EAP in a Foreign Language Context: An Impact Analysis of Course Achievements on Proficiency Gain

Steven Ross
School of Policy Studies
Kwansei Gakuin, Kobe/Sanda, Japan

Abstract

English for academic purposes (EAP) curricula have gradually become widespread in a number of different foreign language contexts. Often, EAP in these contexts is conjoined with other curricular innovations such as thematic, adjunct, and sheltered content teaching approaches. The impact of such foreign language instructional programs is not particularly well known. One such program is the focus of the present study. EAP course achievement scores over a three year period are analyzed for their aggregate and individual impact on Test of English as a Foreign Language (TOEFL) score gains for two cohorts of Japanese university undergraduates. Contrastive regression analyses are used to examine the influence of the course achievement outcomes on changes in learner proficiency, as it is measured by the Test of English as a Foreign Language.

An increasingly utilized model of foreign language instruction is predicated on the idea that learning language can be facilitated through the study of academic content. The English for academic purposes approach to foreign language learning is currently most often based on teaching language and content simultaneously. This approach assumes that content instruction models (Brinton, Snow and Wesche, 1989; Mohan, 1986) entail some form of language curriculum design involving accommodation of learning content to student proficiency level. The spread of content based language instruction has also been propelled by a wide range of empirical studies which by and large have indicated that second language learners can successfully acquire language and content concurrently. Studies of the efficacy of content based second language instruction in Canada, in particular, (Edwards, Wesche, Krashen, Clement, and Krudenier 1985; Hauptman, Wesche and Ready, 1988) have indicated learner success in subject matter content learning as well as second language learning. While the initial studies of content based instruction have been mostly positive,

they were in fact conducted in a primarily bilingual milieu. Similarly, in the context of Hong Kong, where bilingual education has a well established history, Ho (1982; 1985) found corroborating evidence that content and language can be concurrently acquired when second language medium instruction is implemented.

A number of applications of content based instruction have been carried out in more recognizably foreign language contexts. Lafayette and Buscaglia (1985), Peck (1987), Sternfeld (1989), Giauque (1987) report on varying degrees of success in teaching content through the medium of a foreign language. Among foreign language programs there are, however, fundamental differences in the instructional goals and expected outcomes for the participants. In some programs, the students may aim to improve their foreign language knowledge, and might not actually be held accountable for content mastery. In other foreign language contexts, learners may be pressed to acquire language and content concurrently (Fredrickson, Hagedorn and Read 1991; Mayer and Mayer, 1991). Criteria for determining the success of these programs may be more substantively linked to objective evidence of content mastery as well as to observable changes in language proficiency, often on standardized tests (Brown, 1991; Beretta 1992; Des Brisay and Ready, 1991).

Mohan (1986; 1990) has outlined the major differences in the foci and procedural characteristics of content-based instruction models with a wide ranging array of applications in education. Oxford (1993) notes an explicit gradation in the use of the content-based language instruction curriculum, with the task-based (Long and Crookes, 1992) considered applicable to all levels, followed by the theme-based approach, which, according to current practice in second language contexts at

least, promises to yield positive results at all levels of proficiency. In second and foreign language applications of content-based instruction both task-based and thematic contest-based instructional programs have been undertaken. The general consensus appears to be that task-based and thematic instruction offer advantages over more traditional approaches to language instruction.

Less is known about the generalizability of the adjunct model to truly foreign language contexts. In the adjunct model, a content course is taught in the second or foreign language by an subject-area specialist. A content-specific syllabus is constructed in a parallel language course which focuses on the language of instruction used the content course. Readings, lectures and writing assignments from the content course may become the basis of reinforcing tasks in the parallel language course (Brinton, Snow and Wesche, 1989). The language course thus reinforces and recycles language used to teach the thematic content of the subject-matter course and provides contextual language analysis with immediate relevance to the needs of the learners.

