DOCUMENT RESUME

ED 386 939 FL 023 252

AUTHOR Wilhelm, Kim Hughes
TITLE Identification of LL (Language Learning) Background Variables Associated with University IEP Success.
PUB DATE Mar 95
NOTE 26p.; Paper presented at the Annual Conference of the American Association for Applied Linguistics (Long Beach, CA, March 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; College Second Language Programs; College Students; *English (Second Language); Higher Education; *Intensive Language Courses; *Language Aptitude; *Language Proficiency; *Language Role; Learning Motivation; Second Language Instruction; Student Attitudes; Student Characteristics; Student Placement
IDENTIFIERS Indiana University

ABSTRACT

A study examined variability in learner progress through a pre-university intensive English-as-a-Second-Language program as it relates to 36 language learning background variables (e.g., formal English training and exposure, motivation, etc.) and four entry-level proficiency variables. Subjects were 201 Indiana University students. Success was measured by passage from one proficiency level to another/others over the period of program enrollment, and three groups were identified: high, medium, and low success learners. Analysis of the data resulted in identification of nine language learning background variables most closely associated with success patterns. These include: school levels at which English was learned; number of family members who have studied in English-speaking countries; previous exposure to English as the language of instruction; number of years since English was last studied; number of years of university English; teaching focus in English class; first language; attitudes toward English class; and stated purpose for learning English. Results are summarized. The research method applied in this study was found useful for identifying significant variables associated with language learning. Contains 12 references. (MSE)

Identification of LL (Language Learning) Background Variables Associated with University IEP Success

Kim Hughes Wilhelm

Identification of LL (Language Learning) Background Variables Associated with University IEP Success

Introduction

The investigation of variables associated with success in second language learning has long been an area of interest in SLA research, but most studies have been limited to an examination of a small number of variables. This exploratory study utilized an expert system[1] in order to examine a large number of LL variables, with the purpose of identifying those variables most strongly associated with LL success in a university intensive English program.

Chapelle and Green (1992) contrast two types of SLA research, namely "good language learner" (GLL) research which "examines many aspects of the acquisition process, including learner characteristics which influence acquisition (p. 59)" with "language testing research" which describes language proficiency at particular stages of development. Brown (1994) proposes a comprehensive theory of second language acquisition, one which includes "as many relevant factors as possible (p. 278)." Crookes and Schmidt (1991) suggest that a broad taxonomy be developed to answer the question "What factors are relevant to SL motivation?" and note the interrelationship of relevant factors (p. 497-8). Skehan (1991) suggests a greater emphasis on the differential approach (as compared to the experimental approach which has predominated in SLA research), stating the differential approach is more likely to "identify attributes on which people differ and then [suggests relating] such attributes to different performances in, for example, learning (p.

[1] Expert systems are a subset of knowledge base systems, defined by Patrick and Fattu (1986) as "consultation systems, classifications, decision-making systems, expert systems, inference systems, diagnostic systems, hypothesis generating systems..."

275)."

Despite acknowledgements that second language learning models must take into account the interactive nature of a variety of complex variables (e.g. Yorio, 1976; Bialystok, 1978; Spolsky, 1988; Long, 1990; Brown, 1991), researchers typically have been limited by norm-based (Gaussian) statistics to an examination of small numbers of independent variables. However, more recently statistical techniques have become available for examination of a greater number of variables which allow for the possibility that some variables may be inter-related (e.g. path analysis[2] knowledge base systems, expert systems).

This study sought to examine variability in learner progress through a pre-university intensive English program as it relates to 36 language learning (LL) background variables and 4 entry level proficiency variables. A knowledge base was developed consisting of background, entry level proficiency, and program progress data for 201 adult learners. The Outcome Advisor (R) expert system (OA) was used in classification mode to assist in the identification of potentially influential variables for taxonomy refinement. The researcher then used various combinations of selected features to determine the extent to which individual learner success could be accurately predicted, with prediction based only on refined sets of LL variables.

