As part of the foundation for the TOEFL 2000 project, a number of papers and reports were commissioned from experts within the fields of measurement and language teaching and testing. The resulting critical reviews and expert opinions have helped to inform TOEFL program development efforts with respect to test construct, test user needs, and test delivery. Opinions expressed in these papers are those of the authors and do not necessarily reflect the views or intentions of the TOEFL program.

These monographs are also of general scholarly interest, and the TOEFL program is pleased to make them available to colleagues in the fields of language teaching and testing and international student admissions in higher education.

The TOEFL 2000 project is a broad effort under which language testing at ETS will evolve into the 21st century. As a first step in the evolution of TOEFL language testing, the TOEFL program recently revised the Test of Spoken English (TSE®) test and announced plans to introduce a TOEFL computer-based test (TOEFL CBT) in 1998. The revised TSE, introduced in July 1995, is based on an underlying construct of communicative language ability and represents a process approach to test validation. The TOEFL CBT will take advantage of the new forms of assessments and improved services made possible by computer-based testing while also moving the program toward its longer-range goals, which include

- the development of a conceptual framework that takes into account models of communicative competence
- a research agenda that informs and supports this emerging framework
- a better understanding of the kinds of information test users need and want from the TOEFL test
- a better understanding of the technological capabilities for delivery of TOEFL tests into the next century

It is expected that the TOEFL 2000 efforts will continue to produce a set of improved language tests that recognize the dynamic, evolutionary nature of assessment practices and that promote responsiveness to test user needs. As future papers and projects are completed, monographs will continue to be released to the public in this new TOEFL research publication series.

TOEFL Program Office
Educational Testing Service

Abstract

This survey examined reports of research into needs in English for Academic Purposes (EAP) of relevance to the North American higher education context, in order to assess requirements for further, test development-related research. First, a wide range of North American studies were summarised and critically analysed in terms of findings related to the four skills, writing, listening, speaking, and “other” needs. The research methods employed were also reported and evaluated. Overall conclusions were drawn and implications for further research were adumbrated for each of the studies.

Next, a number of major, in-depth British studies of relevance were also summarised and critically scrutinised. All of these reports were concerned with the formulation of a picture of EAP needs as a basis for the development of a test of proficiency in English for Academic Purposes. A number of overall conclusions were once again drawn.

The survey as a whole showed that there was no existing body of research that could form an adequate basis for the development of a test of EAP of relevance to the North American higher education context. Based on a critical analysis of the reports surveyed, a number of proposals for a programme of further research were made.

Acknowledgments

I wish to gratefully acknowledge the initial guidance provided by Gary Buck for the writing of this paper; Isaac Bejar, Joan Carson, Phil Everson, Ann Johns, and John Swales for their penetrating and illuminating comments on the review version; the feedback provided by members of the Lancaster University Communication for Academic Purposes Research Group on a draft of the review version; Charles Alderson and Caroline Clapham, for their very helpful comments and for providing copies of many essential readings; and John McGovern, for several stimulating discussions about the paper's concerns.

Table of Contents

	Page
Introduction.....	1
Research Reports: Review of Findings and Research Methods.....	2
Criteria for Inclusion	2
Structure.....	3
Studies in North America	4
Four Skills Needs	4
Writing Needs	9
Listening Needs.....	22
Speaking Needs.....	27
Other Needs	34
Studies in the United Kingdom.....	39
In-depth Surveys	39
English Language Testing Service Revision Studies.....	45
Discussion and Conclusions	52
Summary of Research Findings	52
Summary of Research Methods	53
Proposals for Further Research.....	55
Overall Conclusion.....	60
Bibliography.....	61
Appendix A	
Summary of North American Research Reviewed	69
Statistical Analysis	72

List of Tables

	Page
Table 1 Summary of Research on Four Skills Needs.....	8
Table 2 Summary of Research on Writing Needs.....	20
Table 3 Summary of Research on Listening Needs	26
Table 4 Summary of Research on Speaking Needs	33
Table 5 Summary of Research on Other Needs	37

List of Acronyms

BALEAP	British Association of Lecturers in English for Academic Purposes
EAP	English for Academic Purposes
ELTS	English Language Testing Service
ELTSREV	English Language Testing Service Revision Project
EPTB	English Proficiency Test Battery
ESP	English for Specific Purposes
GI	A general (that is, non-subject specific sub-test) in the ELTS test
IELTS	International English Language Testing System
ITA	International Teaching Assistant
JYA	Junior Year Abroad (a scheme whereby international students spend their junior year studying at a British university)
M1	One of the subject specific sub-tests of the ELTS test
NNS	Nonnative Speaker of English
NSTA	Native Speaker Teaching Assistant
SPEAK	An institutional version of the Test of Spoken English
TEAP	Test of English for Academic Purposes
TOEFL	Test of English as a Foreign Language
TSE	Test of Spoken English
TWE	Test of Written English

Introduction

The purpose of this paper is to critically review research into needs in English for Academic Purposes (EAP), particularly in relation to the North American context.

This report comprises three main sections. In the first, I describe and discuss a selection of North American research reports. In the second, I review a number of relevant British studies. In the final section, I draw a number of overall conclusions regarding the research findings and methods reviewed in the earlier sections, and I make recommendations for the scope, approach, and overall focus of a programme of further research.

Research Reports: Review of Findings and Research Methods

Criteria for Inclusion

The following criteria have been used for selecting the studies included in this review:

Setting. I have tried to look for close relevance in terms of the setting of the research. The majority of the studies were carried out in the North American higher education context, and the rest in the British higher education setting.

Scope. The studies focus on areas of need that permeate the EAP study curriculum as a whole, for example, the four skills¹, vocabulary, and note-taking. I have tried to include as wide a spread as possible of such studies, in terms of the aspects they focus on, in order to be as comprehensive as possible. Many of the studies have also been selected because they focus on the most crucial stages in the higher education system as far as the overall purposes of this review is concerned, that is, the undergraduate and postgraduate entry points. Also, as far as possible, the studies are concerned with needs in subject-specialism settings rather than in English as a Second Language (ESL) classes, and with students rather than academic professionals.

Importance. I have tried to include all the North American and British EAP studies that have had a widespread influence on the field. Such studies are obviously important in terms of understanding the thrust of subsequent research, as much as for the light they throw on EAP needs.

Focus. The studies are concerned with the identification of actual need, not just descriptive analysis. For example, much of the work on genre analysis (Swales, 1990), critical language awareness (e.g., Clark, Fairclough, Ivanic, & Martin-Jones, 1990; Benesch, 1993), has therefore had to be excluded, since research that shows these to be major areas of need (versus research aimed at refining and developing the basic concepts) is, unfortunately, so far mostly lacking. However, the issues raised by research of the latter kind have informed the framework of thinking behind this review. (See, for example, Proposals for Further Research below.)

Research-based. All the studies reviewed involve actual research of one kind or another, rather than just speculation or theorising. Thus, despite the stimulating nature of many of the ideas-based reports in the EAP literature, they have had to be excluded from the main part of this review. Also, despite their usefulness, I have excluded reports in the form of summaries of other research in order to be able to more directly assess the value of the work reported on.

¹Reading, writing, listening, and speaking.

Validity. This review is obviously far from exhaustive. The items that have been included in this report are only a selection from among the ones I consulted. I excluded many in order to avoid amassing greater detail than strictly necessary, as I believed including items that did not significantly advance understanding would simply cloud rather than clarify the picture. I am confident that the resulting coverage has been sufficient to accurately represent the true nature of the overall picture.²

Structure

The section that follows contains accounts and discussions of first the North American and second the British research reports that I have selected for review. Both main parts are divided into a number of subsections, according to the main focus of the reports. Within each subsection, I have followed a mainly chronological order, as I believe this helps to show how subsequent papers have built on earlier ones. In dealing with each of the reports, my procedure has been to first summarise the purpose and context, then describe the main features of the research method and the findings, and, finally, discuss the significance of the work in terms of the overall concerns of this review. Each section ends with a summary chart and a set of overall conclusions. (A comprehensive summary chart is also included in Appendix A, as are analyses of the data in terms of key variables.)

²I did not interpret my brief as extending to a review of L1 studies, since the term “EAP” normally refers to the L2 situation only. I would, however, concur with the views of those reviewers who felt that to survey this literature would also be a valuable exercise (if time permits).

Studies in North America

Four Skills Needs

Ostler (1980) surveyed the views of 133 international students at the University of Southern California about their EAP needs in terms of major uses of the four skills and their assessment of their oral-aural proficiency, via sentence-combining and summarising tasks and a questionnaire. The majority of the sample were from engineering, business studies, the hard sciences, and the soft sciences. Of the respondents, 72% were undergraduates (mainly freshmen and sophomores), and the remainder were postgraduates (mainly master of arts and master of science candidates).

Findings regarding needs were analysed in terms of (a) majors and (b) class standing.

Findings by major showed that the top three needs were for reading textbooks, taking notes in class, and asking questions in class. There was also a significant need to use English for writing research papers and reading academic journals. However, there was also significant variation across majors in the needs for different skills.

Findings by class standing showed considerable divergence. Only scores for reading text books and taking notes in class were approximately equivalent. Undergraduates indicated greater need for skills in taking multiple-choice examinations, writing laboratory reports, and reading and constructing graphs and charts. Postgraduates put most emphasis on reading academic journals and papers; giving talks in class and participating in panel discussions; writing critiques, research proposals, and research papers; and discussing issues and asking questions in class.

Regarding perceptions of oral-aural proficiency, students generally rated themselves as better able to make themselves understood in nonacademic than in academic settings. As Ostler (1980) points out, this may be because the nonacademic settings require only a mainly formulaic use of language, whereas academic settings require more creative uses of language.

The results of the sentence-combining and summarising tasks showed a need among undergraduates for better sentence-combining and summarising skills on the part of both undergraduates and postgraduates.

Ostler's research is the earliest study of its kind that I have had access to, and subsequent studies are in many ways developments of and departures from it. It was one of the first major attempts to undertake a systematic study of EAP needs in the North American context, and is therefore something of a landmark. However, with the benefit of hindsight, a number of drawbacks in the research design can be detected.

1. It is not reported whether the categories of "academic skills" used in Ostler's questionnaire were based on any prior empirical investigation of what tasks of this kind actually exist. This is not to say that the ones used were necessarily inadequate. However, it is possible that other, equally or more important uses of English exist, but were not surveyed. There is also

the question of to what extent the respondents shared the researcher's understanding of what was meant by the terms used to refer to the skills. Both of these issues resurface as important themes in later research (see, e.g., Horowitz, 1986a, 1986b).

2. No controls appear to have been used regarding the sample surveyed. The findings may therefore be unrepresentative of the population as a whole. (It should also be noted that the return rate for the research instruments is not given.)
3. The main research instrument was a questionnaire. The details of the questions are not provided, so it is not possible to assess their quality. Also, as with all questionnaire-based research, it is impossible to assess the truthfulness of responses.
4. The questionnaire findings are difficult to interpret. Although it is possible to detect some overall trends in the figures, as Ostler indicates, there are also significant differences in the profile of needs from one major to another. It is also not clear whether students are reporting on their perceptions of needs in terms of their frequency or in terms of their importance. A need may be infrequent but very important (e.g., writing examination answers).
5. The research also involved two performance tasks. No rationale is given for their structure or content. There do not appear to have been any controls on the methods by which the students completed the tasks. It is therefore difficult to assess the significance of the results obtained in this part of the research.
6. There was no triangulation. The evidence is therefore uncorroborated by other sources.

Despite these drawbacks, however, Ostler's (1980) research is useful as a starting point for tracing the gradual deepening of enquiry that occurred during the rest of the decade.

Johns (1981) reports on the results of research into views about EAP needs at San Diego State University involving 200 randomly selected classroom instructors (10% of the total faculty). In this study, 140 completed Academic Skills Questionnaires were returned, with a "good" distribution across departments and disciplines and with adequate representation from the two subject areas with the highest proportion of international students (engineering and business studies) and in terms of "lower division" and "upper division/graduate" classes.

Findings indicated that reading and listening were regarded as the most important skills for all levels. Faculty were also surveyed for their views on whether the students' primary need was for "general knowledge of English" or "English specific to the discipline." The preference was for the former, except for engineering.

Many of the same strictures levelled at Ostler's research also apply to Johns' (1981). Ostler's questionnaire is reproduced in Johns' work and seems sound enough, and the sample would appear to have been reasonably representative. But the research model closely resembles Ostler's in other respects, and therefore has similar drawbacks. Furthermore, the fact that some skills were rated higher than others by faculty does not mean that the ones rated lower were necessarily seen to be of less importance in absolute terms, although Johns appears to draw this conclusion. Respondents were asked to number the skills in order of importance — that is, in relative terms. This does not necessarily mean that a skill ranked last was regarded as unimportant. It would therefore be unwise to conclude from Johns' research that speaking and writing were not also regarded as important skills.

On the other hand, Johns' (1981) research provides some corroboration — from the faculty rather than from the students — of Ostler's (1980) findings in the sense that reading is singled out as the first priority in both studies. There may also be a connection between their two second-ranked categories of listening and note-taking respectively. (In light of the former, however, it is somewhat puzzling that very few subsequent studies focus on reading; (but see Christison & Krahnke, 1986). Johns' finding regarding general English is also worth noting as a contribution to the question of to what extent EAP is more specialised than are other forms of English.

In an attempt to compensate for what they saw as the problems of objectivity, sampling, and validity in earlier studies, Christison and Krahnke (1986) used the individual interview as the basis for their research into student perceptions of EAP needs. The subjects were 80 international students studying at five universities in the United States. The undergraduate: postgraduate split was 71%: 29%. The students represented 21 major fields of study, with the main groupings as follows: engineering, 25%; business studies, 25%; sciences, 14%; humanities, 12.5%; computer science, 12.5%. The authors note that "in regard to relevant characteristics, . . . the sample used in this study roughly approximates the overall international student population in higher education in the United States" (ibid., p. 66).

The results in terms of perceived frequency of skill use were as follows (all figures are percentages):

Listening: 50	Reading: 30	Speaking: 10	Writing: 10
---------------	-------------	--------------	-------------

In terms of perception of level of difficulty, the results were:

Speaking: 35	Listening: 32	Miscellaneous: 24	Reading: 9
--------------	---------------	-------------------	------------

In terms of perceptions of level of ease, the results were:

Reading: 38	Nothing: 29	Listening: 27	Writing: 6
-------------	-------------	---------------	------------

The authors point out (ibid., p. 72) that their research procedure was relatively time-consuming, less objective than a multiple-choice questionnaire would have been, and possibly more intimidating (since results were elicited by face-to-face interaction with respondents, who may have feared that their level of English was being assessed).

The most noteworthy feature of this research was that more rigour was applied to selecting the sample and eliciting data than is seen in comparable studies. Also, this research not only focused on needs but also on lacks. Thus, for example, although reading is regarded as the second most frequent need, it is perceived as the easiest, or least difficult, skill to master. It is also worth noting that this study confirms Johns' (1981) findings, and, less directly, Ostler's (1980), regarding the two most frequently needed skills (i.e., reading and listening).

Once again, however, it must be pointed out that no necessary correlation exists between frequency of use and importance. Also, although the research method used in this study is more rigorous in its direction, it is nevertheless not triangulated. Furthermore, the results are more limited in their usefulness than are previous studies, by lack of analysis in terms of variation across disciplines and between the undergraduate and postgraduate levels.

TABLE 1 **Summary of Research on Four Skills Needs**

	Author(s)	Date	Institution(s)	Subjects	N	Level	Subject Area(s)	Research Method
1.	Ostler	1980	University of Southern California	International students	133	u/g: 72%, p/g: 28%	Hard sciences, engineering, business studies, soft sciences	Survey (questionnaire and written language tasks)
2.	Johns	1981	San Diego State University	Faculty	140	lower division & upper division	Cross section	Survey (questionnaire)
3.	Christison & Krahnke	1986	5 U.S. universities	International students	80	u/g: 71%, p/g: 29%	Cross section	Survey (structured interview)

n/a = not applicable

u/g = undergraduate

p/g = post graduate

Conclusions. These studies are all concerned, directly or indirectly, with three main questions of significance to the TOEFL 2000 project:

- Do EAP needs vary significantly by class standing?
- Do they vary significantly by major?
- Do they vary significantly according to point of view (student versus faculty)?

Unfortunately, none of the reports provides conclusive answers to these questions. Although the reports share limited common ground (e.g., regarding frequency of skill use), they also show considerable divergence. Some of the other research reviewed later on in this paper throws further light

on the matter at the level of individual skill areas (writing, etc.), or in relation to other contexts (e.g., higher education in the United Kingdom), but it would appear that a reliable and comprehensive overall survey of needs in the North American situation is currently lacking. This is a lacuna in obvious need of remedy.

Writing Needs

As part of the research that led to the development of the Test of Written English (TWE®), Bridgeman and Carlson (1983) surveyed faculty in 190 academic departments at 34 United States and Canadian universities in order to investigate the types of writing skills needed by first-year undergraduates and postgraduates.

The disciplines surveyed were those with high enrolments of overseas students, as is typical of most of the research discussed in this review. (Potential drawbacks will be discussed later in this report.) The disciplines surveyed at the graduate level were business management, civil engineering, electrical engineering, psychology, chemistry, and computer science. English was the focus at the undergraduate level, on the grounds that “the [preliminary] interviews [with faculty] suggested that most writing by first-year undergraduate students was done in English courses” (ibid., p. 18).

The main research instrument was a questionnaire, which was developed as a result of preliminary consultations with faculty, writing experts, and so on. The questionnaire consisted of questions on types of writing task demands, their importance, writing assignment evaluation criteria, writing problems of native versus nonnative speaking students, writing samples in the admission process, and examples of 10 possible topics for writing samples. The 10 topics were:

Type A:	Personal essay
Type B:	Sequential or chronological description
Type C:	Spatial or functional description
Type D:	Compare and contrast
Type E:	Compare and contrast plus take a position
Type F:	Extrapolation
Type G:	Argumentation with audience designation
Type H:	Describe and interpret a graph, chart, etc.
Type I:	Summarize a passage
Type J:	Summarize a passage and analyze/assess a point of view

Usable responses were received from 82% of the subjects.

Bridgeman and Carlson’s (1983) main findings were as follows:

- Writing was regarded as an important skill at the graduate level, though even more important after graduation.

-
- Even disciplines with lighter writing requirements require some writing from first-year students.
 - There was variation from discipline to discipline regarding which writing skills were regarded as most important.
 - Faculty writing evaluation criteria focus more on the discourse than on the word/sentence level.
 - Nonnative speakers were perceived to have more problems than native speakers at the word/sentence level.
 - There was disagreement among faculty as to the most suitable type of writing task. The two most popular tasks were “description and interpretation of a graph or chart” (Type H) and “comparison/contrast, with a defense of a position” (Type E). The former was the favourite of the graduate faculty surveyed, the latter of undergraduate faculty. (It should be noted that these two task types were subsequently chosen as the basis for the TWE test.)