In some instructional contexts second or even foreign language learners take a content course on their own, where the only overt support learners get is though the content instructor's accommodation to the learners' perceived proficiency. Such sheltered content courses have been found effective in promoting content learning, language proficiency gains, and increased learner motivation (Krashen, 1991). As is the case with the adjunct model, little is known about the applicability of sheltered content in truly foreign language contexts. The few examples of empirical research done on the adjunct or sheltered content curriculum in foreign language context have usually operated under the caveat that learners are of advanced proficiency and are

highly motivated (Brinton, Snow, and Wesche 1989; Oxford, 1993). How the thematic and adjunct content instruction model work in a foreign language context in which the ideal conditions of pre-existing learner proficiency and motivation do not hold, remain to a large extent unknown.

Of particular interest to the language curriculum designer and program evaluator is how achievement in an EAP program translates into generalizable language proficiency. A key reason for this interest is that many stakeholders in an EAP program (students, administrators, and content instructors) may expect outcomes that generalize beyond survival in the immediate context of instruction. In the context of EAP as foreign language education, in which instruction does not necessarily focus exclusively on the structure and usage of the foreign language, the impact of EAP course content mastery on proficiency may well be a crucial source of evidence of program effectiveness. The aim of such formative evaluation is to investigate whether proficiency develops concurrently with content knowledge.

The English language program at the School of Policy Studies at Kwansei Gakuin was designed to implement an English for academic purposes curriculum. The major rationale for the EAP program is based on the larger policy studies curriculum in which the Japanese undergraduates may enroll in a number of upper division (third and fourth years) policy studies courses taught through the medium of English.

The language training program offers five semesters of English language instruction (approximately 324 hours) to undergraduate policy studies majors. All students take the Institutional TOEFL for placement, and then retake the TOEFL at the end of the first and second years of the program. One criterion for program evaluation is based on gains students demonstrate on Institutional TOEFL over the course of their EAP instruction.

The focus of this paper is an analysis of the impact of course achievement in EAP courses on individual differences in TOEFL gain. Here, achievement testing refers to a specific sampling of course content mastery (Bejar, 1983). That is 'achievement' is of the contents of each course's syllabus, as measured with the use of composite grades based on reports, projects, peer evaluations, objective tests, and essays. TOEFL, in contrast, samples a much wider domain of English language, and is therefore conceptualized as a measure of general proficiency, but one sampling sub-domain of English for academic purposes.

The main purpose of the analysis is to explore the validity of different components of the language curriculum as contributors of language learning gain as measured by changes in TOEFL scores. The assumption underlying this analysis is that EAP courses are validated if student success (achievement of course objectives) in them translates to change in language proficiency. A desirable by-product of this type of impact analysis is an empirical basis for conducting frequent curriculum and syllabus renewal so as to optimize positive language learning outcomes.

The method of analysis used in the present study is a regression approach to impact analysis (Mohr 1992). The specific method used here involves contrastive multiple regression modeling that incorporates predictive validity criteria analogous to the Campbell and Fiske (1959) approach to construct validation. However, the approach taken here is different from the usual impact analysis, which is predicated on the quasi-experimental value-added model of program evaluation (Lynch, 1996). Rather, it involves two hypotheses. The first is that same trait achievement in language courses should lead to same trait gain variance

in proficiency if the EAP program can be inferred to have the intended impact on learner proficiency. In contrast, different trait achievement in language courses should lead to comparatively smaller gain in proficiency. It is in this manner of predicting validity coefficients of different magnitudes that there is a similarity to the Campbell and Fiske approach to discriminant validity. The second is that specific same-trait achievement in the EAP courses will lead to unique contributions to variance in TOEFL gain.

Regression Models

The regression model used for the impact analysis is based on a skill-focus categorization of achievement scores from courses in each of six semesters of the language program. Each course's achievement score was categorized as leading to either gain in aural comprehension or to gain in reading or language structure proficiency. Each course achievement score was thus hypothesized at the outset to provide a unique contribution to gain in proficiency. Since students in the program differ widely in proficiency (TOEFL ranges of 320 to 640), individual differences in language proficiency are included in the regression model, in the form of pre-instruction proficiency measures. These pre-instructional scores

function as covariates (Huitema, 1980; Cohen and Cohen, 1983) in that they serve to account for variance in the post-instruction TOEFL scores that could be artefactual, since individual difference always exist before instruction in any given semester. The remaining predictors of proficiency gain are the scores from the EAP language courses, which are categorized a priori as mainly providing input to aural skill development or to literacy/grammar development.