Procedures

The sample population consisted of 242 learners enrolled in the Center for English Language Training (CELT) intensive English program

---

[2] An example of path analysis is LISREL (Analysis of Linear Structural Relationships), used by Gradman and Hanania (1991).

at Indiana University from September of 1988 through August of 1991. Hour-long personal interviews were conducted by researchers Gradman and Hanania, who gathered 75 items of information pertaining to the language learning backgrounds of the subjects in order to identify factors associated with entry proficiency as measured by entry TOEFL scores (see Gradman and Hanania, 1989; 1990; 1991). Data from the Gradman and Hanania interviews provided information for 36 of the features included in the study, with the remaining data collected from student records. Variables examined are listed in Appendix A, according to the broad categories of a) formal learning of English, b) classroom exposure to and use of English, c) extracurricular exposure and use, d) attitudes and motivations, e) entry proficiency and placement, and f) general descriptive information. Sample coding of feature values is provided, as well.

Of the 242 learners interviewed, 41 were excluded from the study because they were only enrolled for one seven-week session; consequently, a total of 201 subjects made up the knowledge base. The sample population was heterogeneous, including pre-university, undergraduate, and graduate level learners from a wide variety of academic fields. There were 120 males and 81 females, ranging from 16 to 45 years of age. Fifteen first languages were represented. English entry level proficiency as measured by entry TOEFL scores ranged from 293 to 580. Appendix B provides general population characteristics and percentages.

Establishment of Success Criteria

During the seven-week CELT sessions, learners participated in twenty-five hours a week of skill-based classes (reading, writing, listening and speaking, and grammar) across seven placement proficiency levels (level 1 = least proficient; level 7 = most

proficient). At the end of each session, learners were either unconditionally advanced to a higher level, probationarily advanced, or retained at the current level on the basis of a) teacher recommendations and course grades, b) proficiency scores as measured by the CELT placement exam, and c) proficiency scores as measured by the TOEFL (Test of English as a Second Language). The final recommendation for unconditional pass, conditional pass, or retention is made by the IEP Director and/or Co-Director, who meet with teachers to discuss each learner and who review placement and TOEFL scores before making their decisions. Learner advancement or retention from level to level formed the basis of the learner success measurement used in the study. For each session enrolled, the subject was assigned a progress score tabulated according to the formula provided in Appendix C, "End of Session Progress Scores." End of session progress scores were summed, then divided by the number of sessions the subject was enrolled in the program. Appendix B also provides examples of Overall Average Success Computation.

Subjects designated "high" or "medium" success, therefore, were consistently passed from level to level, while "low" success learners were retained more than once. Out of the total number of 201 subjects, 28% (n=57) were designated low success, 33% (n=67) were designated medium success, and 38% (n=77) were designated high success.

The IEP director, co-director, and a full-time instructor were also asked to judge the success (high, medium, or low) of fifty subjects, basing judgements on personal knowledge of the learners. There was 100% agreement using only two success categories (high/medium versus low) and 97.6% agreement using three success categories (high, medium, low), with the director designating two

subjects "medium" success while the researcher had designated them "high" success. Appendix D provides a summary of the expert agreement success ratings.

Data were coded so that each subject was represented in the knowledge base as a string of record vectors. Appendix E provides a "translated" version of one subject's record vector. Please take a minute to review this subject's data and to make your judgement as to whether this learner is likely to be a high, medium, or low success learner.

You may have guessed, correctly, that the learner was "low" sucess. To complete this problem, you relied upon your expert judgement and the knowledge you have about second language learners and learning. The OA knowledge base system is able to perform a similar function by "learning" from the stored data provided in the knowledge bank in order to perform sophisticated classification and inference tasks.