Significant features of this research are the size of the sample, the effort made to establish sound questionnaire categories through prior research (though see Greenberg, 1986; Horowitz, 1986a; and Raimes, 1990, below), and the focus on native-speaker/nonnative-speaker(NS/NNS) comparisons. The last is important in order to take into account needs in the sense of lacks rather than just as “necessities” (see Hutchinson & Waters, 1987, pp. 55-56). Thus, although Bridgeman and Carlson (1983) show that faculty attach greater importance to discourse-level features of writing (i.e., these are the main “necessities”), they also show that it is in word/sentence-level skills that nonnative speakers are perceived to be weaker (i.e., these are their “lacks”). To be effective, a test of EAP proficiency must be informed by a picture of needs of both kinds, and therefore so must the research on which the test is based.³

However, Bridgeman and Carlson (1983) were criticised by Greenberg (1986), Horowitz (1986a) and Raimes (1990) for categorising writing tasks on the basis of data elicited by preliminary interviews and questionnaire, rather than by more empirical means (e.g., analysis of documentary evidence). Greenberg and Raimes argue that this has resulted in anomalies, such as the writing tasks in equivalent tests for native speakers being easier than those in the TWE test. As will be seen, a number of later studies attempt to supplement Bridgeman and Carlson’s findings by analysis of actual samples of writing prompts.

³Having said this, however, it is yet another matter as to whether faculty actually attach the same significance to problems of this nature in practice. For example, Wall (1981) found that there was a discrepancy between reported and actual grading criteria used by postgraduate economics faculty (also see Casanave and Hubbard, 1992, below). Research therefore needs to be extended to include a check on such matters.

Greenberg and Raimes also argue that Bridgeman and Carlson's research and the resulting test fail to distinguish adequately between the undergraduate and postgraduate levels, since six of the seven disciplines surveyed were at the graduate level, and, whereas general academic proficiency of native speakers at the undergraduate and graduate levels is distinguished by the use of the Scholastic Assessment Test (SAT®) examination and the GRE® examination, respectively, nonnative speakers of all levels take the same TWE examination (and TOEFL test). Unfortunately, this is an area on which later research has failed to throw much further light (but see Horowitz, 1986a, below).

Greenberg also took Bridgeman and Carlson and the TWE examination to task for, in her view, not having properly honoured the "situational" construct of writing (i.e. the notion that writing demands are task-specific) on which the research was based. She argues that this concept is not adequately represented in the conclusions drawn from the research outcomes and in the form of the resulting test. According to Greenberg (1986, pp. 538-539), despite the weight originally attached to the concept of situationality in writing in the research, the two most popular writing tasks that were identified by the research are used interchangeably in the TWE test, even though there was strong evidence to show that the two tasks call on different writing abilities. The latter would indeed appear to be the case. As Carlson, Bridgeman, Camp, and Waanders (1985) say, in a related report on the validation of the TWE rating procedures:

In the multidimensional scaling, Types H and E were further apart in the space than any other pair of types, suggesting that they were perceived as distinctly different tasks . . . Since departments perceived these two types as distinctly different, it seemed likely that writing samples elicited by Types H and E elicited different writing skills, as well (pp.12-13).

In reviewing Carlson et al.'s findings, Greenberg (1986, pp. 541-542) also argues that, although rater reliability was shown to be high both within and across tasks types, this does not necessarily mean that validity was high. As Greenberg points out, Carlson et al. (1985, p. 58) accept that raters could be reliably assessing something in the tasks that differs from what it is intended should be assessed. Thus, it may be inappropriate to regard the Types H and E writing tasks as equivalents, despite the high levels of rater reliability that were achieved.

However, as Carlson et al. (1985, p. 80) also point out, although "there was no indication of any important differences between the two topic types . . . it is important to remember that both types represent structured, academically oriented writing; results may have been different with a 'What I did last summer' type of topic." In other words, despite the possibility that the two tasks may be calling on different skills at a certain level, the skills are among those relevant to academic writing in general, and likely to be significantly different from those called on in nonacademic writing. Put another way, the extent to which the tasks differ from each other is probably much less significant than the extent to which they both differ from nonacademic writing tasks. From this perspective, the validity issue is therefore something of a red herring.

There is one further criticism of Bridgeman and Carlson's research to be considered. Raimes (1990) points out that the TWE test is intended to be a measure of attained proficiency, yet it uses writing tasks of a type it is assumed students will do in the future. The implication is that the research that underpinned the TWE test was focused in a similar way. This was indeed the case, as the instructions Bridgeman and Carlson give to their questionnaire respondents show: "The main objective of this questionnaire is to obtain a valid description of the kinds of writing tasks that are required of students in your department during 'typical' coursework" (1983, Appendix A, 1).

However, the focus of a needs analysis should ultimately be concerned with deriving the underlying competence required for study at the level in question, as it is this that provides the basis for coping with the study situation "performance repertoire" (see Hutchinson, Waters, & Breen, 1979, and Waters & Waters, 1992). In other words, the study of academic performance at the undergraduate entry level, for example, should be interpreted in order to identify the type of knowledge and skills that it is assumed the typical student will have mastered before entry. It is this type of competence that the subject-specialist will expect to be able to draw on and use as the basis for leading the student toward more specialised forms of communication, since it cannot be expected that the more specialised knowledge will have already been acquired. Although it may be that the types of writing tasks identified by Bridgeman and Carlson (1983) were reasonably close to the level of assumed initial competence anyway, because the focus of their research was on first-year writing, the principle of analysing samples of performance data in order to derive the assumed underlying competence does not appear to have been as explicit in their research as it should have been. (The same point can be made for a number of the other studies reviewed here, it should also be noted, and this issue is discussed below.)

Finally, it should be mentioned that Traugott, Dunkel, and Carrell (1992) attempt to rebut all of Raimes' (1990) points except her first one (i.e., the doubtful validity of the Bridgeman and Carlson's original categorisation of writing tasks). In her turn, Raimes (1992) responds to Traugott et al. by pointing out that it is this very aspect that is the most significant issue, since it is on this base that the credibility of the research as a whole rests.

Horowitz (1986a) criticises previous EAP needs analysis research methodology for its overreliance on questionnaires and the use of preconceived categories for analysing academic tasks, categories about which there is in any case no consensus. The main purpose of his study was therefore to collect and analyse actual samples of academic writing tasks, with a view to establishing an empirical taxonomy.

Out of 750 faculty members at Western Illinois University that were contacted, 36 responded by providing usable data in the form of handouts of writing and test items involving writing assignments. The report focuses on the nontest writing assignments provided (but also see Horowitz, 1986b). These comprised 54 assignments from 29 courses taught in 17 departments. All except one of these courses were open to both undergraduates and postgraduates; that course was open to postgraduates only. The spread of the data in terms of main subject areas is not reported, but it can be inferred that it was predominantly humanities- and social sciences-oriented.

Seven categories of writing task emerged from the data:

1. Summary of/reaction to a reading (9 examples)
 2. Annotated bibliography (1 example)
 3. Report on a specified participatory experience (9 examples)
 4. Connection of theory and data (10 examples)
 5. Case study (5 examples)
 6. Synthesis of multiple sources (15 examples)
 7. Research project (5 examples)
- (Horowitz, 1986a, pp. 449-452)

The data also permitted the formulation of a definition of “the generalized American academic writing task:”

Given a topic, topicless thesis statement, or full thesis statement, an indication of the audience’s expectations (in terms of what questions are to be covered and in what order they should be answered), specified sources of data, and a lexis constrained (to some extent) by all of the above, find data which are relevant to each question and then reorganize and encode those data in such a way that the reader’s expectations of relevance, coherence and etiquette are fulfilled. (Horowitz, 1986a, p. 455)

The research is based on a relatively small and not necessarily representative sample of data, as Horowitz himself acknowledges (*ibid.* p. 460). More competitive universities, for example, might make different demands. However, this research is significant in demonstrating the importance and value of direct inspection of academic needs as a basis for research. It also serves as a useful first step toward formulating an empirically based taxonomy of EAP writing tasks. It is instructive to compare Horowitz’s categories with those of Bridgeman and Carlson above, and with those of Canseco and Byrd (1989) below.

Horowitz (1986b) reports in a similar way on the findings for the examination question part of the data, focusing on one of the examination “prompt” organising variables, that of “the primary organizational instruction.” Four main types of task that students may be asked to perform are identified:

- Display familiarity with a concept
 - Display familiarity with the relation between/among concepts
 - Display familiarity with a process
 - Display familiarity with argumentation
- (Horowitz, 1986b, pp. 110-117)

However, no information is provided regarding the relative frequency or importance of the tasks. Also, as Horowitz acknowledges (1986b, p. 109), it is important to investigate further the nature of

“secondary instructions” (e.g., regarding what persona or point of view to assume, what sources to draw on, etc.; see Swales, 1982).

Because of what they claimed was a lack of research into graduate-level writing assignments in United States universities, Canseco and Byrd (1989) investigated the writing requirements of graduate course faculty in the College of Business Administration at Georgia State University. Fifty-five course syllabuses from 48 different graduate courses with the largest enrolments of foreign students were obtained and analysed.

Seven categories of writing assignments were found:

- Examinations
- Problems and assignments
- Projects
- Papers
- Case studies
- Reports
- Miscellaneous writing assignments

(ibid., p. 308)

Examinations were easily the most frequently occurring type of assignments, followed by problems and assignments, and then projects. Assigned papers and assigned reports were the least frequent. However, the authors conclude that “although an individual department might show a slight preference for a particular type of written work, students can expect many types of writing to be required in any department” (Canseco & Byrd, 1989, p. 309). In analysing writing assignment prompts (Horowitz, 1986a), Canseco and Byrd conclude that, due to the wealth of information about the writing that these tend to provide the student with, “Our analysis . . . supports Horowitz’s (1986a) insight that much of the writing required of students in U.S. universities (at least in the context of graduate courses in business) is highly structured and instructor controlled” (ibid., p. 312).

As a corollary, Canseco and Byrd (1989) point out that students need skill in interpreting syllabuses in order to know what they are supposed to do and when to seek additional information, and foreign students in particular might need to master the vocabulary used for describing the different types of written products. In addition, they stress the need for further research into the type of writing involved in examinations, given the high frequency of this assignment in their data, and in comparable data on graduate engineering courses (see West & Byrd, 1982).

Canseco and Byrd’s research is a useful supplement to the studies of academic writing requirements in the United States already reviewed. It also highlights the value of obtaining data on a first-hand basis, and the importance of distinguishing between writing needs at different levels.

Braine (1989) argues that most previous studies of academic writing assignments in United States universities were unsatisfactory because they used preordained rather than empirically researched categories of analysis and none focused specifically on the writing requirements of science and technology students. Sixty-one writing assignments from 10 randomly selected undergraduate courses (five in natural science and five in engineering) at the University of Texas at Austin were analysed using Horowitz's (1986a) categories. Instances of four of Horowitz's categories were found among these writing assignments. By far and away the most frequently occurring item was the "report on a specified participatory experience" (e.g., lab reports, "brief weekly reports and abstracts" (ibid, p. 7), interim reports on on-going research, etc.). Categories that did not occur were those of annotated bibliography, connection of theory and data, and research project. Braine concludes that since 36 of the 52 samples comprising the "report on a specified participatory experience" category took the form of lab reports, there is a need to more closely analyse what this type of writing assignment involves. He also points out (ibid, p. 10) that since instructors tend to use several different terms for the same basic category of writing assignment, this reinforces the case for inspecting actual documents as a research heuristic, rather than using questionnaires. Braine also confirms Horowitz's (1986a) finding concerning the highly controlled nature of all the writing tasks examined.

This research is useful in confirming the value both of the research methodology and the findings of earlier studies, such as Horowitz (1986a) and Canseco and Byrd (1989). It would seem worthwhile to conduct investigations along similar lines to fill in some of the relevant remaining gaps in terms of the subject areas surveyed (at both the undergraduate and postgraduate levels), and to look more closely at the types of writing tasks involve (e.g., lab reports, as Braine suggests).

Johns and Mayes (1990) point out that although studies such as Horowitz (1986a) have shown that summarising is an important type of academic writing, there has been very little research on the topic. In their investigation, 40 low proficiency (L) and 40 high proficiency (H) ESL students at San Diego State University summarised a low-intermediate level English for business text. The results were analysed in terms of various types of replications and distortions of the original.

The results showed that students in the L group copied from the original to a significantly greater extent than students from the H group. However, there was no significant difference between the two groups in terms of amount of paraphrasing. Numbers of new combinations of idea units (i.e., combining an idea that had taken several sentences in the original into one sentence in the summary) were significantly higher in the summaries of the H group. There was no significant difference however, between the two groups in terms of combinations that drew from more than one paragraph or in the production of "macro-propositions" (i.e., "accurate, writer-invented idea units which provide a generalization about a paragraph or an entire text" (Johns & Mayes, 1990, p. 265). Finally, there were no significant differences between the two groups in terms of distortions and inaccuracies.

The results would appear to indicate that knowledge of language was not the most significant factor involved. Although the H group went some way toward employing the reconstructing strategies

necessary to achieve an adequate summary, “both groups [had] difficulty condensing ideas from the original . . . [and] failed to produce appropriate macro-propositions” (Johns & Mayes, 1990, p. 265). In other words, both groups appeared to be unable to summarise adequately. Such a finding would appear to indicate an important lack on the part of ESL students, especially given the prevalence of summarising as a writing task indicated by Horowitz (1986a) (also see Kirkland & Sanders, 1991). Before this can be confirmed, however, we also need to determine the summary-writing abilities of comparable groups of native-speaker students, in order to ensure that we are not asking more of ESL students than of native speakers. Also, since Johns and Mayes’ research involved the students in carrying out a specially designed summarising task as an end in itself, one needs to also confirm that students display similar (in)abilities when carrying out summarising in the context of the normal study situation.

Casanave and Hubbard (1992) usefully summarise the main findings of recent research into academic writing in the North American higher education context as having shown the following:

First, the types of writing that college students are required to do vary greatly (the personal essay so common in ESL composition classes and placement tests is rare). Second, writing tasks are to some (as yet unknown) extent specific to discipline and to educational level (undergraduate vs. upper division/graduate) and are often highly structured and controlled by instructors. Third, faculty tend to judge content and organisation as more important than local language issues. Fourth, students in general have problems writing, but surface language problems distinguish NSs from NNSs and persist for ESL writers even at very advanced proficiency levels. Finally, writing is only one of the essential academic skills needed by college students, perhaps not even the most important one, though it increases in importance as students move to higher levels of education (p. 24).

In their own research, Casanave and Hubbard used a modified version of Bridgeman and Carlson’s (1983) instrument to survey graduate faculty at Stanford University about the writing requirements of doctoral candidates. In this study, 85 usable questionnaires were returned (about 15% of the total sent out). Responses were divided into humanities and social sciences (HSS; $n = 42$) and science and technology (ST; $n = 43$). The survey represented 28 departments, 208 classes, and 809 assignments.

Because of the low response rate, no reliable statistical evaluation of the data was possible. The results are therefore presented only as an interim resource. Also, examinations were excluded from the survey. The authors report their results in terms of the main sections from their questionnaire: Writing Tasks, Importance of Writing Skills, Evaluation Criteria, Writing Problems of Native- and Nonnative-English Speakers, Explicitness of Assignments, and Open-ended Questions.

- Writing Tasks: Of the HSS assignments, 73% took the form of critical summaries, problem-solving analytical tasks, brief research papers, and long research papers, whereas 60% of the ST assignments were of the problem-solving analytical kind.

-
- Importance of Writing Skills: All skills were regarded as of increasing importance as students moved from the first to the third year. All nine skills surveyed were regarded as important, but HSS faculty gave greater weight to skills of text-based analysis/synthesis from multiple sources, whereas ST faculty gave greater emphasis to describing, defining, and planning, and rated the skills of analysing/criticising and arguing persuasively lower than HSS faculty did. In general, HSS faculty rated all the skills more highly than ST faculty rated them.
 - Evaluation Criteria: Discourse-level criteria were ranked high, word-level and sentence-level criteria were ranked low. HSS faculty generally rated all the criteria higher than SS faculty.
 - Writing Problems of Native- and Nonnative-English Speakers: NNSs were thought to have more problems at the word/sentence level; differences were felt to be smallest at the discourse level.
 - Explicitness of Assignments: The majority of faculty said that their assignments are reasonably explicit, though ST faculty rated “topic” and “content” aspects higher than the HSS faculty rated them.
 - Open-ended Questions: Responses showed that faculty perceptions of the purposes of writing tend to vary considerably; grammar was the most commonly mentioned problem found in NNS student writing, followed by word choice/vocabulary, developing ideas, and spelling; faculty expected students to become familiar with two to four specialised journals in their field of study; and the majority of faculty do not recommend any particular writing style.

In their discussion of their findings, the authors speculate about whether a rough division can be made between text-based and activity-based writing skills, the former perhaps being more preponderant in HSS subjects, the latter in ST disciplines. Also, it appears to be the case that HSS students will be more heavily evaluated by writing from earlier on than ST students.

Although flawed by the small response rate, this research is interesting for its focus on the HSS:STE divide. It would appear that while there is a good deal of common ground within the area surveyed, there are also a number of dissimilarities in terms of what constitutes the typical writing task in each of the main areas, the types of writing skills that are valued, and the stage within the programme at which significant writing demands begin to be made. It is important to research these matters further. Casanave and Hubbard’s (1992, p. 39) research also contains interesting data on perceptions of NS versus NNS writing problems. These show that much the same level of difficulty is experienced by NS students as by NNS students in handling certain aspects of writing at the discourse level. This kind of data is essential to fairly assess the EAP demands made of NNS students.

Jenkins, Jordan, and Weiland (1993) used a questionnaire intended to build on and extend the survey conducted by Bridgeman and Carlson (1983) in order to investigate the beliefs and practices of graduate engineering faculty at Cornell, Drexel, Ohio University, Ohio State, Stanford University, and the University of Cincinnati about writing in graduate engineering education. In this study, 188 responses were received, a return rate of 31%. The focus of the questionnaire was on (a) the features of academic writing regarded as important by faculty; (b) the frequency of types of writing assignments and the weighting given to quality of writing; (c) the amount of time invested by faculty in helping to get students' theses completed; and (d) the experience of faculty in coauthoring and publishing papers in collaboration with their graduate students.

Findings were as follows:

- All the writing skills mentioned were regarded as very important.
- Approximately one-third of faculty said that they used different standards to evaluate NNS students, with a significant proportion doing so not just at the sentence level but at the overall level as well.
- The most frequently assigned writing tasks were reports that require technical writing skills, laboratory reports or descriptions of experiments, and term projects. (It should also be noted that the authors found that the quality of responses to this question to some extent bore out the criticisms of researchers such as Horowitz, 1986a, and Braine, 1989, regarding the potential for confusion and miscommunication of questionnaire categories as a means of eliciting information about types of writing tasks.)
- Quality of writing was said to be taken into consideration in determining grades.
- Advisers reported writing an average of 25% of the NNS students' theses, as opposed to around 10% in the case of NS students.
- Written comments by respondents gave further indication that there was a potential for misunderstanding in the use of terminology by English-language researchers on the one hand, and engineering faculty on the other.

