DOCUMENT RESUME

ED 358 717 FL 021 276

AUTHOR Ginsberg, Ralph B.
TITLE Language Gains During Study Abroad: An Analysis of the ACTR Data. National Foreign Language Center Working Papers.
INSTITUTION Johns Hopkins Univ., Washington, DC. National Foreign Language Center.
PUB DATE Jun 92
NOTE 58p.; For a related document, see FL 021 277.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement Gains; College Students; Databases; Higher Education; *Language Proficiency; Language Skills; Language Tests; Listening Skills; Oral Language; Reading Skills; *Russian; Second Language Instruction; *Second Language Programs; *Study Abroad; Testing
IDENTIFIERS American Council of Teachers of Russian

ABSTRACT

This report presents results of a systematic analysis of an extensive database on study in the former Soviet Union assembled by the American Council of Teachers of Russian (ACTR) over a period of nearly 20 years. The purpose of the report is to document in detail the data used, the analytical strategy and methods employed, and most importantly the grounds on which substantive conclusions from the data rest. Three issues are emphasized: (1) interrelationships among preprogram language measures (standardized listening and reading proficiency tests, the Oral Proficiency Interview, and the grammar-based ACTR Qualifying Exams); (2) predictors of gain on various criteria, with particular focus on gender, knowledge of other languages, other individual attributes and characteristics of previous language learning careers, the Modern Language Aptitude Test, grammar-based ACTR qualifying exams, and proficiency in other language modalities; and (3) interrelationships among gains in different modalities. Data presented related to 658 students who studied in 4-month ACTR programs. Factors affecting gains are described (e.g., gender, age, and high school and college Russian). Analysis of residuals of gains indicates that gains on all modalities are positively interrelated, that gains in listening, reading, and oral proficiency tend to be associated, and that many students gain on one modality but not on the others. Seven figures and 25 tables are included. (JP)

National Foreign Language Center Working Papers

Language Gains During Study Abroad: An Analysis of the ACTR Data

by

Ralph B. Ginsberg
University of Pennsylvania

National Foreign Language Center

Language Gains during Study Abroad: An Analysis of the ACTR Data

1 Introduction

To many it seems almost self-evident that if students are to reach advanced levels of skill in a language, and if they are to acquire some kinds of language skill at any level, they must study in a country where the language is spoken and where they can engage with native speakers in natural, meaningful settings. It is also commonly assumed that students return from study abroad with much improved language skills. Be that as it may there still remains the empirical question of whether students *do in fact* gain from studying in language programs abroad, and if they do, who gains, what skills are acquired, what programs are effective, and how the learning process works.

This is the first of a series of NFLC working papers based on collaborative research sponsored by the National Foreign Language Center and the American Council of Teachers of Russian (ACTR) addressing these questions.* The paper presents the

---

*This research had its origins in a grant to ACTR and the NFLC from the Department of Education and has continued under a grant to the NFLC from the Ford Foundation. The paper draws on ideas developed jointly with my colleagues Richard Brecht and Dan Davidson, although they may not hold to all of my interpretations. Preliminary analysis of the data has been presented at several meetings and has been published in Richard Brecht, Dan Davidson, and Ralph Ginsberg, "The Empirical Study of Proficiency Gain from Study Abroad Among American Students of Russian: Basic Research Needs and a Preliminary Analysis of Data," in A. Barshenkov, T. Garza, et al. eds., Theoretical Problems in Foreign Language Teaching (bilingual edition, Vyshaja Shkola Press, Moscow, in press, 1990); and Richard Brecht, Dan Davidson, and Ralph Ginsberg, "On Evaluating Language Proficiency Gain in Study Abroad Environments: An Empirical Study of American Students of Russian (A Preliminary Analysis of Data)," in Z. Dabars, ed., Selected Papers Delivered at the NEH Symposium on Russian Language and Culture, Bryn Mawr College, Bryn Mawr, PA, May, 1990. A full discussion of the results will appear in our article "Language Gains during Study Abroad: The Case of Russian," Modern Language Journal (in press). These papers are referred to collectively below simply as **BDG**, which the reader can also consult for

results of a systematic analysis of the extensive database on study in the former Soviet Union (Russia) assembled by ACTR over a period of almost twenty years. It is meant to document in detail the data used, the analytical strategy and methods employed, and most importantly the grounds on which substantive conclusions from the data rest. The ACTR database is uniquely important for understanding study abroad because