The OA knowledge bank was generated as data became available upon learner completion of the program, with some learners finishing within 2 sessions while others were enrolled up to 8 sessions. This resulted in seven separate data inputs, with the knowledge bank growing from 52 to 79, then 107, 135, 158, 177, and finally 201 cases. After each data input, the OA was used in classification mode to provide category conditional probability densities for each feature. To better analyze the results, histograms were constructed for each probability density (for each of the three success categories) in order to identify which features "separated" between the high, medium, or low success groups. A "perfect" feature would be one with no overlap between success categories, one which shows disparate patterns for the success outcomes (i.e. maximum interclass

distance). For example, if a very large percentage of males were low success while a large percentage of females were high success, then disparate patterns for the values "male" and "female" would make "gender" a good feature.

Features which showed maximum interclass distance across at least two feature values were selected from the taxonomy after the third data run and again after the sixth data run, with thirteen LL background features and two entry proficiency features emerging as those which most clearly differentiated between the three success categories across 5/6 or 6/6 data runs. These "strongest" features identified from probability density histograms are shown on page 1 of Appendix F.

In addition, statistical analysis of individual feature performance was conducted, providing the relative frequency for each feature value by success category. All 40 features were examined for individual performance and the researcher identified those feature values with .66 or higher likelihood of association with a success category. Individual feature performance analysis was conducted using both three categories of success (high, medium, low) and two categories of success (high/medium, low) upon the recommendation of the OA developers. It was important to examine these results carefully, however, and to identify the number of cases represented within each feature value (e.g. only one subject reported no contact hours of English at the school and the university, but individual feature performance indicated that feature value being strongly associated with low success).

When using three categories of success, LL background feature values identified as being strongly associated with success categories were those with only 1 case, so were not helpful in

identifying "best" features. Other feature values showing high frequency of association with success categories were entry proficiency variables rather than background variables.

When using two categories of success, individual feature performance data was useful as it further confirmed the importance of variables already selected by the researcher and identified features with only one value strongly associated with the high/medium or low success category. Three additional LL background features were identified (contact hours of English at school & university, index of oral exposure in class, and index of oral use in class). Those data are provided on page 2 of Appendix F and can be compared to the "strong" features identified across the six data runs. Individual feature performance for entry proficiency variables is provided in Appendix F, as well. Take a minute to examine the feature values and to list your "best guess" of the associated success category. You were correct if you guessed that those with a "1" rating were low success and those with a "2" were high/medium success.

A further discussion of the nine "strongest" LL background variables follows, using data from probability density histograms resulting from the sixth data run, at N=177, with three categories of success.

School levels at which English was learned

For all learners in the sample, it was most common that English was learned upon entry into secondary school (54% of the low success group, 78% of the medium success group, and 60% of the high success group). A number of both the low (38%) and the high success learners (22%) also were taught English in upper primary school. The medium success group had the highest number who reported learning English upon entry into primary school (15%), followed by 12% of the high

success learners. Only 4% of the low success learners began learning English at such an early age.

Number of family members who've studied in English-speaking countries

About half of each success group had no family member who had studied in English-speaking (ES) countries (54% of low, 42% of medium, and 51% of high). Over a third (35%) of the medium success group indicated two to three family members and 31% of the high success group indicated one family member having studied in ES countries. It would be interesting to investigate which family members (e.g. father, mother, both parents, sibling, etc.) studied in ES countries as a clue to possible home environmental influences.

Exposure to English as the language of instruction

The majority of each success group had no English as the language of instruction (80% of low, 67% of medium, and 55% of high). However, there was a consistent trend across data runs for the more successful learners to have some ELI. Opposite trends for low versus high success learners were quite obvious, with a third (33%) of the medium success and nearly half (45%) of the high success learners having some exposure to English as the language of instruction.

Number of years since last studied English

The majority of learns from all success groups had studied English within the year prior to their entry into the intensive English program. Low success learners were most likely (72%) to have studied English within twelve months before entry. On the other hand, 18% of the low success cases had not studied it for three or more years. It was interesting to note that medium success cases

made up the largest percentage (35%) of those who had not studied English for one to two years, whereas high success cases made up the largest percentage (21%) of those who had not studied for three or more years.