This last point illustrates well one of the main difficulties of conducting research of this kind into EAP needs: the language researcher/subject-specialist divide. Jenkins et al. (1993), while acknowledging the need for clear communication across the division, appear to take a rather one-sided view of the matter in practice. They quote an engineering instructor's comment on students' writing problems as follows:

Correcting syntax mistakes is not the problem — this can be done rapidly by relatively untrained people in the speciality. The real concern is about making valid choices about

what goes into a paper, how it is organised, what the conclusions are, and which ones are valid and which should be qualified (p. 63).

In commenting on this remark, Jenkin's et al. (1993) say:

These statements simply underscore the fact that a person may have excellent knowledge of a language but be unable to identify objectively the language functions that mark logical relationships for 'clarity.'

What appears to have happened here is that the language researchers, in attempting to translate the engineer's words into their own, have distorted the underlying concepts behind the words. The engineer clearly does not see the problem as purely (or even mainly) linguistic; rather, he sees it as primarily cognitive. Jenkins et al., however, see it the other way round: they think the engineer is saying that the problem is to find appropriate expression for the thinking. In other words, in reality, the engineer is looking at the composing aspect of writing, Jenkins et al. at the encoding aspect. But the two, although related, are not the same. The irony is that the engineer appears to be more aware of the fundamental importance of aspects of writing, such as composing, than the language researchers are.

There are thus two points to be emphasised here. First, academic writing, as the remark by the engineering instructor testifies, appears to involve not only a command of formal schemata but also of content schemata (Shih, 1992). That is, one actively deploys one's perceptions in order to establish a meaningful framework of knowledge. We will return to this point in connection with the research reports on reading and listening reviewed below.

The second main point is that great care is needed to avoid miscommunication across the EAP researcher-subject specialist gulf. This calls for more than just looking for equivalent terms in which to express the conceptual categories to which we are accustomed: it also means being sufficiently open-minded to having our conceptual categories reshaped. As a further safeguard, it also calls in the first instance for a style of research that is inherently less prone to miscommunications (e.g., the direct, observational, ethnographic-oriented paradigm deployed in studies such as Jacobson, 1986, Parry, 1991, Rounds, 1987a, and Weissberg, 1993, to be reviewed below) in order to establish the basis needed for more wide-ranging studies.

Table 2 **Summary of Research on Writing Needs**

	Author(s)	Date	Institutions(s)	Subjects	N	Level	Subject area(s)	Research method
1.	Bridgeman & Carlson	1983	34 U.S. & Canadian universities	Faculty	n/a	mainly p/g	p/g: business management, civil engineering, electrical engineering, psychology, chemistry, computer science; u/g: English	Survey (questionnaire)
2.	Horowitz	1986a & 1986b	Western Illinois State University	Faculty	36	u/g & p/g	Predominantly humanities and social sciences	Survey (analysis of documents)
3.	Canseco & Byrd	1989	Georgia State University	Faculty	n/a	p/g	Business administration	Survey (analysis of documents)
4.	Braine	1989	University of Texas (Austin)	Faculty	n/a	u/g	Science and technology	Survey (analysis of documents)
5.	Johns & Mayes	1990	San Diego State University	International students	80	u/g	n/s	Simulated summary task--> analysis of results
6.	Casanave & Hubbard	1992	Stanford University	Faculty	85	p/g	Humanities, science, and technology	Survey (analysis of documents)
7.	Jenkins et al.	1993	6 U.S. universities	Faculty	188	p/g	Engineering	Survey (analysis of documents)

n/a = not applicable
u/g = undergraduate
p/g = post graduate
n/s = not specified

Conclusions. What are the main implications of the research reports reviewed in this section in terms of the TOEFL 2000 project?

First, there can be no doubt that the case made by Horowitz and others for using direct inspection of documentary evidence of writing tasks as a research heuristic is a sound one, given the problems of perception, terminology, and so on inherent in alternative data-gathering methods. It should also be noted, however, that this really just shunts the conceptualisation problem further down the line, rather than completely solving it. This is because, having been collected, the samples still have to be categorised by the researcher. This reintroduces an element of personal judgment into the process. However, there appears to be no way of avoiding this, if the data is to yield some kind of taxonomy. Rather, the problem as I see it is that, so far, no generally agreed upon set of categories for academic writing tasks exists. In other words, we still do not have a universally accepted answer to the question, What are the main types of writing that students in the North American higher education system have to undertake, both in general terms and in terms of level and specialism? It is therefore necessary to conduct further research to determine this, preferably building on the foundations laid by Horowitz and others.

However, clarifying what the writing tasks actually are, although necessary, will be by no means sufficient. It will still be necessary to identify the nature of the writing the students are expected to produce in response to the tasks. This is a matter of first of all finding out what level of performance is acceptable for a range of tasks (judged by the grade earned), and then trying to work out what the main writing skills involved in such scripts are. Such an analysis should also attempt to identify the extent to which the skills required differ or are similar according to level and specialism.

Unfortunately, little light has so far been thrown on this side of the matter. Research by Vann, Meyer, and Lorenz (1984), although useful, focuses only on decontextualised sentence-level writing skills. Santos (1988) investigated faculty judgments of connected writing, but the number and nature of the scripts used makes generalisation difficult. More promising, perhaps, is Leki and Carson's recent study (Leki & Carson, 1994). Their research attempted to find out, among other things, what writing skills undergraduate students from a wide cross section of academic disciplines felt were important for success in academic writing tasks, and what their main areas of weakness were. Leki and Carson's results might be built on by carrying out an analysis of suitable student scripts, at both the undergraduate and postgraduate levels, using the framework of skills that informed their investigation.

It is also worth noting that Leki and Carson's research framework usefully sheds light on not only the skills involved in the written product but also those needed for the writing process itself. Thus, for example, they highlight the importance of being able to synthesise ideas from multiple sources as an input to writing, or being able to select with sufficient speed the most concise form of expression to use. Further research might thus need to look closely at the actual composition process itself, in order to more adequately determine the subskills involved. From this perspective, a suitable writing task for an EAP test might involve the student producing a summary of two or more short readings, under appropriate time constraints.

Whatever the exact nature of the way forward, however, the basic research strategy needed would appear to be as follows:

- Formulate an agreed classification of academic writing tasks, distinguished by level/specialism as necessary.
- Obtain graded samples of student writing produced as a result of undertaking a representative cross section of the tasks.
- Analyse the writing to determine the constituent subskills involved in relation to the grade earned.
- Determine the processes that contribute to the presence of desirable subskills.
- Devise test items that require the student to draw on the underlying processes needed for successful mastery of the subskills.

It should be noted, however, that the order in which the stages of a research and development programme of this kind might be carried out may well differ (see Proposals for Further Research, below).

Listening Needs

Chaudron and Richards (1986) report on their investigation of “macro” and “micro” discourse markers in lecture comprehension. Two groups of subjects were used. The first (the “pre-university” group) consisted of 71 ESL students attending an intensive language programme at a private college in Hawaii. The second (the “university” group) comprised 81 ESL students enrolled at the University of Hawaii. The university group had a higher level of proficiency than the pre-university group.

Four versions of a lecture on United States history were prepared. Each version varied from the others in terms of the presence/absence of “macro-markers” and/or “micro-markers” (i.e., global-level and local-level markers of cohesion, respectively). Subgroups within each of the main groups listened to one of the versions of the lecture, and comprehension was checked using a cloze measure, multiple-choice questions, and a true/false quiz.

Results showed that the + macro-markers/ - micro-markers version of the lecture led to significantly better comprehension on the part of the university group and some of the pre-university subgroups. Macro-markers were therefore shown to be an important element in lecture comprehension for the groups in question. However, although the findings may well be true for other groups of ESL listeners as well, it would be useful for similar research to be carried out with comparable groups of native speakers. Also, the experimental setting of Chaudron and Richards’ study may have had an effect on the listening strategies used by the subjects. In other words, we still need evidence that macro-markers have the same importance when listening occurs in the normal study situation. Furthermore, as Tyler (1992) indicates,

the lack of importance of micro-markers reported may have had more to do with the level of the subjects' English than their intrinsic significance in aural comprehension.⁴

Dunkel (1988) investigated lecture note-taking. The sample consisted of a random selection of 129 undergraduate freshman at the University of Arizona, of which 66 were native speakers of English and 63 were nonnative speakers. Subjects took notes while viewing a video-taped lecture. At the end of the lecture, the notes were collected and a test of retention was administered. The notes were analysed in terms of five main criteria and compared with the test results.

Results indicate that test achievement was not directly related to quantity of notes taken, but instead to terseness and inclusion of the main points.

In general, the analysis of the data indicates that effective L1 and L2 note takers were those who compacted large amounts of spoken discourse into propositional-type information units, transcribed content words (e.g. names, dates, statistics) using abbreviations, symbols and a number of structure words; and detected and wrote down information that subsequently appeared on the post-lecture. The L2 note takers who did not perform as well on the quiz wrote down numerous structure words (e.g. articles and prepositions) so that their notes contained fewer information units overall but a larger quantity of words or notations (Dunkel, 1988, p. 270).

The findings also indicated that no single method of note-taking produced uniformly superior results:

Indexes of note quality did not function in similar fashion across the L1/L2 groups to predict achievement . . . [so] the notion that there is a single, unitary (or universal) note-taking method that is effective for all groups of students does not find support (ibid., p. 271).

There were a number of limitations to this study, as Dunkel herself (ibid, pp. 273-278) acknowledges. For example, the students were not given an opportunity to review their notes before being tested, despite the view of other researchers that it is the review element, rather than the act of note-taking, which is of greater significance in terms of retention of information. Another problem was the laboratory-like conditions of the investigation. As Dunkel says, "A naturalistic investigation of the notes

⁴DeCarrico and Nattinger (1988) usefully extend Chaudron and Richards' research via their concept of the "lexical phrase," i.e., prefabricated chunks of language, some of which correspond to Chaudron and Richards' "macro-markers," but many of which perform other, additional functions. They applied this concept to the analysis of a series of lectures in which examples of all three of Dudley-Evans and Johns' (1981) lecture styles occur (*viz.*: the reading, rhetorical, and conversational styles). They conclude that the colloquial, everyday language of the lexical phrases found in the conversational style of lecture is probably difficult for ESL students to process because of their typical lack of exposure to English of this kind. However, despite its apparent reasonableness, the authors do not provide any research evidence for this view.

students take during actual classroom lectures and their relationship to performance on actual classroom exams is sorely needed" (p. 277) (see Benson, 1989, for a more ethnographic account, based on a case study of a single student, and Dunkel & Davy, 1989, for native- versus nonnative-speaker students' views on note-taking practices).

Nevertheless, Dunkel's research design was careful and thorough, and the limitations are due mainly to its aims and basic structure, rather than to the research procedures used. Her findings appear to show that the good note takers were essentially good summarisers. This, along with Horowitz's (1986a) findings regarding the importance of summarising in academic writing, indicates that skill in abstracting and reformulating the gist of information is a vital, widely applicable EAP skill. There is an obvious need for further research along the lines Dunkel indicates, however, if we are to have a more adequate understanding of how this aspect of EAP fits into the scheme of things. Perhaps it is also worth noting that Gibbs (1981, p. 58) cites several studies that have shown that taking notes is not necessarily associated with better learning outcomes than not taking notes, and Benson (1989) cites Hartley and Davies' (1978) finding that, while 17 studies they surveyed showed that students who took notes had improved examination results, 16 other studies were inconclusive. This may be because notes are only a surface realisation — more important for some students than others — of the play of deeper, more meaningful cognitive processes.

Olsen and Huckin (1990) investigated academic lecture comprehension in order to attempt to throw light on why some students, despite understanding the lecturer's use of "discourse markers" (i.e., global-level markers of cohesion: cf. Chaudron and Richards [1986]) nevertheless do not understand the main points or overall structure of the lecture. The subjects were 10 graduate and four undergraduate international students of engineering. They viewed a video of a first-year graduate course lecture on a mechanical engineering topic and were asked to identify the main ideas and take notes as they would in a normal lecture situation. Afterwards, they were asked to explain what the lecturer said as if to a friend who had missed the lecture, using their notes and taking as much time as needed. These summaries were recorded and analysed for accuracy.

As a result of their analysis of the findings, the authors argue that the failures in understanding that occurred were due to (a) the students being accustomed in their own countries to a more overt and direct signalling of main points in lecture discourse than the kind that is customary in the United States, as typified by the lecture in question; (b) the use by the students of an information-driven listening strategy (i.e., one that simply strives to absorb facts), when what was needed for the lecture in question was a "point-driven" strategy (i.e., one that involves the listener in actively imposing an overall framework of meaning on the data). Effective lecture listening comprehension would therefore seem to involve not just an understanding of macro-markers but of detecting the overall schemata that the speaker is representing.

Unfortunately, while Olsen and Huckin's findings would appear to make sense in terms of what is known about the role of cognitive and affective frames of reference in comprehension (see, e.g., Bransford, Stein, & Shelton, 1984), the research method may have contaminated the results. The comprehension task facing the subjects might well have proved just as difficult for native-speaker

students, since the subject matter might have been insufficiently familiar, in terms of topic and/or level. (It should be noted that the lecture was from a graduate-level class, but four of the subjects were undergraduates.) Furthermore, even when subject matter is familiar, much of what is involved in effective comprehension, for native speakers and nonnative speakers alike, involves drawing on a plethora of contextual clues, both immediate and more remote (see Hutchinson & Waters, 1981, p. 56). In Olsen and Huckin's research, however, the subjects were simply given their viewing task without any immediate contextualization. It is therefore possible that this is what led the subjects to employ an inappropriate listening strategy; that is, their attention was focused to an unnatural degree in the first instance on trying to comprehend the basic facts of the lecture in order to attempt to compensate for the lack of prior knowledge. In the light of this evaluation it is rather unfortunate that Olsen and Huckin (*ibid.*, p. 44) go on to ascribe the predisposition on the part of their subjects to a point-driven listening strategy to what they see as the overly fact-oriented ethos of United States science and engineering education.

Although Olsen and Huckin's research design may be flawed, the issues it raises are important ones both for understanding the nature of lecture discourse comprehension and for conducting further research into this area. While studies such as Chaudron and Richards (1986) have shown the importance of understanding global discourse markers, effective academic listening clearly involves more than this. Depth of understanding of the subject matter also plays a crucial role, for it is this that enables the listener to interpret the discourse not just at the cohesion level but also at the coherence level (Widdowson, 1978), that is, in terms of its underlying structure of ideas, or, in other words, by invoking the relevant content schemata (see Carrell, 1987).

For example, it is perfectly possible to listen to a lecture on say, nuclear physics, in which one understands all the macro-markers, and yet comprehends very little. Conversely, one may be able to grasp the gist of a lecture on a familiar topic in a foreign language without necessarily understanding that language's macro-markers. Coherence, rather than cohesion, is therefore primary. Furthermore, it is not sufficient for the listener simply to have the necessary background knowledge for "listening to learn" (Benson, 1989) to occur: the listener must also want and know how to actively deploy background knowledge to build creatively on the speaker's words, since the straight imbibing of fact will rarely be sufficient for effective listening in an academic context, as Olsen and Huckin (1990) try to show. The role of content knowledge and its effective mobilisation in academic listening comprehension is therefore an area that deserves further research.

Table 3 **Summary of Research on Listening Needs**

	Author(s)	Date	Institution(s)	Subjects	N	Level	Subject area(s)	Research
1.	Chaudron & Richards	1986	University of Hawaii	International students	152	u/g	n/s	Lecture simulation --> analysis of post-lecture results
2.	Dunkel	1988	University of Arizona	International and indigenous students	129	u/g (f/m)	n/s	Lecture simulation --> analysis of post-lecture results
3.	Olsen & Huckin	1990	U.S. university	International students	14	p/g & u/g	Engineering	Lecture simulation --> analysis of post-lecture results

u/g = undergraduate

p/g = post graduate

f/m = freshmen

n/s = not specified

Conclusions. In this area there is once again an inadequate research base for guiding the TOEFL 2000 endeavour. The first problem is quantitative — lack of research. I have not been able to identify a single body of evidence that explains the nature of the North American higher education listening task in terms of the types of input students are exposed to, the status and use of the information, how these matters differ by specialism and level, and so on. Some kind of survey would therefore need to be undertaken in order to obtain this information. The second problem is qualitative. The work that has been done has been exclusively experimental in terms of its research design. While this has yielded useful information, there is also a need for more naturalistic studies that look at academic listening as it actually occurs in real rather than experimental settings.

In addition, for any further research in this area to be useful for testing purposes, it will be necessary to follow the basic procedure sketched out in the discussion above. That is, the essential features of the required EAP competence need to be sifted out of the mass of data, with a view to possibly

operationalising the results in terms of test items. In other words, having identified the most significant kinds of listening tasks students face, we then need to establish what it is that makes for success in coping with such tasks. We may find, for example, that successful academic listening involves, among other matters, a questioning approach to the information being presented. We might be able to assess mastery of this skill through a task in which students are asked to evaluate what they have listened to in the light of some subsequent written input.

Furthermore, in probing for the constituent features of successful academic listening, it would appear important, in the light of research such as Olsen and Huckin's (1990) and Dunkel's (1988), to be prepared to deal with the thorny problem of how the language-specific elements involved can be distinguished from the information-processing side of the matter. It may in fact be the case that the deployment of formal and content schemata are so inseparably intertwined in effective academic listening that one cannot adequately assess academic listening competence solely in terms of command of formal schemata. If so, then it would appear to be the case that test items would need to be devised that could probe competence involving content schemata as well, but not in a way that would disadvantage students who do not have the necessary background subject matter knowledge. This is a tricky problem, but similar to that faced by writers of EAP textbooks in choosing their texts. Some guidance in terms of suitable subject matter might therefore be derived from this source.

Speaking Needs

Rounds (1987a) investigated the oral proficiency needs of international teaching assistants (ITAs) at a United States university. The subjects were two native- and three nonnative-speaking mathematics teaching assistants. On the basis of feedback and other criteria, the ITAs' supervisor had judged two of the subjects to be successful, two to be less successful, and the fifth to have some language difficulties. The ITAs and their students and supervisor, as well as the researcher, analysed the videotapes and transcripts of the subjects' teaching, with an open-ended invitation to comment on anything "unusual, interesting or problematic" (Rounds, 1987a, p. 647). In this way it was hoped that significant features of the performances would be identified more validly than by the imposition of a prior, researcher-oriented framework.

The findings showed that the more successful ITAs:

- used inclusive pronouns more frequently (cf. Rounds, 1987b);
- avoided apparently haphazard periods of silence;
- appropriately matched symbolic representation (e.g., writing an equation on the blackboard) with its verbalisation;
- elaborated stages in explanations by;
 - a) clearly marking important junctures,
 - b) using repetition and other devices in order to establish greater cohesion,
 - c) "information chunking" (clearly organizing information),

-
- d) using “dual-function” conclusions (i.e., ones that comment on the information and the students’ responsibility with respect to it),
 - e) asking questions, including questions aimed at the management of the learning,
 - f) using a persuasive style of speaking (i.e., one that attempted to not only convey the information but also make it seem natural and obvious (p. 665).