In each academic year of the School of Policy Studies English language program, all language course achievement scores were labeled as being members of the 'R' Set (literacy/grammar) or as members of the 'L' Set (aural comprehension), which correspond to the proficiency traits (skills) that the courses were designed to develop.

The main hypothesis tested in the impact analysis was that the two respective sets of course achievement scores would provide multiple correlations of different magnitudes with TOEFL post-test scores at the end of the instructional year, and that the magnitudes of the multiple correlations, which all include the pre-test scores, would match the same trait order predicted. Table 1 outlines the expected outcomes of the contrastive regression models, which include pre-test TOEFL scores.

Table 1

Outcome		R^2	R^2
TOEFLΔLC	=	L Set >	R Set
TOEFLΔST	=	R Set >	L Set
TOEFLΔRC	=	R Set >	L Set

A secondary goal of the impact analysis is to identify course individual achievement scores that do not fit into to expected prediction order, or that do not contribute any unique variance to gain in proficiency.

EAP Course Achievement

Course Achievement score data reflect on-going curriculum development at the School of Policy Studies English language program. While most courses fit neatly into either the R Set or L Set classification scheme, a few potentially offer input to both aural and literacy/ grammar proficiency development. Such courses are 'wild cards' in the current classification scheme, and are therefore included in both the R and L Sets, as are the pre-tests in each proficiency measured by TOEFL.

Impact Study 1

The first impact study is focused on the first nine courses taught in the EAP program1 in the 1995 academic year. The course achievement scores included in the analysis are listed in Table 2.

Table 2

	Code	Course	Domain	Test Method
	L1S	Listening	Ecology	Multiple Choice
	N1S	Note Taking	Thematic	Multiple Choice
L Set	B2F	Study Skills	Ecology	Mixed
	L2F	Listening	Ecology	Multiple Choice
	Eco	Human Ecology	Ecology	Multiple Choice
	R1S	Reading	Thematic	Mixed
	W1S	Writing	Thematic	Multidraft Essay
R Set	W2F	Writing	Thematic	Multidraft Essay
	R2F	Reading	Thematic	Mixed
	Eco	Human Ecology	Ecology	Multiple Choice

As is outlined in Table 1, the hypothesized impact for the L Set is on changes in TOEFL listening subtest variance. In order to test this hypothesis, the magnitudes of observed contrastive multiple correlations (R^2, or 'coefficients of determination' (Cohen and Cohen, 1983) are compared with the order hypothesized. For year 1 of the 1995 cohort of students, Table 3 lists the outcomes.

Table 3

Outcome(N=327)		L Set R^2	R Set R^2	Hypothesis
TOEFLΔLC	=	.461 >	.415	Supported
TOEFLΔST	=	.288 >	.302	Supported
TOEFLΔRC	=	.327 >	.301	Not Supported

The results of the 1995 year 1 achievement scores suggest that changes in TOEFL listening scores are most validly related to achievement in courses providing mainly input in listening (the L Set). The magnitude of differences between the competing coefficients

of determination diminishes, however, when changes in TOEFL structure and reading are compared.

A close look at the TOEFL gain L Set is in order. It can be readily observed that the individual courses in the L Set all contribute unique variance to listening gain (Table 4), with the exception of the 'wild card' Ecology content course.

Table 4 TOEFL Listening Gain 1995 Year 1

L Set			R Set		
VARIABLE	**Beta**	P	VARIABLE	**Beta**	P
PRETEST	0.501	0.000	PRETEST	0.538	0.000
L1S	0.092	0.057	R1S	0.154	0.001
N1S	0.160	0.001	W1S	-0.011	0.813
B2F	0.142	0.001	W2F	0.108	0.025
L2F	0.127	0.009	R2F	-0.063	0.175
ECO	-0.030	0.551	ECO	0.085	0.080

The results of the contrastive regression model for the first year suggest that the EAP curriculum has some degree of impact validity on the development of listening proficiency. There is less clear evidence for the other two skills measured by TOEFL -- structure and reading comprehension. It can be seen (Table 2) that there is very little difference between the L Set of predictors and R Set (.288 vs. .302) in accounting for the observed differences on the TOEFL Structure subtest. The impact of the R Set on the TOEFL Reading subtest is, contrary to the first hypothesis, not as large as the L Set. This finding suggests a lack of validity in the construction of the R Set curriculum in the first year, a lack of coherent reading achievement testing, or both.