Number of years of university English

The majority (60%) of low success learners had no years of university English, with an addition 24% having only one year. Comparatively, nearly a third (30%) of the medium success learners had two years of university English, and an additional 17% had three or more years. High success learners were more evenly distributed, although the majority (54%) had at least one year of university English.

When comparing number of years since last studied English, age upon entry into the IEP, academic status, and number of years of university English, a general trend appeared. Low success learners were more likely to be younger than 22 years old, to be pre-undergraduates with no university English, and to have studied English quite recently. In comparison, medium success learners were more likely to be within the 22-30 age bracket, to already hold Bachelors degrees, to have two or more years of university English, and to have not studied English for a year or two. The highest percentage of learners over 30 years of age fell into the high success group, with that group more likely to have earned Masters degrees and least likely to have studied English within the last three years. In addition, the low success learners tended to have lower entry proficiency in English, despite their exposure to English study more recently.

It is possible that the more successful learners may have more

highly developed cognitive academic skills, contributing to their success in a language program geared for university admittance. It would be interesting to examine the funding sources for the groups of learners to see if medium and high success learners are more likely to be sponsored students. If so, it may be that they are more academically gifted (and perhaps more motivated) when compared to the low success groups. Older learners also may be more mature and more adapted to living independently, helping them to cope more effectively with a new culture and with the pressures of an academic program.

Teaching focus in English class

Low success learners were most likely (40%) to report an "all skills" teaching focus, medium success learners most likely (42%) to report "grammar, translation" as the teaching focus, and high success learners most likely to report "grammar and reading" as the teaching focus. As already mentioned, strong grammar and reading scores on the subsections of both the TOEFL and the placement tests were also associated with success. There was also some indication that oral and written exposure to English were associated with learner success, but these skills were not measured by the TOEFL and placement tests.

First Language

On all six data runs, at least half of the low success learners were Arabic speakers; an additional their (34%) were native speakers of Korean or Japanese. Medium success learners were more evenly distributed among the first language groups, with the majority (38%) being speakers of Korean or Japanese and a fourth (25%) speakers of Arabic. An additional 22% spoke a Romance language or German. In contrast, nearly half (43%0 of the high success learners were native

speakers of a Romance language or German. An additional 28% were Korean or Japanese speakers, and only 13% were Arabic speakers.

Liking of English class

This feature showed opposite trends for low as compared to medium and high success learners across the two feature values. The majority of low success learners (60%) reported not liking English or else starting out liking it but eventually disliking it. In contrast, the majority of high success learners (57%) reported an opposite trend (liking it or else starting out disliking it and eventually liking it). Medium success learners were rather evenly split between those who disliked English class (52%) and those who likes it (48%).

Stated Purpose/Reason to Learn English

The majority from each success group reported learning English for academic reasons only (56% of low, 58% of medium, and 61% of high). Low success and medium success learners were more likely, however, to feel also that they were learning English because it was "important in the world" (36% of low and 27% of medium success learners), as compared to high success learners being more likely to give work and travel reasons in addition to study reasons for learning English (28%).

Limitations

The data used to construct this knowledge bank was limited by reliance on existing data. Subsequently, the features and values used were also limited to those under investigation in the Gradman and Hanania studies. There were a number of variables which the

researcher would have liked to include but could not.

Another problem was in the clear differentiation of success categories. Initial investigation used three categories of success (high, medium, low), but the definition of high versus medium success, in particular, seemed "fuzzy," especially since both sets of learners did show continuous, consistent progress from level to level. It would help if success measurements had greater "gaps" between the success categories or if "borderline" cases were moved into a separate success category or were omitted from the study. There was also a tendency for learners who entered at the low level to be less successful, while those who entered at upper levels to be more successful, a function of entry proficiency and the use of TOEFL scores and TOEFL gains when designating entry placement and end- of-session level progress.