In discussing her findings, Rounds concludes that

... communicatively competent classroom discourse is based on (a) an understanding of the student-teacher relationship expected in American university classrooms; (b) an understanding of the ability of silence to contribute to or detract from the creation of fluency; (c) an awareness of what students are doing while the teacher is performing, especially a sensitivity to their note-taking task; and (d) an acceptance of the fact that teaching involves more than efficient transmission of information and that elaboration is highly valued by students” (p. 666).

Rounds’ research design is significant for its careful, ethnographic reading of the study situation under investigation, and for the close attention paid to the relationship between the linguistic elements analysed and their place in the overall discourse setting. In other words, one is left with the impression that the linguistics has not imposed itself on the discourse, and thus left out of account or distorted important elements in the picture. Rather, it has been used as a heuristic to deepen understanding of the picture as a whole. At the same time, however, it has been possible for a number of generalisable points to emerge. We will return to the value of this type of research design for investigating EAP needs in “Other Needs,” below.

Micheau and Billmyer (1987) report on their investigation of the spoken discourse of MBA case analysis discussion sessions at the University of Pennsylvania. Using a framework derived from conversational analysis, case study sessions involving international students and similar sessions involving native speakers were observed. In addition, interviews were held with some of the native-speaker students and with MBA and other faculty.

The data was analysed in terms of strategies used by speakers for coordinating turn exchange, obtaining a turn, behaviour within the turn itself, and cooperation in interaction. The native-speaker student and faculty data identified a preferred pattern for the overall structure of the discourse and a number of discourse strategies for its realisation. Analysis of the international student data showed a number of discourse strategies that did not conform to the native-speaker norms, for example, violations of turn taking, delayed response to a turn allocation, taking an excessively long turn, and so on.

This research would appear to show that there is a clear need for nonnative-speaker students involved in such interactions to understand and be able to operate the relevant interactional rules appropriately. Although level of language no doubt plays a role in the realisation of the latter, the student clearly also needs the right kind of interactional competence as well. It is also worth noting that this research shows,

as does much of the research on other skills (e.g., academic writing and listening) reported elsewhere in this review, that it would appear to be strategies at the genre level (Swales, 1990) that are perceived to be crucial in demarcating the line between communication in academic settings versus more everyday ones. In other words, what the student needs to be able to cope with are settings that are characterised by specific conventions established by the “owners” of the setting, that is, Swales’ (1990) “discourse community.” The question of what kind of competence this calls for on the part of the student will be discussed in the final section of this paper.

Williams (1992) studied the effect of discourse marking on the intelligibility of 24 International Teaching Assistants (ITAs) at a major United States university. The study also included five native-speaker TAs (NSTAs). Two teaching performances of each of the ITAs were videotaped, “In the first task, planning was both possible and encouraged, whereas in the second task, little planning was possible” (Williams, 1992, p. 698). Segments of planned and more extemporaneous teaching by the NSTAs were also recorded. The videos were analysed in terms of the occurrence of macro-markers (Chaudron and Richards, 1986) in conjunction with certain kinds of key statements (e.g., definition, example/illustration, restatement/rephrasing, etc.) and rated in terms of comprehensibility.

The results indicated that

- percentages of NSTA-marked key statements were similar for planned and unplanned discourse;
- in planned production, the number of markers used by the ITAs increased;
- the NSTAs were rated as considerably more comprehensible;
- in general, the planned production of the ITAs was rated higher than their unplanned production;
- morphological and syntactic complexity differed significantly for planned versus unplanned ITA discourse, but accuracy did not.

In her discussion of the results, Williams points out that despite the similarities between the NSTAs and ITAs’ level of marking in unplanned discourse, the NSTAs were clearly more comprehensible, indicating that the NSTAs were exploiting resources other than macro-markers to make themselves clear. On the other hand, increased use of markers by the ITAs enhanced their comprehensibility. Williams concludes that ITAs should be encouraged to use more overt marking of discourse moves in order to compensate for their inability to acquire other aspects of NSTA competence, native-like pronunciation, for example.

This research usefully complements Rounds’ (1987a) research by throwing further light on the gap between ITA and NSTA performance. More overt marking is probably associated with more successful ITA/TA teaching performance because it contributes to greater elaboration, in Rounds’ sense of the term (p. 658 ff).

Tyler (1992) compared the use of discourse-structuring cues by an ITA and an NSTA at a United States university. Videos were made of both speakers giving a short lecture in their area of specialisation to an audience of fellow trainee ITAs/TAs. In order to remove the nonnative-like pronunciation of the ITA as a factor in the investigation, transcripts of the two lectures were made and the two texts were read

as lectures to another audience (of graduate linguistics students) by a native speaker of English. The linguistic content of and the audience reactions to the two texts were analysed.

The findings indicated that the second audience found the ITA text difficult to follow. Analysis of the text revealed that the speaker used discourse markers in a disorganised manner, failed to provide an adequate level of "lexical specificity" (semantic cohesion), used a level of syntax that was insufficiently complex to signal the interconnections between ideas adequately, and produced prosodic clues that conflicted with semantic information. On the other hand, the second audience found the NSTA's text well-organised and clear. Although the global structure of the two talks was in fact similar, the use of discourse markers by the NSTA was consistent and predictable, there was much greater use of repetition at word and phrase level to establish lexical specificity, and complex syntactic structures were used much more frequently, making logical connections more explicit.

Tyler concludes that both micro- and macro-markers have a strong effect on native-speaker comprehension, and that therefore both levels are important to consider when evaluating ITA performance. She argues that Chaudron and Richards' (1986) finding that micro-markers were less significant was probably a reflection of the lack of development of these features in the subjects' own English proficiency. However, although Tyler's conclusions seem intuitively sound and coincide with many of Rounds' (1987a) findings, it should be pointed out that the research design contains a number of artificialities. The subject of the ITA's lecture was traffic signalling devices, and the NSTA's lecture discussed the diffusion of perfume molecules. The audience, however, was made up of linguistics students. The effects of knowledge of and attitude to the subject matter must therefore be considered intervening variables. Also, the manner of delivery of the lecture was artificial. The results can therefore only be generalised from with caution.

Weissberg (1993) investigated the discourse of a number of seminars in graduate departments of animal science and agronomy at New Mexico State University. In particular, he focused on the use of the graduate seminar as a forum for thesis/dissertation defence and oral presentation of research in progress. Ten such seminars were observed, half of them presented by nonnative speakers. Data was also obtained via inspection of course syllabi and through interviews with some of the instructors and nonnative-speaking students.

Weissberg characterises graduate seminars of the type under consideration as being essentially a "hybrid" genre, a point during which the student is in a period of transition from "apprentice" to "peer" (Swales, 1990). Their discourse is therefore ambiguous, displaying conventional student: instructor distance on the one hand, and the more convivial tenor of the (incipient) peer relationship on the other. Instructors indicated that their preference was for a relatively extemporaneous, audience-friendly presentational style.

Analysis of student discourse showed that most of the nonnative-speaker students used a more formal style of delivery, while most of the native speakers spoke more conversationally. This applied to the use of visual means of expression (slides and overhead projector transparencies) as well. The interviews with the nonnative-speaking students revealed that in their own academic culture, it was more normal for a

research report to take the form of written text being read aloud. Also, their level of confidence in handling reasonably spontaneous spoken English was generally low. Here, as elsewhere in the literature on EAP, (see, e.g., Hutchinson & Waters, 1981; DeCarrico & Nattinger, 1988; Parry, 1991), the need emerges for nonnative-speaking students to be able to deal with a reasonably general style of English within the academic setting. However, this need has tended to be overlooked, and greater importance has traditionally been attached to the need for the student to master more formal academic discourse. This need is also undoubtedly a major one, but, in practice, students are often less well-prepared for handling the more informal style of much oral academic communication, as Weissberg's research indicates. One must distinguish between EAP needs in the sense of academic English on the one hand and English for communicating about academic topics on the other. It is the latter, not just the former, that EAP involves, as is shown by the variation in the type of language used in academic settings that research such as Weissberg's reveals.

Smith (1989) investigated the effect of topic on the oral proficiency of 38 University of Minnesota intending ITAs from the departments of physics, mathematics, and chemistry. The subjects were tested with Form 1 of the original SPEAK[®] test⁵ and a subject-specific version of the test designed by Smith. The field-specific versions were validated by three NSTAs. However, the ITAs felt that the field-specific versions involved a higher cognitive load than was required by the general version.

The results showed that some of the ITAs did not perform as well on the subject-specific version as on the general one. There were no significant differences between the mean scores on the two tests. Analysis of variation in individual performance across the two tests showed no consistent pattern. This study was therefore unable to show that oral competence in academic settings depends to a significant extent on specialised topic knowledge.

In a related study, Douglas and Selinker (1993) attempted to investigate the effect of "test method facets" (i.e., features of the context of the test, such as instructions, test language (including vocabulary), distribution of information, etc.) on performance by ITAs in general as opposed to field-specific tests of oral performance. The subjects were 15 international graduate students at Iowa State University. Twelve of these subjects were mathematics specialists, and, in order to throw light on the specificity of the field-specific test, the others were statistics graduates and economics specialists. The subjects had all taken the SPEAK test within the previous 12 months. The researchers prepared a mathematics-specific version of SPEAK (MATHSPEAK). In order to foster a more mathematics-oriented situation the instructions for MATHSPEAK differed from those for SPEAK by emphasizing the intended context of use, and so on. For similar reasons, the MATHSPEAK test was administered to the subjects under the aegis of the mathematics department.

⁵The SPEAK (Speaking Proficiency English Assessment Kit) test is an institutional version of the Test of Spoken English (TSE). The test requires examinees to answer orally a variety of general questions presented in print and recorded form.

Results indicated that while grammar and fluency scores for MATHSPEAK were significantly higher than for SPEAK, comprehensibility and pronunciation scores were not. Also, "of the 12 mathematics subjects, 7 scored higher on the SPEAK than on MATHSPEAK" (Douglas & Selinker, p. 244). However, since the nonmathematics subjects scored consistently lower on MATHSPEAK than on SPEAK (though not significantly so), the researchers conclude that MATHSPEAK appeared to show itself to be a "test of ability to talk about mathematics in English" (Douglas & Selinker, p. 242).

Further analysis of transcripts of the subjects' performances on SPEAK and MATHSPEAK showed no consistent grammatical differences, despite the higher ratings given for this to subjects on MATHSPEAK. However, MATHSPEAK discourse showed greater rhetorical complexity, which raters may have included in their grammar assessment. "Inter-rater reliability for MATHSPEAK was not particularly high" (Douglas & Selinker, p. 243).

Overall, the authors conclude that the results are characterised by considerable variation, and no firm conclusion can be drawn regarding the desirability of field-specific versus general tests of oral proficiency. Like Smith's (1989), this study would therefore seem to indicate that the oral competence needed in academic settings is a general, rather than subject-specific one. We will return to this issue.

Table 4 **Summary of Research on Speaking Needs**

	Author(s)	Date	Institutions	Subjects	N	Level	Subject area(s)	Research method
1.	Rounds	1987	U.S. university	ITAs & NSTAs	2 & 3	p/g	Mathematics	Analysis of teaching performance
2.	Micheau & Billmyer	1987	University of Pennsylvania	Faculty, international students, and indigenous students	n/s	p/g	Business administration	Observation and interviews
3.	Williams	1992	U.S. university	ITAs & NSTAs	24 & 5	p/g	n/s	Analysis of simulated and real teaching performance
4.	Tyler	1992	U.S. university	ITA & NSTA	1 & 1	p/g	Engineering & science	Analysis of simulated lecture
5.	Weissberg	1993	New Mexico State University	Faculty, international students, and indigenous students	n/s	p/g	Animal science and agronomy	Observation, interview, analysis of documents
6.	Smith	1989	University of Minnesota	ITAs	38	p/g	Physics, mathematics, & chemistry	Analysis of test results
7.	Douglas & Selinker	1993	Iowa State University	ITAs	15	p/g	Mathematics	Analysis of test results

p/g = post graduate

n/s = not specified

Conclusions. The most striking overall feature of research in this area is the preponderance of studies relating to the needs of the ITA. This is an important area, of course, and the studies that have been conducted have been valuable in their own right as well as in terms of illustrating methods of research that might be used to investigate other areas of speaking need. It is clear, however, that research into other uses of speaking at the postgraduate level and, at the undergraduate level, research into speaking needs in general, seem to be largely lacking. Further research is therefore urgently needed. Given the current dearth of information, this might necessitate carrying out, first of all, a number of in-depth studies along the lines of those performed by Micheau and Billmyer (1987) and Weissberg (1993), followed by an overall survey to determine how the results vary according to level and specialism. Although it is perhaps premature to speculate, on the basis of the evidence currently available, one might eventually be looking at test items that focus on students' ability to participate in an appropriate manner in reasoned academic discussion, and to make a short oral presentation on an academic topic.

Other Needs

Jacobson (1986) analysed the EAP needs of international students in the physics laboratory at the University of Minnesota Institute of Technology. The research methods involved direct inspection of instructors' teaching and evaluation material and observation of lab sessions, as well as interviews with the instructors. The design of the study also made it possible to compare the communication strategies of international students with those of indigenous students. In the interests of maximising the possibility for generalisation, a basic-level, common-core engineering course was chosen as the focus of the research.

Four main communicative strategies important for operating effectively in the environment in question were identified:

- Evaluating and selecting information for a specific purpose
- Synthesising information from more than one source
- Applying information to new or different situations
- Establishing working relationships with others in the lab

Jacobson also points out that similar findings have been made for other contexts [e.g., graduate seminars in agriculture (see Schmidt, 1981) and undergraduate lectures in business studies (see Sorensen, 1982)].

Jacobson's study is to be commended for its rigorous research design, including as it does three main points of reference for the collection and analysis of data — instructors, international students, and indigenous students — as well as a number of other checks and balances. Indeed, in overall terms, it approximates an ethnographic research paradigm much more closely than do any of the other studies reviewed so far. This study is also valuable for having yielded a useful framework for the kind of spoken and written strategic competence needed for academic tasks. As such, it expands the focus of previous research and provides a sound basis for further research in this area.

Jacobson's work is also significant as an early example of an attempt to look closely at a relatively small part of the overall picture, as opposed to the wider, survey-based approach of much of the research that came before it. In fact, the trend in most North American studies from the mid-1980s onward appears to have been away from viewing the broader canvas and toward looking in increasing detail at a smaller part of the picture.

In my view, both types of studies can be seen as usefully complementing each other, similar to the way that altering the focus on a telescope can provide different perspectives, enabling different information to be obtained. The more varied the perspectives, the more likely the picture we see will be an accurate one. The most accurate picture will be the one that tries to synthesise and reconcile as many differing views as possible. However, as discussed earlier, it would seem best, as a general strategy for further research, to first of all properly establish the nature of a small part of the field, and only then to broaden the enquiry. Unless this basic strategy is followed, there is a danger that there will be insufficient depth of understanding and communication for the wider type of survey to produce reliable results.

Benson (1991) describes a study of the academic reading experiences of a graduate international student following a course in public administration at a United States university. One-sixth of the students' readings were randomly sampled and their constituent "thematic units" identified [a category similar to Trimble's "conceptual paragraph" (Trimble, 1985)]. The thematic units were then coded in terms of four variables: content, sources of authority, values, and text types.

Benson's main findings were that (a) "extensive reading is an essential part of the EAP curriculum;" (b) dealing with such texts involves understanding the author's underlying value system, which will usually be of the Western-progressive-intellectual type; (c) such texts make use of various sources of authority, and the student needs to be aware of what these are and how they operate; (d) readings of the type in question contain significant amounts of description and argument, in addition to exposition (though this did not appear to have impeded the subject's understanding). In addition, Benson argues that his study highlights the importance of academic reading not just as a process of comprehension but also of learning. In other words, what his subject had to do in order to succeed academically involved not only absorbing facts but gradually linking these to his existing framework of knowledge and, ultimately, arriving at a new synthesis based on his own ideas and those in the readings (see also, Clark et al., 1990).

Benson's research, like that of Jacobson's (1986) reviewed above, shows once again what can be learned from a close, ethnographic account of academic practices as experienced by the student. Also, Benson's concept of "reading to learn" usefully complements Olsen and Huckin's "point-driven" listening strategy and Benson's (1989) own parallel "listening to learn" concept; it would appear that comprehension in academic settings involves not just the ability to decode the linguistic data ("listening to comprehend," Benson, 1989), but also the ability to impose a meaningful framework of understanding on it. In other words, a general academic frame of mind, involving active cognitive processing strategies (and probably an equally or even more important affective dimension as well) seems to play a significant role in successfully studying English (see also Waters & Waters, 1992).

Parry (1991) reports on a longitudinal study of several international students' vocabulary acquisition experiences in an introductory anthropology course at Hunter College, City University of New York (CUNY). The subjects were given a preliminary vocabulary test and were asked, as they read their anthropology textbooks, to (a) list any words that gave them difficulty, (b) write down what they guessed the words to mean, and (c) record any dictionary definitions they had recourse to. The glosses were analysed by comparing them with the original words.

The findings showed that words specific to anthropology were not the main cause of difficulty. A much larger proportion of words in the students' lists were what Parry (after Martin, 1989) terms "bridging" vocabulary, that is, general items that characterise formal prose and often express relationships or other abstractions (e.g., "ensuing," "vitality," "rudimentary," "circuitous," "supplant," "lurid," "repertoire," "viable," etc., Parry, 1991, p. 637). Parry therefore concludes that "where students most need to build up their vocabularies is not in any particular subject matter but in the register of formal expository prose" (Parry, 1991, p. 637). She also points out that each student's list was highly individual (no word was identified by more than two of the four students). She therefore argues that teaching should focus on techniques for inferring word meanings. What students really need to be able to do, she goes on to say, is to "recognize those words whose meaning they must know accurately in order to understand a particular text" (Parry, 1991, p. 650).

One might quibble at the definition of terms underpinning some aspects of Parry's research. For example, it is difficult to see what the defining criterion is for her category of words "specific to anthropology," since these include words such as "betrothal," "kinship," "emic," "etic," "sorcery," "stratification," "primate," and so forth. While some of these words are reasonably specialised, their occurrence is by no means confined to anthropology alone, and some of the terms are quite general. Likewise, the term "bridging" vocabulary, although defined, appears to encompass a very wide range of lexical items indeed, with the "general" perhaps insufficiently distinguished from the "sub-technical" (Cowan, 1974).