Impact Study 2

The second year of the program was evaluated with the same type of regression model. This time, the post-test from the first year served as the pre-test for the second year of the curriculum. In the second year, a different focus was offered with less emphasis on the ecology theme, and more variety in the form of thematic content courses offered. The impact of those courses is the main object of interest in the second year analysis of the 1995 cohort of students. The 'wild card' course in the second year analysis is the thematic 'mini-lecture' course, which varied considerably in content and focus across a number of class sections. For this reason, it could not be clearly classified as having any particular input focus.

Table 5

	Code	Course	Domain	Test Method
	L3S	Listening	Thematic	Mixed
L Set	N3S	Note Taking	Thematic	Multiple Choice
	M4F	Multifocus	Thematic	Mixed
	R3S	Reading	Thematic	Mixed
	W3S	Writing	Thematic	Multidraft Essay
R Set	S3F	Discussion	Thematic	Mixed
	W4F	Writing	Thematic	Multidraft Essay
	M4F	Multifocus	Thematic	Mixed

The results of the second year impact analysis are listed in Table 6. Once again, the TOEFL pre-test scores are also entered into the regression model. As in the first year impact study, the achievement scores from the L Set lead to a higher R2 than the 'rival' R Set of different skill predictors. The hypothesiz-ed order is also observed for gain on the TOEFL skills of structure and reading.

Table 6

Outcome(N=258)		L Set R^2		R Set R^2	Hypothesis
TOEFLΔLC	=	.459	>	.433	Supported
TOEFLΔST	=	.284	>	.298	Supported
TOEFLΔRC	=	.343	>	.376	Supported

Of special interest here is the change in TOEFL Reading gain. Table 3 lists the larger R^2 for the L Set relative to the R Set in the first year. In the second year, the order was reversed, as was the original hypothesis. We can look to the content of the R Set in the second year curriculum in order to see what individual achievement scores had a unique impact on the TOEFL gain. Table 7 shows the contribution of individual achievement scores (standardized regression weights, or 'Betas') on the regression model.

Table 7 TOEFL Reading Gain 1995 Year 2

L Set			R Set		
VARIABLE	**Beta**	P	VARIABLE	**Beta**	P
Pretest	0.518	0.000	Pretest	0.484	0.000
L3S	0.003	0.956	W3S	0.114	0.054
N3S	0.079	0.176	R3S	0.086	0.207
M4F	0.182	0.002	S4F	0.159	0.023
			W4F	0.102	0.084
			M4F	0.084	0.228

As might be expected, those achievement scores with the largest input focus on reading and literacy lead to the largest impact in reading proficiency gain. That is, the writing and seminar (topical reading and discussion) course components in the R Set impact on the gain most. It is also worthy to note that there are more predictors in the R Set than the L Set. This can lead to bias when the usual R2 is used as the basis for comparison. In the present study, the bias is controlled for with the interpretation of the adjusted R2 as the basis for comparing rival hypotheses. This criterion is used in all analyses.

One course component was entered in both sets. The thematic 'mini-courses' were neither clearly categorizable as mainly listening input or literacy/structural input dominant, and were therefore tested for their major impact. As can be seen in Table 7, the standardized regression weight (beta) is largest for the 'mini-course' achievement scores when the outcome is the TOEFL Listening post-test.

Impact Study 3

The final analysis in the study was conducted on the language achievement records of the second cohort of students. The 1996 cohort underwent a curriculum very similar to that of the 1995 cohort, although there appears to be some degree of drift away from the 1995 core theme of ecology more towards a thematic smorgasbord. The main overt difference was the replacement of the academic reading component with the seminar (reading and discussion) component in the first and second terms of the first year. Table 8 outlines the courses and shift toward the thematic focus.

As in the foregoing analyses, the hypothesized relationships were tested using the pre-test TOEFL and course achievement grades as predictors of gain in proficiency. The model differs from that of 1995 Year 1 in that the Ecology course grade is not modeled. This provides symmetry between the two sets. Table 8 lists the components of the L Set and R Set for the 1996 Year 1 analysis.