Whenever a large number of variables are under investigation, especially since many have multiple values, a large number of cases must be collected in order for results to be considered with confidence. The developers indicate that data analysis will begin to show consistent outcome predictions with a minimum of 150 cases in the knowledge bank, but with a problem that is this complex, many more cases need to be collected and added to the knowledge bank before outcomes are reliable and can be reported with confidence. In addition, for results to be valid, the features and values must be carefully defined, data must be carefully collected, and results used with caution. Much refinement of the knowledge bank is required and there is the danger of feature values being either too broad or too general. The developers of the expert system (Patrick and Fattu) emphasize that it be used in combination with expert judgement (OA Manual, A-3). They stress the interactive nature of the OA as the

user converts the problem (questions and possible answers) into features with feature values then classifies data and deduces outcomes. Researchers must examine the data, determine what makes logical sense (given their expertise in the field), and try out combinations of features and feature values which are most likely to provide outcomes which can be used with confidence. Some may question this combination of expert judgement and statistical analysis in pursuing research into SLA, but I would contend that the value of knowledge base systems are as tools for analysis of problems which require human problem-solving and interpretation, those which employ complex, interdependent data.

Conclusions

This study hypothesized that there were elements in learners' language learning backgrounds which were conducive to learner progress in an academically-oriented intensive English program. The purpose of the study was to develop a limited set, or taxonomy, of those LL variables most clearly associated with variable learner success in the program, thereby contributing to our theoretical knowledge of variables associated with successful second language acquisition.

On a practical, applied level, investigations of this type can help to identify the characteristics of learners best served by particular programs. For example, learner proficiency information from this study indicated that students who have already demonstrated medium or better language proficiency (i.e. those with an entry TOEFL above 400 and who entered at level 3 or higher) tended to progress consistently through the program. It may be that the needs of less proficient learners may be better served by different program or by

courses which focus less on learning English for academic purposes. Results of this nature could provide valuable information for program assessment and for an evaluation of the "match" between learner needs and instructional objectives.

The taxonomy development process itself could be used to compare second and foreign language learning within different contexts. The IEP population was very heterogeneous in terms of first language, cultural backgrounds, age, etc. It would be interesting to conduct a similar enquiry into LL background variables with more homogeneous subjects to see what patterns emerge. It would also be interesting to note how the profile of the successful learner changes when the language context changes from an academic orientation to a more instrumental, vocational orientation (for example, with newcomer immigrant populations).

Despite limitations, this study was useful as an application of new statistical technology in an exploratory investigation of variables associated with SLA acquisition. Given further investigation and refinement of procedures and features/feature values, studies of this type will further advance our knowledge of the "good" language learner.

References

Bialystok E. (1978). A theoretical model of second language learning. Language Learning 28, 69-83.

Brown, H.D. (1991). Breaking the Language Barrier. Yarmouth, ME: Intercultural Press.

Brown, H.D. (1994). Principles of Language Learning and Teaching, 3rd edition. Englewood Cliffs, N.J.: Prentice-Hall Regents.

Chapelle, C. & Green, P. (1992). Field independence/dependence in second language acquisition research. Language Learning 42:1, 47-83.

Crookes, G. and Schmidt, R.W. (1991). Motivation: Reopening the research agenda. Language Learning 41:4, 469-512.

Gradman, H. and Hanania, E. (1989). Language learning background and ESL performance: Factors promoting success. In V. Bickley (Ed.), Language teaching and learning styles and learning styles within and across cultures. Hong Kong Government Education Department.

Gradman, H. and Hanania, E. (1990). Learner profiles, reading, and second language proficiency. Paper presented at the 9th World Congress of Applied Linguistics, Thessaloniki, Greece.