More positively, however, Parry's research shares with, for example, Benson (1991), the virtue of being grounded in a direct observation of the student's academic experiences. It also points usefully to one particular type of reasonably general vocabulary ("bridging" words) as being of particular importance in EAP, and the need to pay attention to not only the type of vocabulary but also the kinds of vocabulary reading strategies needed. It is likely that both indigenous as well as international students need to be able to decide which unknown words are important enough to try to understand and how to locate clues to understanding from within the text itself. In other words, the lexically related knowledge needed for successful academic study in English may depend not only on having acquired a certain stock of vocabulary but, equally crucial, on being able to go on adding new items of importance to the stock without undue difficulty. Here, as elsewhere, we need to see the capacity for study not only in product-oriented terms — as a body of knowledge — but also as a set of processes that enable existing knowledge to be deployed strategically and further developed and refined (see Waters & Waters, 1992).

Table 5 **Summary of Research on Other Needs**

	Author	Date	Institution(s)	Subjects	N	Level	Subject area(s)	Focus	Research method
1.	Jacobson	1986	University of Minnesota Institute of Technology	Faculty, international students, & indigenous students	45	u/g (f/m)	Engineering	Communication strategies	Observation, interview, analysis of documents
2.	Benson	1991	U.S. university	International student	1	p/g	Public administration	Reading	Analysis of readings
3.	Parry	1991	Hunter College, City University of New York	International students	4	u/g	Anthropology	Written vocabulary	Analysis of glosses

u/g = undergraduate
p/g = post graduate
f/m = freshmen

Conclusions. The most obvious feature of the research reported in this section is how little of it there has been. First of all, at the skills level, less research has been conducted on reading than on any of the other skills. As a result, it would appear we do not know with any certainty what the higher education academic reading requirements are in North America, in terms of the types required for the main levels and specialisms, and, at each of these "nodes," the most important subskills and processes. Since it is clearly difficult to construct any new test of EAP proficiency in the absence of this information, it seems that this is, once again, an area requiring further research.

Second, this section and the review as a whole shows that there has been very little research on areas of need other than in terms of one or another of the four skills. This results in at least three major problems. First, it ignores the fact that a good deal of EAP is carried out within a multiskill or integrated-skills framework. For example, lectures involve not only listening but frequently also writing, both in the form of note-taking and possible subsequent incorporation of the information into, for example, an essay, as well as a certain amount of reading (from the blackboard, overhead projector, and/or lecture handouts). Similarly, writing an academic paper clearly also involves a good deal of reading. Research therefore also needs to capture the integrated nature of skill use in academic settings. Unfortunately, very little work of this kind seems to have been done so far.

The second main problem with using a four-skills framework for investigating EAP needs is that it tends to obscure aspects of need that cut across the division between the skills. Benson's work (Benson, 1989 & 1991) shows that academic listening and academic reading, although clearly different in a number of respects, can both be thought of as involving the student in a process of learning, not just comprehension. In other words, at the deeper, subskill level, both of these skills probably involve similar general information gathering, evaluating, and resynthesising strategies. With the addition of an information-presentation dimension, academic writing and speaking can be thought of in similar ways. Thus, beneath the level of the four skills, there exists a body of underlying EAP competence, made up of general information processing strategies, which it is also important to identify and recognize as a major part of a picture of EAP needs. Once again, however, research into this dimension appears to have been neglected, and therefore further investigation needs to be undertaken. To this end, Jacobson (1986) provides a good model.

Finally, the preponderant focus on the four skills also means that other important aspects of EAP, such as vocabulary, may not have been given the attention they deserve. In addition to Parry's work (1991), there are indications in Santos' research (1988) that command of the type of vocabulary that differentiates academic discourse from other types of communication may be more important than has been realised. It is also worth mentioning that, in the Australian context, Bowyer (1993) found that conveying the right tone in academic writing is as much a matter of using appropriate phrases (e.g., "based on these assumptions") as individual words (see also, DeCarrico & Nattinger, 1988). It may well be the case that certain grammatical constructions play a similar role as well.

In short, there is a need for a research framework that focuses on matters that lie beyond the boundaries of the four skills (such as vocabulary and grammar) and that also encompasses integrative and underlying aspects of skill use.

Studies in the United Kingdom

In order to throw further light on the research from the United States reviewed so far, I shall now turn to similar research done in the United Kingdom. As with the previous section, however, I will concentrate on reviewing in detail only those studies that I feel are of major significance to the task in hand. These comprise the two major recent studies of EAP needs in the British higher education context and a cluster of studies associated with the English Language Testing System Revision Project (ELTSREV). There are, of course, numerous other studies, particularly those available in the collections of papers from the succeeding British Association of Lecturers in English for Academic Purposes (BALEAP) conferences, such as Williams, Swales, and Kirkman (1984); Robinson (1988); Dudley-Evans and Henderson (1990); Adams, Heaton, and Howarth (1991); and Blue (1993). Despite the intrinsic quality of these bodies of research, however, I have not felt that they add significantly to the concerns of this paper. This is because these papers cover much the same ground, with respect to the British and other English-medium higher education systems, as that covered by the United States papers already reviewed. To review these studies at length would therefore merely add further detail without increasing the level of insight. Thus, I have deliberately chosen in this section to review work I feel expands the picture significantly.

In-depth Surveys

Hawkey (1982) investigated the role of cognitive/affective and social factors in EAP success in a longitudinal study of 27 overseas postgraduate students studying a wide range of mostly scientific, technical, and social science subjects at 28 different universities and other institutions in the United Kingdom.

The research involved two phases. Phase 1 took place while the subjects were undergoing an intensive English language course after their arrival in the United Kingdom and prior to starting their subject studies. Phase 2 occurred during the subjects' period of study in their receiving institutions. Both phases used a series of quantitative and qualitative measures of language proficiency, cognitive/affective characteristics, and motivational/attitudinal variables. The results of the findings from the two phases were compared, and the extent to which the Phase 1 measures had predicted events in Phase 2 was analysed.

The main findings indicated the following:

- Language measures that include both performance-oriented and competence-oriented elements are better predictors of necessary language ability than are measures that include only one of these elements. Of the two, however, the competence-based measures proved the more effective.
- Measures of cognitive/affective traits (such as radicalism, and field independence) and of social factors (such as attitude and motivation), when combined with language measures, predicted presence of necessary language ability more strongly than did language measures alone.

Hawkey therefore concludes that:

TL level is a powerful but not sufficient predictor of C2 EAP success. . . . My evidence is that selection (and, therefore, assessment) should be on the basis of U.K. academic awareness as well as TL level (Hawkey, 1982, p. 443)⁶.

Hawkey's research has a number of significant strengths, including its blend of quantitative and qualitative research methods, and the types of data uncovered. This allows the study to be much thicker, richer, and, despite the superficial anonymity of the subjects, much more personalised than is typical of studies of this kind. Another is its longitudinal construction, enabling both a before and after view of events and also, therefore, comparisons between the two views. If possible, these are obviously attributes worth emulating in any further major research work that may be undertaken.

Another important feature of Hawkey's research is the wide-ranging and comprehensive nature of the language assessment battery he developed as part of his research. This consisted of nine main components, tapping a cross-section of elements related to the students' levels of underlying competence and performance ability. These instruments were also valuable in yielding detailed and informative profiles of the students' knowledge and abilities. That subsequent tests such as the ELTS test, the Test of English for Academic Purposes (TEAP), and the International English Language Testing System test (IELTS) have all seen fit to include a profiling element speaks to the value of this innovation.

A further important aspect of Hawkey's research is its focus on the connection between language and personal and social factors in EAP needs analysis. There can be no doubt that Hawkey's findings confirm a view widespread among EAP teachers, namely that a significant number of overseas students with reasonable language proficiency do not appear to be able to deploy this knowledge sufficiently well in the English-medium study situation because of factors to do with their personal make-up, socio-cultural attitudes, and so on (also see Chapelle & Roberts, 1986).

Such anecdotal evidence is also of a piece with relevant theoretical constructs. EAP involves the real-life use of language, and therefore draws upon the student's level of academic communicative competence in the language. Academic communicative competence, like communicative competence in general, consists of not only the linguistic knowledge and abilities important for academic study, but also those of a psychological and social nature as well. There is thus a valid case for undertaking research that attempts to address more than just the linguistic variable in EAP communication, with a view to establishing, for example, whether a measure of "North American academic awareness" should (and could) be included in a revamped TOEFL test. Such a construct might take the form of a development of the "Metacognitive Strategies" component in Bachman's model of communicative competence (Bachman, 1991, p. 685).

⁶TL = target language; C2 = second culture.

It might be argued, however, that whatever truth there is in this picture of EAP competence, it is another matter for a test of linguistic competence to be concerned with such extra-linguistic aspects of English-medium study. Indeed, there is no reason why measures of academic and cultural adaptability need even necessarily involve English at all, since information about such traits might be more reliably obtained through measures in the test taker's mother tongue. However, given that not only Hawkey but a number of the other studies reviewed so far (e.g., Benson, 1991; Jacobson, 1986; Jenkins et al., 1993; Micheau & Billmyer, 1987; Olsen & Huckin, 1990) all point toward the close connection between language use in academic settings and awareness of academic cultural norms and expectations, it would appear important to take this dimension into account in assessing the preparedness of international students for the study in United States higher education system.

A language test may well be inappropriate for this purpose. However, it may be that supplemental measures, aimed at tapping North American academic awareness, should also be used for screening international students, and promoting and/or providing for the use of such measures should be seen as a responsibility of the language-testing organisation. In other words, does it make sense to see one's goal as only to provide information about the linguistic side of the international student's ability to cope with the North American higher education system, when it would appear that other important dimensions of the test taker's foreignness may also need to be taken into account?

The danger is that if this responsibility is not accepted by the language-testing organisation, there may be no other body with the insight, expertise, and commitment to play this role. Ultimately, if the aim is to assess EAP proficiency, the question of what measure(s) are adequate for this purpose depends, of course, on what we understand the nature of EAP to be. If EAP is more than a purely linguistic construct, as would appear to be the case, then it follows that measures of language proficiency alone will be inadequate. The risk is that unless this nettle is grasped, we may end up with ever more sophisticated language tests, but these will be inadequate predictors of EAP success, because there is no supplementary measure of the student's preparedness for being able to handle the academic cultural context in which the language will be used.⁷

Finally, to come back to matters more directly concerned with Hawkey's research per se, it should be pointed out that his work can be criticised for failing to include indigenous students as a direct point of reference and comparison. It would have been interesting to compare findings for Hawkey's subjects for all the main factors investigated — linguistic, personal, and social — with those of a comparable group of indigenous students. This would have enabled an even richer understanding of the EAP needs of the

⁷It should also be pointed out that the scope of any such supplementary measures would not extend to assessment of subject-matter knowledge, since this would not be an attribute of the student's foreignness, and, in any case, other measures already exist for this purpose. Rather, the point is to provide adequate measures of all the major attributes of the international students that separate them from their indigenous North American counterparts, i.e., first, and most obviously, proficiency in English, and, second, as has been argued here, "North American academic awareness."

overseas student, as it might have indicated more clearly what was more likely to cause problems for overseas as opposed to British students.

A further drawback of Hawkey's research is that the concept of language needs on which his tests were based was informed in construct terms by the theoretical model of Munby (1978), and, in content terms, by a review of the research findings and other evidence about EAP needs extant at the time (rather than by a direct, empirical study of the matter). As will emerge in this paper, there are now widespread reservations about the adequacy of the Munby model (see, e.g., Hutchinson & Waters, 1987, p. 54, though also see Alderson & Clapham, 1992a, pp. 155-6, 161). There have also been increasing misgivings about the wisdom of proceeding to the design of an EAP test without the benefit of an empirically established database (Criper & Davies, 1988). To try to be fair to Hawkey, however, neither of these failings appear to have invalidated his research results. Nevertheless, the construct of EAP needs informing more recent test design has moved on, and this will also be discussed below. We will revisit the second issue, that of empirically based test content, in the work of Weir (1983), which is discussed next.

As part of the design input to the Associated Examining Board's Test in English for Academic Purposes (TEAP), Weir (1983) surveyed the EAP needs of a large number of overseas students following science, engineering, and social business/administrative courses at a wide range of institutions of further and higher education in the United Kingdom. The fields selected were those with the highest proportion of overseas students.

The survey of needs involved two main phases. In the first, observations of science, engineering, and social science courses were carried out at three British universities and four colleges, in order to identify the language tasks faced by the students. The observation instruments were based on Munby's Communicative Needs Processing model (Munby, 1978) and the Schools Council Science Teaching Observation Schedule (Egglestone, Galton, & Jones, 1975).

In the second main phase, and on the basis of the findings from the observations, a series of questionnaires were drawn up in which the three categories of potential respondents — overseas students, British students, and staff — were asked to estimate the frequency of occurrence of certain academic tasks and the amount of difficulty the tasks and their attendant constraints caused. Out of 5,947 questionnaires sent, responses were received from 940 overseas students, 530 British students, and 559 staff in respect of 43 postgraduate courses, 61 undergraduate courses, and 39 A-level centres.

The questionnaire findings were summarised in a chart for each of the main skill areas (reading, speaking, etc.) in terms of perceptions of frequency and difficulty of task and associated constraints from the points of view of (a) the three main categories of respondents and (b) the main subject areas. This data was then used

. . . to establish what might be considered as "key" activities and constraints across levels and disciplines . . . we included in the [test] battery those tasks which were common and frequent

across disciplines and levels and/or seemed, from the available evidence, more likely to cause problems for the overseas as against the British students (Weir, 1984, p. 147).

Unfortunately, the questionnaire findings are too extensive and detailed to be reported here. However, published versions are available in Weir (1984).

The third and final stage of Weir's research involved the design and validation of various test formats in order to investigate which of them were best for the assessment of student performance in terms of the kinds of tasks and constraints identified in the earlier stages of the research. The resulting test consisted of two main sessions. All students took the first session, which involved a series of reading-writing, listening, listening-writing, and grammar components, linked by a theme from popular science. The grammar component was subsequently dropped in order to reduce the size of the test (Weir, 1983, p. 521). The second session consisted of two parallel versions, one aimed at engineering/science students, the other intended for arts, social studies, administrative, and business studies students. Each consisted of a thematically linked series of reading, listening, and writing components. The texts used in each version were selected from the students' subject areas. It is interesting to note that the need for subject-specific versions of Session 2 was not borne out by the research findings. That is, students' performance was not significantly affected by content from outside their specialist area:

In our investigations of the language events and activities overseas students have to deal with in British academic environments and the difficulties they encounter therein, we discovered much that was common between students of different disciplines and at different levels. This did not remove the possibility though that the subject content of texts employed in our tests might unduly affect performance. Whilst we attempted to take account of this in our sampling, we were unable to produce any conclusive evidence that students were disadvantaged by taking tests in which they had to deal with texts other than those from their own subject area. The case for a variety of E.S.P. tests therefore remains unproven (Weir, 1984, pp. 549-550).

Rather, the subject-specific nature of Session 2 was deemed necessary because of marketing factors (Weir, 1984, p. 528).

Weir's research is significant as a thorough attempt to base the design of a major EAP proficiency test battery on an empirically determined understanding of needs. As such, there is much to be learned from his study concerning both the potential and the limitations of such an approach.

First, the methodology of Weir's approach to needs analysis provides some useful pointers for carrying out similar work in other contexts. For example, the detailed and thorough observation studies he undertook in the initial phase of the research had an important effect on refining the quality of the questionnaire categories used in the second stage. Furthermore, the second stage involved a triangulation of viewpoints and sought to establish a picture of lacks by first of all identifying areas of difficulty and then comparing findings for overseas and British students.

However, in assessing the validity of the findings, it must be remembered that it was impossible to predetermine who returned the questionnaires, as Weir himself points out (Weir, 1984, pp. 153 ff). Also, in the observation stage of the research it was noted that the students who were weakest in English had the greatest difficulty in completing the pilot version of the questionnaire, and this may have also affected the responses made by some of the respondents to the later version. Furthermore, the responding sample was not representative of the research population as a whole.

But there can be no doubt that Weir's study provided him with firm research evidence for a number of the EAP tasks and constraints that subsequently informed his test design. Even if "much of what we established could, perhaps, have been accomplished by sophisticated armchair speculation" (Weir, 1984, p. 547), it would have been impossible to be sure of this, except through prior empirical verification. Thus, although many of Weir's research findings might have been anticipated, this needed to be confirmed empirically for the a priori validation of the test content, given the dearth of comprehensive research into language use in academic settings prior to Weir's study. The dangers of uninformed speculation concerning the nature of academic language use is well-documented in studies such as Hutchinson and Waters (1981) and by critiques such as Alderson (1981a).

However, Weir found that there were sometimes severe problems in establishing a match between his needs analysis findings and the requirements of his test, a reflection ultimately, it would appear, of the use of the Munby Communicative Needs Processor (CNP) (Munby, 1978), the nature of which is "descriptive, not generative" (Weir, 1984, (1) p. 23). For example, with reference to the category of "task dimensions," Weir (1984) says:

During the observations . . . it became clear that the range of complexity of text that students were exposed to in the same course and the difficulties involved in adjudging complexity, meant that no easy answer was available for the question 'how complex should the texts selected for the test battery be?' (p. 326)

Similar problems were encountered for other major categories, such as "functional" and "referential" range, subject-specific analysis of discourse (*ibid.*, p. 328), lower-order skills in spoken discourse (*ibid.*, p. 356), and so on.

Such problems would appear to cast some doubt on the feasibility of using the results of a needs analysis as a basis for EAP test design. In fact, the main source of the problem appears to have been with the construct informing the needs analysis, rather than with the needs analysis concept itself. As has already been pointed out, the Munby CNP model, albeit somewhat adapted and perceived only as *faute de mieux*, was the main basis for Weir's research into needs. Munby's model, however, is basically a device for gathering data about features of language performance in communication settings (see also, West, 1994, p. 3); Weir, on the other hand, was interested in trying to gather data about the nature of the competence underlying such performance, since the aim of his test was to sample communicative ability across a wide range of academic settings (Weir, 1984, p. 145).

Furthermore, since Weir's purpose was to gather information about needs in order to design a test, he required a needs analysis system that would enable him to operationalise the data. However, the Munby model deliberately sets aside factors concerned with the practical purpose of carrying out the needs analysis (Munby, 1978, p. 217). As a result, as Alderson (1988, pp. 223-224) puts it:

Despite considerable summarising of the mass of resulting data, the end product was relatively unmanageable for purposes of test construction. The test developers still faced the major problem, after all the needs analyses, of having to decide what to put into a test and what to leave out. They also faced the major problem of how to test what had to be tested: how, in other words, to operationalize the specifications. Moreover, the needs analysis provided no information on which such decisions could be based. The test developers were forced to rely upon their intuitions and experiences, both as test developers and as teachers of the students for whom the test was intended.

In other words, the problems that emerged stemmed from the use of an inappropriate needs-gathering instrument, rather than because of some fundamental flaw in the concept of basing test design on the results of a needs analysis. What is called for, therefore, is a more appropriate framework for the analysis of needs, one that matches more closely the purposes for which it is being used.