Table 8

	Code	Course	Domain	Test Method
	L1S	Listening	Thematic	Mixed
	N1S	Note Taking	Thematic	Multiple Choice
L Set	B2F	Lecture Review	Ecology	Multiple Choice
	L2F	Listening	Ecology	Mixed
	R1S	Reading	Thematic	Mixed
	W1S	Writing	Thematic	Multidraft Essay
R Set	W2F	Writing	Thematic	Multidraft Essay
	S2F	Reading	Thematic	Mixed

The impact of the achievement scores on TOEFL differs from that observed in the 1995 Year 1 analysis. It can been observed in Table 9 that L Set's impact on TOEFL gain is as hypothesized. The validity of the R Set, however, varies. As in the previous two analyses, the impact of the R Set on TOEFL structure relative the impact of the L Set, is negligible. It appears that the overall impact of the EAP curriculum and achievement testing has a consistent relationship with listening gain only.

Table 9

Outcome(N=253)		L Set R^2	R Set R^2	Hypothesis
TOEFLΔLC	=	.583 >	.536	Supported
TOEFLΔST	=	.361 >	.342	Not Supported
TOEFLΔRC	=	.449 >	.352	Not Supported

As in the analysis of the 1995 Year 1 cohort, we will examine the achievement scores that have a significant impact on gain in listening proficiency. Once again, the major achievement impacting on gain is that accruing from the listening/note taking course, in which students are taught strategies for comprehending authentic academic lectures. This course and the achievement scores associated with it appear to have the most salient impact in the first year of the EAP program. This finding may reveal Japanese students' general lack of sophistication in listening, which is most readily developed with sustained, focused input as that provided in the listening/note taking courses.

Table 10   TOEFL Listening Gain 1996 Year 1

L Set			R Set		
VARIABLE	**Beta**	P	VARIABLE	**Beta**	P
Pretest	0.621	0.000	Pretest	0.687	0.000
L1S	0.077	0.132	W1S	-0.002	0.962
N1S	0.231	0.000	R1S	0.154	0.003
B2F	0.021	0.693	S2F	0.060	0.267
L2F	-0.006	0.914	W2F	-0.035	0.522

It is worthy to note that the impact of the dedicated listening achievement courses' scores is surprisingly small. Only one of five semesters' listening achievement scores had a significant beta weight on the TOEFL Listening post-test regression. This phenomenon may indicate some degree of syllabus 'drift' away from the principle skill area the course was designed to address, into some other nexus of sub skills that do not show up as listening gains. Consistent use of materials or language teaching technology (language/video lab facilities) may also be the objects of closer scrutiny.

The foregoing analyses have been of the three dimensions of language proficiency measured by the TOEFL test. There is another important area of language proficiency not directly assessed by TOEFL that is systematically featured in the EAP curriculum. This crucial skill is academic writing. The writing course achievements, as members of the R Sets, typically have middling-sized beta weights on TOEFL Structure and Reading post-

tests. These regression weights, however, can be considered to underdetermine the impact of the writing achievements, chiefly because there is no direct assessment of academic writing currently used for SPS program evaluation. Future evaluations will need to redress the current under-sampling of EAP proficiency outcomes such as writing if this aspect of the curriculum is to be accurately evaluated. Future impact studies may need to include pre-and-post tests of the Test of Written English in addition to the TOEFL.

The impact analyses presented here give a general picture of the interface between EAP course achievement and proficiency gain. Variation in the impact of different course achievements from term to term can serve as a criterion for determining which courses need to be considered for immediate rather than delayed revision and restructuring. The fact that the hypothesized bundles of courses as skill 'sets' did not consistency have an impact on proficiency gain indicate the need for serious rethinking of how current EAP courses can be improved to lead to greater proficiency gain.

It is also important to consider that for the purposes of thorough program evaluation, more information is necessary in order to determine the extent of other types of desirable outcomes of the current program (Jackson, 1994). For this, surveys and interviews, or other qualitative methods of program evaluation may prove useful (Lynch, 1996).