Gradman, H. and Hanania, E. (1991). Language learning background factors and ESL proficiency. The Modern Language Journal, 75(i), 39-51.

Long, M. H. (1990). The least a second language acquisition theory needs to explain. TESOL Quarterly 24, 649-666.

Skehan, P. (1991). Individual differences in second language learning. Studies in second language acquisition: Series on issues in second language research, 13:2, 275-298. Cambridge University Press.

Spolsky, B. (1988). Bridging the gap: A general theory of second language learning. TESOL Quarterly:22, 377-396.

Yorio, C. (1976). Discussion of "Explaining sequences and variation in second language acquisition." Language Learning:4, 59-63.

Appendix A: Taxonomy Categories and Features Examined

GENERAL INFORMATION
Age when Entered the Intensive English Program
Gender
First Language
Academic Status (e.g. pre-undergraduate, during undergraduate)
Field of Study in University

FORMAL LEARNING OF ENGLISH
School Levels at which learned English (school levels E)
English as the Language of Instruction (ELI)
Years Since Last Studied English (Yrs. last E)
Average Number of Students per Class in School
Private Schooling (private S)
Number of Years of Universitiy English
Months of Intensive English
Contact Hours with English (school and university)
Contact Hours with English (Intensive English Program)
Contact Hours with English (Private Instruction)
Total Contact Hours with English

CLASSROOM EXPOSURE AND USE
Number of Native English-Speaking (NS) Teachers
Index of Oral Exposure in Class
Index of Oral Use in Class
Index of Written Use in Class
Index of Communicative Use in Class
Audio-Visual Index
Major Teaching Focus

EXTRACURRICULAR EXPOSURE AND USE
Travel to English-Speaking (ES) Countries
Opportunity to Use English Outside of the Classroom
Index of Extracurricular Reading
Index of Extracurricular Writing
Index of Extracurricular Speaking
Index of Extracurricular Listening

ATTITUDES AND MOTIVATIONS
Family Encouragement to Learn English
Family Members who've Studies in ES Countries (family study ESC)
Liking of English Class (historical) (like E C)
Recognition of Need for English (historical)
Number of Favorite Teachers
Purpose to Learn English (reason learn)
Future Need for English

ENTRY/EXIT PROFICIENCY
Entering TOEFL Score (TOEFL entry)
Strongest Entry Skill (TOEFL)
Entry Placement Level (Place entry)
Strongest Entry Skill (Placement)

Appendix A, continued:

Sample Coding of Features

Feature 2: SCHOOL LEVELS AT WHICH LEARNED ENGLISH
1 none
2 secondary school only
3 upper primary and all of secondary
4 both primary and secondary
9 no evidence or not able to judge

Feature 27: MAJOR TEACHING FOCUS
1 grammar, translation
2 grammar, reading
3 all skills
9 no evidence or not able to judge

Feature 29: ACADEMIC STATUS
1 pre-undergraduate
2 during undergraduate
3 Bachelor's degree
4 Bachelor's and Master's degrees
5 PhD, Engineering, Medicine
9 no evidence or not able to judge

Feature 36: PERCEIVED NEED TO USE ENGLISH IN THE FUTURE
1 no need
2 occasional need, as in travel
3 regularly need (for profession, to travel)
4 need for possible immigration
9 no evidence or not able to judge

Feature 18: INDEX OF ORAL EXPOSURE IN CLASS

1 0
2 1-2
3 3-4
4 5-6
5 7-8
6 9-10
7 11-15
8 over 15
9 no evidence or not able to judge

Step A: Teacher's Instructions (totalled for levels: Elementary, Intermediate, Secondary, University, IEP)

	Code
never	0
sometimes	1
always	2

Step B: Teacher's Explanations (as above)
Step C: Index of Exposure = sum of A & B

Appendix B: General Characteristics of Sample Population

FIRST LANGUAGE	
Japanese, Korean	35%
Arabic	30%
Romance/German	23%
Chinese/other Asian	9%
Other	3%