Having said this, however, empirical needs analysis of itself, however appropriate the instrument, will probably be unable to provide us with a sufficient level of understanding of EAP proficiency on which to base test design. This is because, as Hutchinson and Waters (1987, pp. 150-151) and Seaton (1987) point out and Alderson and Clapham (1992a) show, the current state of theoretical knowledge about language use in general, and in academic settings in particular, and on which needs analysis has to draw, is still insufficiently developed. Until this situation is remedied, there will also be a need to give due weight to concepts of needs based on soft data, for example, Alderson's "intuitions and experiences." As we will see, this approach was used to good effect in the ELTSREV studies that are reviewed next.

English Language Testing Service (ELTS) Revision Studies

As the final part of this section, I will briefly review some of the key aspects of the long series of studies and development tasks that have occurred in connection with the design of the two best-known recent British tests of EAP, namely the ELTS and IELTS. It is not easy to do justice to the many interesting points that emerge from the large number of documents of relevance, and so the reader is urged to refer for further information to the very useful summaries of much of the work to be found in Westaway, Alderson, and Clapham (1990) and Alderson and Clapham (1992b).⁸

⁸In the context of reporting on these studies, and those related to the development of the TEAP, mention should also be made of another recent attempt to design a major test of EAP, namely the Ontario Test of ESL (OTESL) (Wesche, 1987). This test was developed in order to act as a placement and proficiency measure for international students entering Ontario colleges and universities. The tests contain a number of features that distinguish them from tests such as TEAP, ELTS, and, to a lesser extent, IELTS, such as thematization of content and avoidance of nonauthentic item types in order to enhance the direct, performance-related nature of the test. On the other hand, the test construction process and many other features of the test itself are derived from the models established in

The ELTS test originated at a time when the notion of ESP in the sense of subject-specific English language teaching was in the ascendancy. Added impetus was given to this movement by Munby (1978), and the concept of subject-specific, needs-analysis-based test design emerged as a seemingly natural complement to the direction increasingly being taken by ESP/EAP course design. The ELTS test thus included as the second of its two main components a series of subject-specific (M) papers. Six subject-specific modules were developed: life sciences, medicine, physical sciences, social studies, technology, and general academic. However, the needs analysis profiles that underpinned these modules were drawn up largely on the basis of British Council staff expert opinion, rather than by direct observation (Carroll, 1981, pp. 69-71).

Some of the drawbacks in this approach are trenchantly criticised in Alderson and Hughes' work (1981). Chief among the reservations noted were the lack of solid research evidence for the inadequacy of previous measures (such as the "Davies Test" English Proficiency Test Battery [EPTB]) on the one hand, and on the other, the lack of adequate data to support the ESP language proficiency construct on which the ELTS specification was based.

Hamp-Lyons (1987) investigated to what extent the inclusion of a subject-specific writing question in the ELTS test was justified. Her subjects were 111 overseas postgraduate students, most of whom were studying at the University of Edinburgh, Scotland. Their specialisms, in terms of the ELTS categories, were as follows: general academic, 24; life sciences, 41; medical, 11; physical sciences, 7; and social studies, 28 (*ibid.*, p. 194).

The writing of the subjects on two subject-specific (SAP) questions and one general academic (GAP) one were analysed. No consistent distinction between performance in the two types of writing could be detected. On further investigation, it was revealed that raters applied general criteria to both SAP and GAP writing tests. Also, there was no clear SAP/GAP distinction among the three writing tests in terms of writers' responses, and difficulty levels and task demands were not uniform.

Hamp-Lyons' research throws considerable doubt on the validity of the content-specific part of the construct of EAP proficiency on which the ELTS test was based. Her findings indicate that, at least at present, there is no justification for the use of SAP rather than GAP writing tasks in EAP tests. Taken with the findings of other studies discussed in this review, for example, Johns (1981); Weir (1983); Criper and Davies (1988); Smith (1989); and Douglas and Selinker (1993); (also see Alderson & Urquhart 1983, 1985a, 1985b, and Clapham 1993), Hamp-Lyons' research adds further support to the view that there is no conclusive evidence that EAP proficiency differs significantly across subject areas.⁹

the development of the TEAP and ELTS tests, and so I have felt that to also discuss the OTESL at the same level of detail is redundant.

⁹Hale (1988) found a statistically significant relationship between student major and text content in an extensive sampling of TOEFL test results. However, as Hale himself observes, the effect was negligible in terms of test takers' overall TOEFL scores. Also, as Clapham (1993) points out, the sheer size of the sample reduces its

Criper and Davies (1988), in their report on the validation of the ELTS test, cast further doubt on the adequacy of the construct underlying the test:

Our analyses of ELTS confirm that the test can properly be described as a test of ESP and that it does set out to draw on a needs analysis. There are, however, in our view limitations in both areas. In the first, the lack of specificity as well as the uncertainty as to level . . . is in part a reflection of the weak content validity of the test, drawing too little on subject specialist opinion — in part a flaw in the theory of ESP itself. Like register analysis before it, ESP, both in teaching and in testing, falls down once it moves from the process of variation, variety and specific purpose to discrete entities which appear impossible to delineate and keep apart. The failure then is not in ELTS, but in the theory.

In the second area, that of needs analysis, as we have seen, ELTS was constructed with something of a needs analysis blueprint but in what was, it appears, a highly unsystematic way and also in a thoroughly non-empirical manner. Since needs analysis stands or falls by the empiricism it demands, it is regrettable that the needs analysis used for the construction of this first version of ELTS was not in itself the result of empirical investigation but rather, as far as we can see, the result of best guesses by language teachers. This activity may have helped in the construction of a good language test but in no way can it be regarded as an exercise in needs analysis (p. 99).

It is also worth noting that Criper and Davies (1988, p. 49) showed that, in terms of concurrent validity, the correlation between ELTS and the EPTB was .81, indicating a high measure of overlap in what the two tests were apparently testing. This finding obviously casts doubt on whether the subject-specific component of ELTS added significantly to the overall assessment picture.¹⁰

However, in terms of face validity, Criper and Davies (1988, p. 99) point out that “the modular approach as well as the needs analysis framework is hugely popular among subject specialists in universities and elsewhere who believe that language should be tied to its subject, a belief that should be open to investigation but which we have observed in many cases does not appear to invoke the scepticism that many academics rightly prize in their own research.”

true significance: “Once the number of subjects becomes very large almost anything can appear significant” (ibid., p. 259).

¹⁰Criper and Davies also report that “over 55% of the students who reported the ‘study skills’ (MI) test as easier than the ‘general reading’ (GI) test had score bands of 5.5 and below in both reading sub-tests with the higher scores being in MI. Of equal interest is the fact that over 58% of the students who perceived GI as easier had scores of 6.5 and above in both reading sub-tests with equal or higher scores in GI. Such results . . . tend to show that specific texts are particularly (if not more) beneficial to candidates with lower reading proficiencies” (Criper & Davies, 1988, p. 85).

Criper and Davies (1988, p. 108) also reported on the practical problems experienced by administrators of the ELTS test. It was perceived to be unnecessarily long and complicated, and there were difficulties in deciding on the choice of subject-specific modules.

As a result of Criper and Davies' report, a project to revise the ELTS test was undertaken. The revision process was informed by a wide variety of views. First, information was gathered from:

- receiving institutions
- British Council headquarters staff
- overseas test administrators
- EAP teachers
- language testers
- applied linguists

(Alderson & Clapham, 1992b, p. 3)

Next, a conference of British and overseas language testing researchers was held to review the data received and recommend what form a revised ELTS might take.

Finally, the Project Steering Committee made a number of decisions regarding the design of the revised test. Of particular interest to the concerns of this paper is the account of the thinking behind the decision about the fate of the subject-specific (M) component. The data collected showed that:

... there was no body of evidence yet available to support the ESP testing claims that different disciplines consistently demanded different linguistic skills, and that students were disadvantaged if they took a test which was not in their subject area (Alderson & Clapham, 1992b, p. 11).

However, Alderson and Clapham (p. 12) then go on to report that:

... almost all participants felt that one of the attractions of ELTS was the choice of subject modules. The receiving institutions, in particular, were very much in favour of them.

In addition, they point out that some of Alderson and Urquhart's (1985b) and Criper and Davies' (1988) findings also lent some support to the ESP construct. For both of these reasons, therefore, the steering committee decided that the revised form of the test would retain a modular (i.e., subject-specific) component.

The question then arose of how many subject modules there should be. Viewpoints on this issue varied widely. In the end, however, it was decided there should be three subject modules: Arts and Social Science, Physical Sciences and Technology, and Life and Medical Sciences. This tripartite division was chosen on the grounds that it corresponded closely to the main subject groupings revealed by an analysis of 1,000 ELTS Test Report Forms, and it was also the grouping recommended by

receiving institutions that were in favour of a rationalisation of the previous six-module structure (Alderson & Clapham, 1992b, p. 14).

Although the data revealed a feeling among some ELTS users that a different battery was needed for undergraduates than what was needed for postgraduates, the committee decided that for undergraduate candidates the rationalisation of test modules would lessen the problem of choice. Also, no body of research indicated that needs differed radically in terms of level, and so the committee decided that only one battery would be required for undergraduates and for postgraduates (ibid., pp. 14-15).

The consultative conference recommended that a test of lexis and structure be introduced into the revised battery on the grounds that there had been perhaps an overreaction against such a component in the past, and because of Hawkey's (1982) finding that a mixture of competence/performance measures produced superior results, and Weir's (1983) finding that such a component produced high internal consistency scores in relation to the rest of his test battery and was seen to have high face validity by testees. It was therefore decided that such a component should be introduced (Alderson & Clapham, 1992b, p. 16). However, subsequent trials showed that this sub-test correlated highly with the results of the test as a whole, and thus provided no additional information. It was therefore removed from the final version of the test (ibid., p. 19) (also see Alderson, 1993, for further discussion).

Alderson and Clapham (1992a) discuss further two main aspects of the revision process outlined above. First and foremost, they review the data collected from applied linguists as part of the consultation process in terms of views about matters such as "current testing theory," the "need for more research," "commonality," "variability," and so on. However, their overall conclusion was that, unfortunately, "the survey revealed that there appeared to be no dominant theoretical model that the ELTS Revision Project could use as a basis for test construction and construct validation" (ibid., p. 162). The authors go on to conclude that "the only consensus identified in our search for a model of language proficiency was on the need for more research" (ibid., p. 165).

With regard to the issue of needs analysis, Alderson and Clapham (1992a) report that:

. . . more than half the applied linguists wanted candidates to be given tasks which were as similar as possible to those they would meet during their future courses. Since several analyses have been carried out into the language needs of tertiary level students (in particular, Weir, 1983) we used these for the test specifications and tests, which we adjusted once they had been shown to a range of specialists in different disciplines. . . . In the event the tasks required in the three academic subject areas appeared to be so similar that the specifications for the test *tasks* are now virtually identical; only the reading *texts* differ (p. 163).

In other words, in reality, the needs analysis findings favoured an EAP rather than an ESP construct, even though, for reasons of face validity, it has been necessary to maintain a "fiction" to the contrary. In this respect, the position ultimately arrived is very similar to that reached by Weir (1983).

Conclusions. In this section, I wish to briefly summarise what I feel are the major implications, in overall terms, of the United Kingdom studies, as far as the TOEFL 2000 project is concerned.

First, although obvious, it is worth pointing out that a cardinal principle of EAP test design is that it should be based on an empirically researched concept of needs. However, there are two important corollaries of this position. The first is that, as Weir's experience in particular shows, the needs analysis instruments need to be designed with their eventual purpose kept as clearly in mind as possible. In other words, although the needs analysis may well be the first step in the test construction process, it should be informed by the overall characteristics of test development. In particular, the form of the needs analysis procedures must be such that the data they will yield can be readily operationalised in test construction terms. The second corollary is that any empirical form of needs analysis is only as powerful as the construct of EAP informing it. As we have seen, there are still significant gaps in our understanding in this area. As a consequence, it may be necessary to conduct needs analyses of both the soft, informed-opinion gathering kind as well as of the hard, empirical type.

Second, the majority of the studies in this section lend support to the EAP rather than the ESP construct. In other words, there seems little evidence that EAP needs vary significantly in subject-specialism terms. On this basis at least, there would appear to be good grounds for any new test of EAP being unitary rather than modular in construction. However, there is rather less evidence in the United Kingdom studies concerning the question of to what extent EAP needs vary by level (undergraduate vs. postgraduate), and therefore no firm conclusions can be drawn from these studies concerning this issue.

Furthermore, the United Kingdom studies appear to be based on an assumption that EAP involves uses of English that differ significantly from those required in nonacademic contexts, and therefore a test that differs from the kind used to assess general English language proficiency is needed. However, despite the apparent reasonableness of this position, it is important to note that this position rests largely on assumption rather than on proven fact. Research effort has been devoted mainly to analysing the use of English in academic settings rather than to comparing the use of English in academic settings with use in nonacademic settings. In this connection, it is also worth bearing in mind that indigenous British students, like their North American counterparts, are not expected to demonstrate proficiency in EAP as a condition of entry into higher education; rather, their general, native-speaker knowledge of English is deemed sufficient. Why then should the situation be different for international students?

However, the third major point that emerges from the United Kingdom studies is that, despite what may or may not be the case in theoretical terms, there is a strong feeling among applied linguists, EAP teachers, subject specialists, and students (Cripser & Davies, 1988) alike that, to be valid, the subject matter of an EAP test should be distinguished at least in terms of a broad science/nonscience split, and that the items contained in the test should be direct and based on performance. The former point may well be a function of the early specialization tendency of British higher education, and so comparable pressures may well not be felt in the North American context. However, the latter point appears to reflect a more widely shared stance and may be more difficult to resist, despite the apparent difficulties of reconciling the inclusion of items of this kind with an attempt to measure general EAP competence. In

other words, in the development of TOEFL 2000, there may be aspects of test face validity that it is felt important to take into account and that, at first sight, conflict with aspects of test construct validity that are felt to be equally important.

However, on the basis of the TEAP and IELTS experiences, it would appear that it is feasible to devise test items that have the necessary performance-related appearance but are general enough in their underlying characteristics to yield useful indications of overall EAP competence. Furthermore, the demand for performance-derived test items would not appear to be absolute. Both the TEAP and IELTS tests originally contained items of this kind, and, although they were eventually discarded, this was not due to the pressure of market forces but rather a result of the items not appearing to add significantly to the information being yielded by the other parts of the test. Test users and other interested parties therefore appeared to be satisfied with a mixture of so-called performance-related test items and other items more concerned with measuring general competence (cf., Hawkey's, 1982) finding that a mixture of performance-oriented and competence-oriented items tended to yield the most reliable results, a happy — however lamentably rare — coincidence of research and “lay” views! In any case, one would anticipate that the much more widespread faith (in the United States at least) in multiple-choice and other similar test-item formats often used for competence-related test items, would lead to relatively lower resistance to their inclusion as one of the item types in any new United States test of EAP. The overall point is that the United Kingdom experience indicates that, in the current climate, any major new EAP test will need to be artfully constructed so that potentially opposed research findings and theories on the one hand, and lay opinion on the other, are satisfactorily reconciled.

Finally, the fourth main implication of the United Kingdom research reviewed here, particularly that of Hawkey, relates to the question of whether the concept of EAP needs is only (or even mainly) satisfactorily conceived of as a language problem, or whether it should also be thought of in terms of a problem of becoming acculturated to a foreign academic milieu. In addition to the evidence for the latter view that exists in the North American research I have reviewed, there is a large body of other United Kingdom work that points in a similar direction (see, e.g., the papers in collections such as Greenall & Price, 1980, and Adams, Heaton, & Howarth, 1991). In my opinion, there is in fact already ample evidence for the view that EAP needs are both linguistic and cultural. The question is thus whether the TOEFL 2000 project should concern itself with this issue. As I have already made clear, I feel that it is difficult to see how an initiative of one kind or another in this direction can be avoided.

Discussion and Conclusions

What are the overall implications for the TOEFL 2000 project of the work reviewed above?

Summary of Research Findings

Unfortunately, it has to be said at the outset that because of its provenance, it is difficult to draw any overall conclusions from the data reviewed in this paper. The studies of North American EAP needs have not formed part of a region-wide, systematic, coordinated overall programme of enquiry. The “story-line” therefore tends to be discursive rather than straightforward. Also, the vast majority of the studies discussed in this paper are pedagogic or research-oriented, rather than test-based. In other words, the data has usually been produced in order to throw light on a teaching problem or to explore the effectiveness of certain kinds of research methods. Although relevant, the research findings resulting from such areas of enquiry, and emerging in a relatively ad hoc way, cannot adequately substitute for an organised programme of investigation aimed directly at the concerns of the TOEFL 2000 project.

Research into writing skills is a good case in point. As should be clear from the work reviewed above, this has been one of the areas where a good deal of relatively thorough research has occurred, much of it in response to or based on earlier studies. There is therefore more of a connection between individual pieces of research here than is the case for some of the other main areas that have been investigated. Unfortunately, however, we still do not have a taxonomy of North American higher education writing skills about which there is general agreement, nor, would it appear, is there any consensus about to what extent and in what ways writing requirements vary by specialism or level. This is not to say that the existing research is not useful. Given its limitations, however, it is better viewed as a basis for further research rather than as a full-fledged database.

Much the same point can be made for all the main areas of research reviewed. It should be clear from the summaries provided that, in research terms, we simply lack a coherent, overall picture of what EAP proficiency in North American higher education involves in each of the other main skills (listening, speaking, and reading), as well as with respect to areas such as vocabulary and communication strategies. Furthermore, for certain other areas, such as skills-integrated activities and subskills that cross skills boundaries, there appears to be no substantial body of work extant (although the latter is adumbrated by some of the studies that focus on other matters, Jacobson (1986) for example). It is also the case for all these skills that none of the constituent parts of the picture is fully developed. That is, there is no truly authoritative body of research about how these skills may vary by specialism and/or level. The need here is therefore also for a programme of research that builds on what has been done and extends the enquiry into areas that have so far been largely neglected.

In short, research must be undertaken that would provide a North American test database equivalent to that produced in the United Kingdom by studies such as those undertaken by Weir (1983) and Hawkey (1982).¹¹

¹¹Because such a database is currently lacking, I have been unable to reach any meaningful conclusions regarding similarities and differences between EAP needs in the United States and United Kingdom. However,

The only possible exception to this rather depressing state of affairs is the existing body of work that explores the extent to which EAP proficiency varies according to topic. Here, there appears to be reliable evidence that the level of variation is insignificant. In other words, an EAP proficiency test, in terms of construct validity, need not contain subject-specific components. However, as has also been shown, the need in terms of face validity may be another matter. Also, there has been little research concerning the extent to which academic settings involve a type of language proficiency different from that needed in nonacademic settings, and this issue therefore obviously merits further investigation.

Summary of Research Methods

The studies surveyed evince a wide variety of research methods. Any further research should find this a rich resource on which to draw, in terms of illustrating the many methods of enquiry that are possible in this area and the potential and limitations of each of these methods. In the course of reviewing the reports, I have already tried to indicate what I feel are a number of the important research methodology issues presented by these studies. In what follows, I will briefly recapitulate on these. It should be noted, however, that I have deliberately oversimplified the nature of and the differences between the various research approaches, for the sake of clarity.