Notes

1)

The introduction to human ecology course can be considered part of the EAP curriculum in the first year of the program. Since this content course is taught in English, its impact on proficiency development is potentially large. For this reason it was included in the regression model for 1995 Year 1.

References

Bejar, I. (1983) Achievement Testing. Sage university paper on quantitative methods in the social sciences. Thousand Oaks, CA: Sage.

Beretta, A. (1992). Evaluation of language education: an overview. In J.C. Alderson and A. Beretta (Eds.) Evaluating second language education. Cambridge: Cambridge University Press.

Brinton D. M. , Snow, M.A. & Wesche, M.B. (1989). Content-based second language instruction. New York: Newbury House.

Brown, J.D. (1991). Content-based language testing. Cross Currents 18, 174-178.

Campbell, D. and Fiske, D. (1959). Convergent and discriminant validity by the multitrait, multimethod matrix. Psychological Bulletin 56. 81-105.

Cohen, J. and Cohen, P. (1983) Multiple Correlation/Regression Analysis for the Behavioral Sciences. NJ: Lawrence Erlbam.

Cummings, J. (1982). Tests, achievement, and bicultural ambivalence. Rosslyn, VA: Clearinghouse on Bilingual Education.

Cummings, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. Clevedon: Multilingual Matters.

Des Brisay, M. & Ready, D. (1991). Defining an appropriate role for language tests in intensive language programs. In S. Anivan (Ed.) Issues in language program evaluation. Singapore: RELC 15-25.

Edwards, H., Wesche, M., Krashen, S., Clement, R. and Kruidenier, B. (1984). Second language acquisition through subject-matter learning: A study of sheltered psychology classes at the University of Ottawa. The Canadian Modern Language Review 41, 268-282.

Enochs, K. (1994). Content-based language teaching: A survey of the literature. Language Research Bulletin 9, 7-18.

Fredrickson, T.L., Hagedorn, J.J. & Reed, H. (1991). The content-based curriculum at ITM/MUCIA. Cross Currents 18, 200-205.

Giauque, G. (1987). Teaching for content in a skills course: Greek mythology in French. Foreign Language Annals 20, 565-569.

Haputman, P.C., Wesche, M. B. & Ready, P. (1988). Second language acquisition through subject matter learning: A follow-up study at the University of Ottawa. Language Learning 38, 439-482.

Ho, K. (1982). Teaching physics through English. Language Learning and Communication, 1, 283-288.

Ho, K. (1985). The paradox of immersion in a second language. NABE Journal, 10, 51-64.

Huitema, B. (1980) The Analysis of Covariance and Alternatives. NY: John Wiley and Sons.

Jackson, E. (1994) Non-Language Outcomes in the Adult Migrant English Program. Research Report Series 5. Sydney: National Centre for English Language Teaching and Research.

Krashen, S. (1991). Sheltered subject matter teaching. Cross Currents 18, 183-189.

Lafeyette R. & Buscaglia, M. (1985). Students learn language via a civilization course - A comparison of second language classroom environments. Studies in Second Language Acquisition 7, 323-342.

Long, M. H. & Crookes, G. (1992). Three approaches to task-based syllabus design. TESOL Quarterly 26, 27-56.

Lynch. B. (1996) Language Program Evaluation. London: Cambridge University Press.

Mayer, L. & Mayer, M. (1991). Four years later -- Are we content with content? Cross Currents 18, 206-209.

Mohr, L. (1992). Impact Analysis for Program Evaluation. Thousand Oaks, CA: Sage.

Mohan, B. (1986). Language and Content. Reading, MA: Addison-Wesley.

Mohan, B. (1990). LEP students and the integration of language and content: Knowledge structures and tasks. In C. Simich-Dudgeon (Ed.) Procedings of the First Research Symposium on Limited English Proficient Students' Issues. Washington DC: U.S. Department of Education.

Oxford, R. (1993). Progress in tertiary content-based ESL instruction. TESL Canada Journal 11, 75-97.

Peck, B. (1987). Spanish for social workers: An intermediate-level communicative course with content lectures. Modern Language Journal 71,402-409.

Sternfeld, S. (1989). The University of Utah's immersion multiliteracy program: An example of an area studies approach to the design of first-year college foreign language instruction. Foreign Language Annals 22, 341-353.