AGE	
16-22	38%
23-30	38%
31-45	14%

ENTRY TOEFL	
293-400	30.3%
401-500	47.3%
501 or above	22.4%

GENDER	
Males	60%
Females	40%

Appendix C: Success Criteria

End of Session Progress Scores

SCORE	INDICATED
0	Retention at previous level
0.5	Probationary pass to next level
1	Unconditional pass to next level
1.5	Probationary skip a level
2	Unconditional skip a level

Overall Average Success Computation

AVERAGE PROGRESS SCORE	SUCCESS	PROGRESS INDICATED
1.0 or above	high	consistent unconditional passes
.60 to .99	medium	consistent passes (some probational)
less than .60	low	inconsistent passes, > 1 retention

For example:

Subject 10 entered at level 1 and enrolled 4 sessions:
level 1 --> 2 --> 2 --> 2 --> 3 probationary
code +1 +0 +0 +.5 = 1.5/4 = .375 (low success)

Subject 1 entered at level 1 and enrolled 6 sessions:
level 1 --> 2 --> 3 --> 4 --> 5 --> 5 --> 6
code +1 +1 +1 +1 +0 +1 = 5/6 = .833 (medium)

Subject 72 entered at level 3 and enrolled 4 sessions:
level 3 --> 5 --> 6 --> 7 --> unconditional pass (exit program)
+2 +1 +1 +1 = 5/4 = 1.25 (high success)

Appendix D: Expert Agreement with Success Ratings

POSITION	YEARS EXPERIENCE in Field	YEARS EXPERIENCE at CELT	# AGREE	/	# JUDGED
Director	18+	15+	33	/	35
Co-Director	26+	13+	31	/	31
Instructor	14+	6+	14	/	14

# of Experts Judging Case	# AGREE	/ # CASES	% AGREEMENT
3/3	13	13	100%
2/3	15	17	88%
1/3	9	9	100%

Appendix E: "Translated" Record Vector for One Subject

Record Vector is: 32111112334423344231234422112211336112133253311

1 AGE
3 over 30
2 SCHOOL LEVELS AT WHICH LEARNED ENGLISH
2 secondary school only
3 GENDER
1 male
4 FAMILY ENCOURAGEMENT TO LEARN ENGLISH
1 none
5 NUMBER OF FAMILY MEMBERS WHO'VE STUDIED IN ENGLISH-SPEAKING COUNTRIES
1 none
6 EXPOSURE TO ENGLISH AS THE LANGUAGE OF INSTRUCTION
1 none
7 YEARS SINCE LAST STUDIED ENGLISH
1 none or less than 1
8 AVERAGE NUMBER OF STUDENTS PER CLASS IN SCHOOL
2 31-40
9 PRIVATE SCHOOL
3 all of schooling
10 NUMBER OF YEARS OF UNIVERSITY ENGLISH
3 2
11 MONTHS OF INTENSIVE ENGLISH
4 7-12
12 CONTACT HOURS OF ENGLISH: SCHOOL AND UNIVERSITY
4 1001-1500
13 CONTACT HOURS OF ENGLISH: INTENSIVE ENGLISH PROGRAM
2 1-100
14 TOTAL CONTACT HOURS OF ENGLISH
3 1001-1500
15 TRAVEL TO ENGLISH-SPEAKING COUNTRIES
1 never
16 NUMBER OF NATIVE-SPEAKING TEACHERS
2 1 or 2
17 INDEX OF ORAL EXPOSURE IN CLASS
3 3-4
18 INDEX OF ORAL USE IN CLASS
4 5-6
19 INDEX OF WRITTEN USE IN CLASS
4 5-6
20 INDEX OF COMMUNICATIVE USE IN CLASS
2 1-2
21 INDEX OF EXTRACURRICULAR EXPOSURE AND USE
2 1-4
22 EXTRACURRICULAR READING
1 0
23 EXTRACURRICULAR WRITING
1 0
24 EXTRACURRICULAR SPEAKING
2 1-2
25 EXTRACURRICULAR LISTENING
2 1-2