Bird's eye and worm's eye views. Some studies have adopted a panoramic perspective (e.g., Ostler, 1980, Canseco & Byrd, 1989, etc.), and this is obviously useful for getting the kind of general, overall perspective that is essential if the input to a test such as the TOEFL is not to reach unmanageable proportions. On the other hand, only by taking a close look at the constituent parts of the overall picture can any kind of survey point in the right direction. The worm's eye view of studies such as Jacobson (1986), Rounds (1987a), Benson (1989, 1991), and others can provide this essential prior perspective.

Perception-based versus documentary-based studies. All research studies are ultimately perception-based, of course, but some are more so than others. As with all the issues discussed in this section, however, the point is that both perception-based and documentation-based research studies have an important role to play, and are best seen as complementary. By perception-based research is meant studies such as those of Johns (1981), Bridgeman and Carlson (1983), and so on, in which the primary source of data was a survey of views. However, in deciding what views to survey, adequate prior investigation of documentary evidence is essential, as evidenced in research such as Horowitz (1986a). In other words, both hard and soft data need to be collected, compared, and synthesised for the picture to begin to emerge properly. Apparently opposed research approaches need to be seen as complementary, if a reasonably sound concept of needs is to be identified.

"Situated" versus experimental. By "situated" I mean a research approach that starts from the inside, and tries to work outwards. Such an approach attempts to study the phenomenon to the greatest possible

work could be done on this point by, for example, interviewing Junior Year Abroad (JYA) students and the North American Office Director at Lancaster University.

extent within its natural ecology, and strives to make the research procedures as unintrusive as possible within the constraints of the variables. The work of Rounds (1987a) and Jacobson (1986) are good illustrations of this approach. The experimental approach, on the other hand, is typically a matter of working from a pre-determined external point of reference, and then proceeding inwards. A preordained framework is used to limit the variables, so that a clear understanding of a particular aspect of the subject under investigation can be obtained. Chaudron and Richards (1986) and Dunkel (1988) provide clear examples of this type of research approach. Once again, a more valid picture is likely to emerge when a combination of these approaches is employed.

Triangulation. Research that incorporates multiple points of reference is likely to give a more reliable picture of events, especially if any of the research is perception-based. Variety in points of reference can occur in terms of the research methods used (e.g., quantitative vs. qualitative), the perspectives taken into account (e.g., faculty, international student, indigenous student), and the type of data collected (hard vs. soft). Studies such as those by Jacobson (1986), Hawkey (1982) and Weir (1983) are good examples of the use that can be made of such a framework within the study of EAP needs. In addition, there is a particular need to compare the perspectives of findings related to nonnative-speaker students with findings for native speakers, if the resulting picture is to be at all realistic.

Empirical research versus expert opinion. There is no doubt that empirical research is necessary to account for the nature of the communicative competence needed for academic settings. This is because, unlike the Chomskyan notion of linguistic competence, which is a property of the mind, communicative competence is a product of the psychological and social characteristics of situations in which language is used for communication.¹² Ultimately, therefore, the nature of the communicative competence required can only be determined through empirical study of those communication situations. However, this does not necessarily mean that our understanding of the nature of communicative competence is limited to evidence obtained by formal research. As Alderson and Clapham (1992a, 1992b) have shown with reference to the ELTS revision project, valuable information about EAP needs, at all sorts of levels, can also be obtained by surveys of expert opinion. Indeed, as with all the issues discussed in this section, the relationship between research data and expert opinion needs to be seen as a complementary one. For research to be given direction, it has to be informed by some kind of underlying hypothesis, however ill-formed. Not even so-called “hypothesis-generating” research will, of itself, produce the concepts that are needed to guide it. Therefore research must always be informed, to a varying extent, by expert opinion. However, it is equally vital to refine and develop intuition through research data, if it is to have empirical grounding and be liberated from “the tyranny of the particular.”

¹²I should point out that I am, of course, assuming linguistic competence to be part of communicative competence.

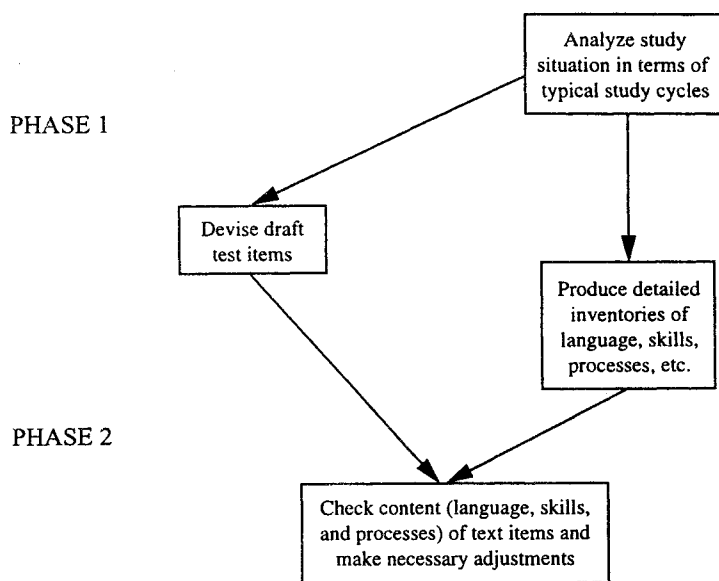
Proposals for Further Research

In this section, I wish to outline the main features of a programme of further research designed to remedy the gaps identified above. (These proposals should be taken in conjunction with the other recommendations for further research that I have made throughout this report.)

The needs analysis instruments and procedures should be geared to the overall purposes of the research. In other words, it should be possible to operationalise the findings of the needs analysis, in test item terms, in a relatively straightforward manner. In practical terms, this might involve, for example, an initial focus on broad and general categories of analysis, such as the typical sequence of study activities students need to go through in order to transform inputs from lab sessions, lectures, group discussions, readings, and so on into some kind of written/spoken output. Examples of cycles of this kind could then be used to generate putative test items.

Meanwhile, as a subsequent step, a more detailed analysis of the skills and processes involved in the actual study cycles could be undertaken. Then the test items themselves could also be analysed in a similar way. Finally, the two inventories could be compared, and any necessary adjustments to the test items made.

The basic principle behind this approach is that the needs analysis procedure should be flexible and dynamic, in keeping with the different phases and the overall purposes of the research. The key to achieving this is not to allow too much detail to enter the picture at the initial, item-writing stage. This process as a whole may be represented thus (cf. Hutchinson & Waters, 1987, p. 93):



A comprehensive and representative framework for data-gathering should be established. This feature of the research programme involves thinking through who exactly, in student terms, should be the focus of the research; what type(s) of institution(s) will be appropriate for the research; what levels within the higher education system should the research concern itself with; what subject specialisms should be targeted; and what teaching-learning methods should be surveyed.

With regard to the student population, as Swales (1994, p. 2) points out, the situation is a complex and changing one. There would appear to be a need, in particular, to distinguish between permanent-resident or immigrant students, on the one hand, and international students proper on the other. These two groups may well differ significantly in terms of linguistic and cultural knowledge. Furthermore, as Swales (1994) also points out, student demographics in United States higher education have changed significantly since the early 1980s, and there is every possibility that they may once again change significantly over the next 10 to 15 years. Thus, the question of who the needs analysis should focus on will depend on carefully thinking out who the test takers are likely to be.

The second issue in this area is concerned with what type(s) of institution the research should focus on. As Swales (1994, p. 3) points out, there are considerable differences between North American higher education institutions in terms of "size, wealth, levels of degree, admissions policy, amount of research, teaching loads, and so forth." Given such heterogeneity, ensuring that any study of needs is representative is tricky. Much previous research has attempted to narrow the field by focusing primarily on institutions with relatively high proportions of international students. As a further strategy, it would be worthwhile adopting Swales' (1994) recommendation that

... undergraduate and graduate populations are separated for test-development purposes. It seems reasonable to posit a fairly elite-type institution for NNS graduate applicants, but a more "open" one for NNS undergraduates, if only because this division accords quite well with current realities (p. 3).

Which level(s) the research should focus on is the third issue that should be considered. The evidence in the literature points in both directions. For example, Bridgeman and Carlson (1983) has been widely criticised for failing to adequately address this issue and thus ignoring what others have perceived to be important differences in writing requirements between the undergraduate and the graduate levels. On the other hand, as Alderson and Clapham (1992b) point out, the research supporting the ELTS revision project failed to produce clear evidence of the need to distinguish levels, although whether this issue was actually addressed by the majority of the research, or simply overlooked, is not clear. On balance, it would seem wise to develop a pilot study for investigating this issue further, with additional research contingent on the results.

The fourth question is that of subject specialism. As with type of institution, it has been traditional to focus on the subject areas that attract the highest numbers of international students. However, as Carson (1994, p. 5) points out, "undergraduate students must be successful in core courses to graduate (or to even proceed to upper division major courses), and so studies that focus on popular NNS majors at the expense of general education courses seem to me to be of little value" (see also Hutchinson & Waters,

1981). It is obviously important to study the entire curriculum the student will experience, not just the more specialized aspects of that curriculum. Also, as with the issue of student demographics discussed above, it will be important to try to take into account likely future trends in areas of study.

The fifth issue is to do with subject grouping. Should subject specialisms be studied at the level of the department and the faculty, or in terms of very broad groupings, such as arts/humanities and science/engineering? My own view is that, in order to reflect the current level of understanding and debate on this matter, it would be best to find a position that takes both of these views into account. In other words, the research might be divided up into three broad areas in terms of subject specialism — arts, social sciences, and hard sciences.

Finally, what teaching-learning methods should the research focus on? Swales (1994) argues that a historical survey of needs analysis research has to be set against evolving views about educational objectives:

We should be aware, I believe, that quite a lot has changed in U.S. universities over the last 15 years, particularly at the undergraduate level. The spread of computers, the attempts to develop critical reading and critical thinking, the proliferation of diverse writing tasks and options, p.c. sensitivities to ways of talking and writing, increase in student collaborative projects, peer-critiquing, and some decline in the use of multiple-choice test formats may expose further and deeper “lacks” in international students applying from traditional educational cultures (p. 3).

Thus it would seem that, here again, research must try to anticipate and take into account current and likely future trends.

The research design should be triangulated, in terms of sources and types of data and methods of data-gathering. Most of the North American research reviewed in this paper looks at EAP needs from only one of the three possible principal points of view: the student, the faculty, and the observer/researcher. This inevitably means that the resulting understanding of needs is not as rich as it might be. I therefore recommend that future research follow a pattern similar to that of Weir (1983) and Wesche (1987) in this regard, that is, data is gathered from all three angles.¹³ A similar approach should be adopted with respect to types of data, so that both hard data (e.g., documentary evidence) and soft data (e.g., opinion) are taken into account, since both types of data are necessary but, on their own, insufficient. The same can be said for methods of data-gathering. Some aspects of needs lend themselves best to quantitative forms of data gathering, while others are better suited to a more qualitative style of investigation. Research that employs both approaches in appropriate ways is therefore likely to produce better results.

¹³Also well worth examining for the light it throws on student self-evaluations as a source of data about EAP needs is Banko (1994).

The research should be based on a rich concept of needs. The literature on types of need makes a basic distinction between needs in the sense of “necessities,” on the one hand, and “lacks” on the other (see e.g., Hutchinson & Waters, 1987, and West, 1994). Necessities refers to what the student needs to know and do in order to function in the target communication situation. Lacks are the necessities that the student does not already possess. An EAP test should therefore focus on needs in the sense of lacks. As far as the research programme in question is concerned, in terms of identifying necessities, a base-line study of what constitutes an acceptable level of native-speaker performance needs to be conducted (cf., e.g., Micheau & Billmyer, 1987). Only in this way can we begin to identify the typical lacks of the nonnative-speaker student with any accuracy.

However, it is also vital to take into account needs in the sense of “wants” (see, e.g., Hutchinson & Waters, 1987). Wants are necessities and lacks as seen from the student’s point of view. This is the student’s view of needs and is an important complement to the researcher’s or teacher’s perspective of necessities and lacks. Although a number of North American research studies have adopted this perspective with respect to student views about necessities (e.g., Ostler, 1980, and Christison & Krahne, 1986), less work has been done in the area of student views about lacks. However, the value of such research is well-illustrated in the recent studies of Leki and Carson (1994) and Banko (1994). In other words, the student’s own perspective on the realities of study, in terms of necessities, and in particular, in terms of lacks, can often provide important insights regarding priorities for needs, and the research programme I am advocating would include methods for taking these perspectives into account.

The research should identify the general competence needed for study in the North American context. The merest glance at the results of extant research into EAP needs reveals a picture characterised by heterogeneity (cf. Johns, 1988). Differences and particularities abound, in terms of research categories, student levels, and specialisms. What is more, the potential for variation is ultimately such that a comprehensive description of EAP needs would have to go well beyond the level of investigation reached so far and arrive at a description of what it is like to be a typical student doing assignment A while being taught by Professor B in course C at institution D in country E! As Alderson puts it, “The danger is that in specifying communicative performance, one might end up describing an impossible variety of situations, which one cannot encompass for testing purposes” (1981b, p. 58-59).

The point, however, is that for testing purposes, the study of academic communication situations is not an end in itself. Rather, its purpose is to derive from such study the attributes of the underlying academic communicative competence that enables the student to cope, despite the myriad particularities of each individual situation. That such a competence exists is surely self-evident, since the qualifications for entering the system do not involve already having done “assignment A while being taught by Professor B in course C at institution D in country E.” Instead, at the undergraduate level it is expected that the requisite high school level of knowledge has been obtained, and, at the graduate level, that the right kind of undergraduate level of knowledge has been reached. These more general types of knowledge and abilities are viewed as the resource that enables the student to successfully handle the next, more specialised, stage. It therefore follows that a test of EAP should focus on this type of competence, at the relevant level. Thus, although academic situations and tasks undoubtedly vary

enormously, as much EAP research shows, it is important to look beyond the specifics of the performance data in order to identify the area of common ground.

Examples of such an approach in the literature reviewed in this paper are unfortunately rare. However, Rounds' research into the "elaborated" style of spoken discourse needed by TAs (Rounds, 1987a), Parry's conclusion that students need a stock of "bridging" vocabulary and strategies for "sizing up" the relative importance of new terminology (Parry, 1991), and Canseco and Byrd's identification of the need for students to actively interpret writing requirements (Canseco & Byrd, 1989) are cases in point. Also, another area of increasing importance in EAP — that of "genre analysis" (Swales, 1990) — provides a very relevant further illustration of the main point being made in this section. In terms of the genre concept, the academic world as a whole and each of its constituent parts can be seen as consisting of a series of "discourse communities." The chief characteristic of a discourse community is that it shares certain common purposes, and, in order to achieve them, a common set of conventions. Another major feature of a discourse community is that it possesses one or more genres, or distinctive forms of communication, which it uses as a means of realising its purposes. The student is seen as someone who, through accomplishing a series of tasks (Swales, 1990, p. 76), gradually approximates more and more closely to full membership of the discourse community, by mastering its genres (conventional modes of discourse). Swales (1990, Part III) provides numerous examples of what the concept of genre means in academic discourse communities, particularly in terms of the various productions associated with the academic research process.

Genres are characterised by their distinctiveness. They act as boundary markers, a way of setting off the interests and pursuits of one part of the academic world from another, as well as the academic world from the lay world. There is thus considerable variation in types of genres (Swales, 1990, pp. 61 ff). As a result, it would appear that the main result to emerge from genre analysis — the most recent and in many ways the most powerful of our tools for analysing academic discourse — is a compounding of the original problem. In other words, the picture of needs now becomes even more heterogeneous than before. But this is to look at the matter only in terms of its surface manifestations. The real point is that the academic discourse community functions as a learning system. Its purpose is thus to enable the student to be inculcated into its practices. It is not expected that the student will have already mastered the practices that the discourse community teaches. What the student therefore needs to bring to the discourse community is the ability and willingness to learn what the discourse community exists to teach.

A good example of what this can involve in practice is illustrated by Johns (1993). In the first part of her article, she describes the research she pursued into the nature of the "grant proposal" genre. Basically, this showed that getting the genre right was a matter of patient detective work on the part of the writers, in terms of discovering what constituted an effective grant proposal in the eyes of the intended audience. She then describes how she transferred this principle to an ESL writing class in which the students were given the task of writing to the trustees of their university to protest against a planned increase in their tuition fees. This involved the students in a great deal of research into the background of their intended audience, in much the same manner as that conducted by the grant proposal writers. In other words, what specificity of genre really implies is that, whatever the particular genre the

student seeks to master, he or she must have the resources and be willing to identify the distinctive features of the genre in question. The student is thus not expected to have already mastered the specifics of the genre. Rather, the student is expected to have the competence necessary to eventually master the genre.

What might constitute the major components of an “academic communicative competence” of the kind outlined above? Several taxonomies already exist, and a synthesis of them might be used as a starting point for researching this question. The taxonomies I have in mind are those to be found in Weir (1983), Council of Ontario Universities (1986) (as reported in Wesche, 1987), Banko (1994), and Waters and Waters (1995). The resulting synthesis could be refined and developed through further empirical research and/or by eliciting the views of concerned parties (cf. Alderson and Clapham 1992a).

Overall Conclusion

In this review, I have tried to show that the existing literature on EAP needs in the North American higher education context, despite its intrinsic quality, is in most respects an inadequate basis for determining the kinds of EAP needs that should underpin the TOEFL 2000 project. To remedy this problem, I have therefore recommended that a programme of further research should be carried out, which, as far as possible, builds on existing work and is informed by the lessons it offers in terms of formulating an appropriate research approach.

A research programme of this kind is a major undertaking and can therefore only be justified if truly essential. I believe, however, that I have provided ample grounds for my recommendations. I feel confident that because of the concept of needs and the design informing the proposed research, the results would provide the kind of database needed for the TOEFL 2000 project. I have also been at pains in my proposals to bear in mind the realities of test construction, especially with regards to the interfacing of needs analysis and test-item development. In the past, proceeding in a coherent way from a concept of needs to test development has been the least successful aspect of attempts to devise empirically validated tests. It is to be hoped that, in the research in question, the solution to this problem has been found, and the results of the research will therefore not only be of value in underpinning the development of TOEFL 2000, but also in the development of empirically based tests in general.

Bibliography

- Adams, P., Heaton, B., & Howarth, P. (Eds.). (1991). Socio-Cultural issues in English for academic purposes, *Review of ELT*, 1(2).
- Alderson, J. C. (1981a). Report of the discussion on testing English for specific purposes. In J. C. Alderson, & A. Hughes (Eds.), *Issues in language testing* (pp. 123-134). (ELT Documents No. 111). London: The British Council.
- Alderson, J. C. (1981b). Report of the discussion on testing English for specific purposes. In J. C. Alderson, & A. Hughes (Eds.), *Issues in language testing* (pp. 55-65). (ELT Documents No. 111). London: The British Council.
- Alderson, J. C. (1988). New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing*, 5(2), 220-232.
- Alderson, J. C. (1993). The relationship between grammar and reading in an English for academic purposes test battery. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 203-219). Alexandria, VA: TESOL.
- Alderson, J. C., & Clapham, C. (1992a). Applied linguistics and language testing: A case study of the ELTS test. *Applied Linguistics*, 13(2), 149-167.
- Alderson, J. C., & Clapham, C. (Eds.). (1992b). *Examining the ELTS Test: An account of the first stage of the ELTS revision project*. (IELTS Research Report No. 2). Cambridge, England: The British Council/University of Cambridge Local Examinations Syndicate.
- Alderson, J. C., & Hughes, A. (Eds.). (1981). *Issues in language testing*. (ELT Documents No. 111). London: The British Council.
- Alderson, J. C., & Urquhart, A. H. (1983). The effect of student background discipline on comprehension: A Pilot Study. In A. Hughes & D. Porter (Eds.), *Current developments in language testing* (pp. 121-127). London: Academic Press.
- Alderson, J. C., & Urquhart, A. H. (1985a). The effect of students academic discipline on their performance on ESP reading tests. *Language Testing*, 2(2), 192-204.
- Alderson, J. C., & Urquhart, A. H. (1985b). This test is unfair: I'm not an economist. In R. Hauptman, R. leBlanc, & M. B. Wesche (Eds.), *Second language performance testing* (pp. 25-43). Ottawa: University of Ottawa Press.
- Bachman, L. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671-704.

-
- Banko, R. (1994). *Self-assessment of academic language ability: Its relationship to performance and the effects of learner variables*. Unpublished doctoral thesis, Lancaster University, England.
- Benesch, S. (1993). Critical thinking: A learning process for democracy. *TESOL Quarterly*, 27(3), 545-548.
- Benson, M. J. (1989). The academic listening task: A case study. *TESOL Quarterly*, 23(3), 421-445.
- Benson, M. J. (1991). University ESL reading: A content analysis. *ESP Journal*, 10(2), 75-88.
- Blue, G. (Ed.). (1993). Language, learning and success: Studying through English. *Review of ELT*, 3(1). London: Modern English Publications/The British Council.
- Bowyer, L. (1993). Individual pursuits: Preparing students for tertiary study through EAP programs. In R. Khoo (Ed.), *The practice of LSP: Perspectives, programmes and projects* (pp. 196-208). (Anthology Series No. 34). Singapore: SEAMEO Regional Language Centre.
- Braine, G. (1989). Writing in science and technology: An analysis of assignments from ten undergraduate courses. *ESP Journal*, 8(1), 3-16.
- Bransford, J. D., Stein, B. S., & Shelton, T. (1984). Learning from the perspective of the comprehender. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 28-44). London: Longman.
- Bridgeman, B., & Carlson, S. B. (1983). *Survey of academic writing tasks required of graduate and undergraduate foreign students*. (TOEFL Research Report No. 15). Princeton, NJ: Educational Testing Service.
- Canseco, G., & Byrd, P. (1989). Writing required in graduate courses in business administration. *TESOL Quarterly*, 23(2), 305-316.
- Carlson, S. B., Bridgeman, B., Camp, R., & Waanders, J. (1985). *Relationship of admission test scores to writing performance of native and nonnative speakers of English*. (TOEFL Research Report No. 19). Princeton, NJ: Educational Testing Service.
- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21(3), 461-481.
- Carroll, B. (1981). Specifications for an English Language Testing Service. In J. C. Alderson & A. Hughes (Eds.), *Issues in language testing* (pp. 66-110). (ELT Documents No. 111). London: The British Council.
- Carson, J. (1994). Review of EAP paper. Unpublished paper, Educational Testing Service, Princeton.

-
- Casanave, C. P., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *ESP Journal*, 11(1), 33-49.
- Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field independence as predictors of proficiency in English as a second language. *Language Learning*, 36(1), 27-45.
- Chaudron, C., & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7(2), 113-127.
- Christison, M. A., & Krahne, K. J. (1986). Student perceptions of academic language study. *TESOL Quarterly*, 20(1), 61-81.
- Clapham, C. (1993). Is ESP testing justified? In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 257-271). Alexandria, VA: TESOL.
- Clark, R., Fairclough, N., Ivanic, R., & Martin-Jones, M. (1990). Critical language awareness Part 1: A critical review of three current approaches to language awareness. *Language and Education*, 4(4), 249-260.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389-99.
- Criper, C., & Davies, A. (1988). *ELTS validation project report* (Research Report No. 1 [i]). Cambridge, England: The British Council/University of Cambridge Local Examinations Syndicate.
- DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *ESP Journal*, 7(2), 91-102.
- Douglas, D., & Selinker, L. (1993). Performance on a general versus a field-specific test of speaking proficiency by international teaching assistants. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 235-256). Alexandria, VA: TESOL.
- Dudley-Evans, A., & Henderson, W. (Eds.). (1990). *The language of economics: The analysis of economics discourse*. (ELT Documents No. 134). London: Modern English Publications/The British Council.
- Dudley-Evans, A., & Johns, T. F. (1981). A team teaching approach to lecture comprehension for overseas students. In *The teaching of listening comprehension* (pp. 30-46). London: The British Council.
- Dunkel, P. (1988). The content of L1 and L2 students' lecture notes and its relation to test performance. *TESOL Quarterly*, 22(2), 260-279.

-
- Dunkel, P., & Davy, S. (1989). The heuristic of lecture notetaking: Perceptions regarding the value and practice of notetaking. *ESP Journal*, 8(1), 33-50.
- Egglestone, J. F., Galton, M. J., & Jones, M. E. (1975). *A science teaching observation schedule*. (Schools Council Research Studies). London: Macmillan Educational.
- Gibbs, G. (1981). *Teaching students to learn*. Milton Keynes, England: Open University Press.
- Greenall, G. M., & Price, J. E. (Eds.). (1980). *Study modes and academic development of overseas students*. (ELT Documents No. 109). London: The British Council.
- Greenberg, K. L. (1986). The development and validation of the TOEFL writing test: A discussion of TOEFL Research Report No. 15 and No. 19. *TESOL Quarterly*, 20(3), 531-544.
- Hale, G. A. (1988). *The interaction of student major-field group and text content in TOEFL reading comprehension*. (TOEFL Research Report No. 25). Princeton, NJ: Educational Testing Service.
- Hamp-Lyons, E. M. (1987). Testing second language writing in academic settings. Unpublished doctoral thesis, University of Edinburgh, Edinburgh, Scotland.
- Hartley, J., & Davies, I. K. (1978). Notetaking: A critical review. *Programmed Learning and Educational Technology*, 15, 207-224.
- Hawkey, R. (1982). An investigation of inter-relationships between cognitive/affective and social factors and language learning. Unpublished doctoral thesis, University of London, London, England.
- Horowitz, D. M. (1986a). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20(3), 445-462.
- Horowitz, D. M. (1986b). Essay examination prompts and the teaching of examination writing. *ESP Journal*, 5(2), 107-120.
- Hutchinson T., & Waters, A. (1981). Performance and competence in ESP. *Applied Linguistics*, 2(1), 56-69.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge, England: Cambridge University Press.
- Hutchinson, T., Waters, A., & Breen, M. (1979). An English language curriculum for technical students. In *Practical papers in English language education*, Vol. 2 (pp. 146-171). Institute for English Language Education, Lancaster University, England.

-
- Jacobson, W. H. (1986). An assessment of the communication needs of nonnative speakers of English in an undergraduate physics lab. *ESP Journal*, 5(2), 173-187.
- Jenkins, S., Jordan, M. K., & Weiland, P. O. (1993). The role of writing in graduate engineering education: A survey of faculty beliefs and practices. *ESP Journal*, 12(1), 51-67.
- Johns, A. M. (1981). Necessary English: A faculty survey. *TESOL Quarterly*, 15(1), 51-57.
- Johns, A. M. (1988). The discourse communities dilemma: Identifying transferable skills for the academic milieu. *ESP Journal*, 7(1), 55-59.
- Johns, A. M. (1993). Written argumentation for real audiences: Suggestions for teacher research and classroom practice. *TESOL Quarterly*, 27(1), 75-90.
- Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, 11(3), 253-271.
- Kirkland, M. R., & Sanders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105-121.
- Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.
- Marshall, L. A., & Rowland, F. (1983). *A guide to studying independently*. Milton Keynes, England: Open University Press.
- Martin, A. V. (1989). *Bridging vocabulary: An essential component of ESL proficiency*. Paper presented at the 23rd Annual TESOL Convention, San Antonio, TX.
- Marton, F. (1974). What does it take to learn? In N. J. Entwistle & D. Hounsell (Eds.), *How students learn* (pp. 125-138). Institute for Post-Compulsory Education, Lancaster University.
- Micheau, C., & Billmyer, K. (1987). Discourse strategies for foreign business students: Preliminary research findings. *ESP Journal*, 6(2), 87-97.
- Munby, J. (1978). *Communicative syllabus design*. London: Cambridge University Press.
- Olsen, L. A., & Huckin, T. N. (1990). Point-driven understanding in engineering lecture comprehension. *ESP Journal*, 9(1), 33-47.

-
- Ostler, S. E. (1980). A survey of academic needs for advanced ESL. *TESOL Quarterly*, 14(4), 489-502.
- Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly*, 25(4), 629-653.
- Raimes, A. (1990). The TOEFL Test of Written English: Causes for concern. *TESOL Quarterly*, 24(3), 427-442.
- Raimes, A. (1992). The author responds to Traugott, Dunkel, and Carrell. *TESOL Quarterly*, 26(1), 179-180.
- Robinson, P. (Ed.). (1988). *Academic writing: Process and product*. (ELT Documents No. 129). London: Modern English Publications/The British Council.
- Rounds, P. L. (1987a). Characterizing successful classroom discourse for NNS teaching assistant training. *TESOL Quarterly*, 21(4), 643-671.
- Rounds, P. L. (1987b). Multifunctional personal pronoun use in an educational setting. *ESP Journal*, 6(1), 13-29.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22(1), 69-90.
- Schmidt, M. (1981). Needs assessment in English for specific purposes: The case study. In L. Selinker, E. Tarone, & V. Hanzeli (Eds.), *English for academic and technical purposes: Studies in honour of Louis Trimble* (pp. 199-210). Rowley, Massachusetts: Newbury House.
- Seaton, I. (1987). Comments on Wesche: Second language performance testing: The Ontario Test of ESL as an example. *Language Testing*, 4(1), 48-50.
- Shih, M. (1992). Beyond comprehension exercises in the ESL academic reading class. *TESOL Quarterly*, 26(2), 289-318.
- Smith, J. (1989). Topic and variation in ITA oral proficiency: SPEAK and field-specific oral tests. *ESP Journal*, 8(2), 155-167.
- Sorensen, K. (1982). Modifying an ESP course syllabus and materials through a teacher-planned needs assessment. In *ESL Working Papers No. 2* (pp. 1-17). University of Minnesota, Minneapolis.
- Swales, J. (1982). Examining examination papers. *English Language Research Journal* 3, 9-25.
- Swales, J. (1990). *Genre analysis*. Cambridge, England: Cambridge University Press.

-
- Swales, J. (1994). Commentary on "A review of research into needs in English for academic purposes of relevance to the North American higher education context." Unpublished paper, Educational Testing Service, Princeton.
- Traugott, E., Dunkel, P., & Carrell, P. (1992). An acknowledgement of concern and a concern for lack of acknowledgement. *TESOL Quarterly*, 26(1), 180-185.
- Trimble, L. (1985). *English for science and technology*. Cambridge, England: Cambridge University Press.
- Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly*, 26(4), 713-729.
- Vann, R., Meyer, D., & Lorenz, F. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly*, 18(3), 427-440.
- Wall, D. (1981). A Pre-sessional academic writing course for postgraduate students in economics. In *Practical papers in English language education, Vol. 4* (pp. 31-105). Institute for English Language Education, Lancaster University, England.
- Waters, M., & Waters, A. (1992). Study skills and study competence: Getting the priorities right. *ELT Journal*, 46(3), 264-273.
- Waters, M., & Waters, A. (1995). *Study tasks in English*. Cambridge, England: Cambridge University Press.
- Weir, C. J. (1983). Identifying the language problems of overseas students in tertiary education in the United Kingdom. Unpublished doctoral thesis, University of London, London.
- Weir, C. J. (1984). The Associated Examining Board's Test in English for Academic Purposes (TEAP). In R. Williams, J. Swales, & J. Kirkman (Eds.), *Common ground: Shared interests in ESP and communication studies* (pp. 145-158). (ELT Documents No. 117). London: The British Council/Pergamon.
- Weissberg, B. (1993). The graduate seminar: Another research-process genre. *ESP Journal*, 12(1), 23-35.
- Wesche, M. B. (1987). Second language performance testing: The Ontario TEST of ESL as an example. *Language Testing*, 4(1), 28-47.
- West, G. K., & Byrd, P. (1982). Technical writing required of graduate engineering students. *Journal of Technical Writing and Communication*, 12, 1-6.

-
- West, R. (1994). Needs analysis in language teaching. *Language Teaching*, 27(1), 1-19.
- Westaway, G., Alderson, J. C., & Clapham, C. M. (1990). Directions in testing English for specific purposes. In J. H. A. L. de Jong & D. K. Stevenson (Eds.), *Individualizing the assessment of language abilities* (pp. 239-256). London, England: Multilingual Matters.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford, England: Oxford University Press.
- Williams, J. (1992). Planning, discourse marking and the comprehensibility of international teaching assistants. *TESOL Quarterly*, 26(4), 693-711.
- Williams, R., Swales, J., & Kirkman, J. (Eds.). (1984). *Common ground: Shared interests in ESP and communication studies*. (ELT Documents No. 117). London: The British Council/Pergamon.

Appendix A: Summary of North American Research Reviewed

	Author(s)	Date	Institution(s)	Subjects	N	Level	Subject area(s)	Focus	Research method
1.	Ostler	1980	University of Southern California	International students	133	u/g: 72%, p/g: 28%	hard sciences, engineering, business studies, soft sciences	4 skills/sub-skills needs; oral-aural proficiency	survey (questionnaire and written language tasks)
2.	Johns	1981	San Diego State University	Faculty	140	lower & upper divisions	cross section	4 skills needs; need for GE vs. ESP	survey (questionnaire)
3.	Christison & Krahnke	1986	5 U.S. universities	International students	80	u/g: 71%, p/g: 29%	cross section	4 skills needs and lacks	survey (structured interview)
4.	Bridgeman & Carlson	1983	34 US & Canadian universities	Faculty	n/a	mainly p/g	p/g: business management, civil engineering, electrical engineering, psychology, chemistry, computer science u/g: English	writing tasks	survey (questionnaire)
5.	Horowitz	1986a & 1986b	Western Illinois State University	Faculty	36	u/g & p/g	predominantly humanities and social sciences n/s	writing tasks	survey (analysis of documents)
6.	Canseco & Byrd	1989	Georgia State University	Faculty	n/a	p/g	business administration	writing tasks	survey (analysis of documents)
7.	Braine	1989	University of Texas (Austin)	Faculty	n/a	u/g	science and technology	writing tasks	survey (analysis of documents)

u/g = undergraduate
p/g = postgraduate
GE = General English
f/m = freshmen
n/s = not specified

ESP = English for Specific Purposes
ITA = International Teaching Assistants
NSTA = Native-speaker Teaching Assistants

	Author(s)	Date	Institution(s)	Subjects	N	Level	Subject area(s)	Focus	Research method
8.	Johns & Mayes	1990	San Diego State University	International students	80	u/g	n/s	summarising	simulated summary task--> analysis of results
9.	Casanave & Hubbard	1992	Stanford University	Faculty	85	p/g	humanities, science, and technology	writing tasks	survey (analysis of documents)
10.	Jenkins, et al.	1993	6 U.S. universities	Faculty	188	p/g	engineering	writing tasks	survey (analysis of documents)
11.	Chaudron & Richards	1986	University of Hawaii	International students	152	u/g	n/s	listening skills	lecture simulation --> analysis of post-lecture results
12.	Dunkel	1988	University of Arizona	International and indigenous students	129	u/g (f/m)	n/s	lecture note-taking	lecture simulation --> analysis of post-lecture results
13.	Olsen & Huckin	1990	U.S. university	International students	14	p/g & u/g	engineering	lecture comprehension	lecture simulation --> analysis of post-lecture results
14.	Rounds	1987	U.S. university	ITAs & NSTAs	2 & 3	p/g	mathematics	oral skills	analysis of teaching performance
15.	Micheau & Billmyer	1987	University of Pennsylvania	Faculty, international students, and indigenous students	n/s	p/g	business administration	oral skills	observation and interviews
16.	Williams	1992	U.S. university	ITAs & NSTAs	24 & 5	p/g	n/s	oral skills	analysis of simulated and real teaching performance

u/g = undergraduate
p/g = postgraduate
GE = General English
f/m = freshmen
n/s = not specified

ESP = English for Specific Purposes
ITA = International Teaching Assistants
NSTA = Native-speaker Teaching Assistants

	Author(s)	Date	Institution(s)	Subjects	N	Level	Subject area(s)	Focus	Research method
17.	Tyler	1992	U.S. university	ITA & NSTA	1 & 1	p/g	engineering & science	oral skills	analysis of simulated lecture
18.	Weissberg	1993	New Mexico State University	Faculty, international students, and indigenous students	n/s	p/g	animal science and agronomy	oral skills	observation, interview, analysis of documents
19.	Smith	1989	University of Minnesota	ITAs	38	p/g	physics, mathematics, & chemistry	oral skills	analysis of test results
20.	Douglas & Selinker	1993	Iowa State University	ITAs	15	p/g	mathematics	oral skills	analysis of test results
21.	Jacobson	1986	University of Minnesota Institute of Technology	Faculty, international students, & indigenous students	45	u/g (f/m)	engineering	communication strategies	communication strategies
22.	Benson	1991	U.S. university	International student	1	p/g	public administration	reading	reading
23.	Parry	1991	Hunter College, City University of New York	International students	4	u/g	anthropology	analysis of glosses	written vocabulary

u/g = undergraduate
p/g = postgraduate
GE = General English
f/m = freshmen
n/s = not specified

ESP = English for Specific Purposes
ITA = International Teaching Assistants
NSTA = Native-speaker Teaching Assistants

Statistical Analysis

a. by subjects/number of studies

International students (including ITAs)	Faculty	International students (including ITAs)/indigenous students (including NSTAs)	International students/indigenous students/faculty
9	7	4	3

b. by level/number of studies

Postgraduate	Undergraduate	Both undergraduate & postgraduate
11	6	6

c. by subject areas/number of studies

Science/engineering	Cross section	Business administration	Other	Unknown
9	6	2	2	4

d. by focus/number of studies

Writing tasks	Oral skills	Listening skills	4 skills	Other
7	7	3	3	3

e. by research method/number of studies

Analysis of experimental/test data	Survey of study materials	Survey of views (questionnaire/interview)	Observation/analysis of study materials/interviews
8	7	6	2



Cover Printed on Recycled Paper

58701-14166 • Y116M.75 • 253704 • Printed in U.S.A.