Appendix E: "Translated" Record Vector for One Subject, continued

```
26    AUDIO-VISUAL INDEX
      1     0
27    MAJOR TEACHING FOCUS
      1     grammar, translation
28    NATIVE LANGUAGE
      3     Korean, Japanese
29    ACADEMIC STATUS
      3     Bachelor's degree
30    FIELD OF STUDY IN UNIVERSITY
      6     Social or Behavioral Sciences
31    LIKING OF ENGLISH CLASS (when in primary/secondary school)
      1     no (negative)
32    RECOGNITION OF NEED TO LEARN ENGLISH (when in primary/secondary
      school)
      1     no (no need)
33    NUMBER OF FAVORITE TEACHERS
      2     one
34    OPPORTUNITY TO USE ENGLISH FOR COMMUNICATION
      1     little
35    STATED PURPOSE IN LEARNING ENGLISH
      3     need TOEFL score for admission
36    PERCEIVED NEED TO USE ENGLISH IN THE FUTURE
      3     regularly need (for profession, to travel)
37    ENTERING TOEFL SCORE
      2     351-400
38    STRONGEST ENTRY SKILL AREA (TOEFL SUBSECTIONS)
      5     combination 1 & 3 (listening & reading)
39    ENTERING PLACEMENT SCORE/LEVEL
      3     3
40    STRONGEST ENTRY SKILL AREA (PLACEMENT SUBSECTIONS)
      3     reading
```

Appendix F: Identification of Limited Set of Taxonomy Features

"Strongest" Features Identified through Analysis of Probability Densities (3 categories of success):

Key:
+ Feature discriminated at minimum 15% difference across at least 2 success groupings and across at least 2 feature values
- Feature did not meet discrimination criteria

LL Features which discriminated in 5/6 data runs

	n=52	79	107	135	158	177
Gender	-	+	+	+	+	+
Private School	+	-	+	+	+	+
# NS Teachers	+	+	+	+	-	+
Index Written Use	+	+	+	+	+	-

LL Features which discriminated in 6/6 data runs

School Levels at which English was Learned
Family Members who've Studied in English-Speaking Countries
Exposure to English as the Language of Instruction
# of Years since last Studied English
# of Years of University English
Teaching Focus in English Class
First Language
Liking of English Class
Stated Purpose/Reason to Learn English

Entry Proficiency Features which discriminated in 6/6 data runs

TOEFL Entry Score (overall)
Placement Entry Score/Level

Additional LL Features Identified Through Individual Feature Performance

Contact Hours of English: School and University
Index of Oral Exposure in Class
Index of Oral Use in Class

Appendix G: Identification of Limited Set of Taxonomy Features

LL Features with values showing the greatest likelihood of accurate success predication (>= .66 likelihood), excluding value 9 "no evidence, not able to judge" (2 categories of success):

Feature & Value #	Value	Description	Decision Success Probability	Category (yours)	(OA)
2	3	Learned English upper primary & all of secondary	.68	_____	1
2	4	Learned English both primary & secondary	.88	_____	2
5	4	More than 3 family members studied in ES countries	.70	_____	2
7	2	1-2 years since last studied English	.73	_____	2
10	4	3 or more years of university English	.80	_____	2
12	1	no contact hours of English: school & university	.83	_____	1
12	2	1-500 contact hours of English: school & university	.66	_____	2
16	4	6 or more native English-speaking teachers	.75	_____	1
17	7	index of oral exposure in class 11-15 (high)	.69	_____	1
18	2	index of oral use in class 1-2 (low)	.68	_____	2
18	4	index of oral use in class 5-6 (medium)	.68	_____	2
19	1	index of written use in class 0 (none) n=1	.72	_____	1