- it has carefully collected before and after measures of three of the four basic language modalities (speaking, listening, and reading), so that gains can be assessed;
- it is rich in other pertinent variables, so that determinants and correlates of gain can be assessed; and
- it is extensive — 658 cases are used in the current analysis — so that conclusions are statistically defensible (i.e., not based on small numbers) with considerable control exercised.

Moreover, the data can be linked to that collected in the NFLC study of Language Learning Abroad, making it possible to determine what students actually do while abroad and hence to understand how gains come about.[1]

The paper is organized around three basic issues:

1. **Interrelationships among preprogram language measures** (standardized listening and reading proficiency tests, the OPI, and the grammar-based ACTR Qualifying Exams)

The interrelationships between language modalities and skills are interesting in their own right. Moreover, since it is rare to have such a comprehensive set of measures, it is of some practical value to determine whether (for people who apply for and are admitted to programs like ACTR's) a complex and problematic test like the OPI can be predicted from easier and less expensive measures.

---

references to the relevant literature. I also wish to acknowledge my debt to Paul Wheeling, who organized and managed the ACTR database, and without whose skill and attention to detail no analysis would have been possible.

[1] Of course the ACTR database does not have all of the variables one could think of, and in particular there are no standard motivational measures (although the RD ratings discussed below may be proxies for these). See **BDG** for discussion.

2. **Predictors of gain** on various criteria (OPI, listening, reading)

Predictors of gain is the central issue of the paper from both a policy and a scientific point of view. From a policy point of view, especially in conjunction with the NFLC study of Language Learning Abroad, they provide the empirical grounds on which decisions concerning when students should go abroad, which students should go, what preparation they should have, what skills should be emphasized, and what programs should be offered. From a scientific point of view they speak to basic hypotheses in the field that have heretofore been addressed on a rather intuitive basis, or at best on the basis of very small samples with little rigorous measurement. Particular attention is given to several factors including:

- Gender

It seems obvious that men and women (students) are treated differently by Russia (as they are in many countries) and that this *might* affect language learning. Gender differences turn out to be a very important theme in the NFLC/Ford study so it is useful to get as specific an idea as possible about what needs to be accounted for.

- Knowledge of other languages

The argument here is that people who know other languages have previous experience in language learning, and perhaps general linguistic knowledge, or that knowing other languages is an indicator of an aptitude for language learning, i.e that they are "expert" learners.

- Other individual attributes and characteristics of previous language learning careers

For the ACTR data these include the undergraduate college attended (previous training and learning culture), degree, major (a surrogate for unspecified motivational and attitudinal variables), the Institute attended in Russia(where ACTR students get five hours per day of formal language training), and program (changes in the pool of students who apply, the opportunities available in Russia, and ACTR's selection criteria).

- Modern Language Aptitude Test (MLAT)

The MLAT is a widely used instrument in the field but its predictive power for learning during study abroad has never been investigated.

- Grammar-based ACTR Qualifying Exams

The role of grammatical knowledge in language acquisition, an important issue in second language acquisition, is discussed at some length in **BDG**.

- Proficiency in other language modalities

Does, for example, good preparation in listening and/or reading proficiency facilitate the acquisition of oral skills in the study abroad milieu?

This in turn leads to the closely related, third major issue addressed in the paper:

3. Interrelationships among gains in different modalities

Here the implications for such important policy questions as program design and the place of study abroad in students' larger language learning careers (articulation) are paramount. Do students (in programs like ACTR's) tend to do well across the board, or is their learning skill-specific and individualistic? Does learning in one modality reinforce learning in the others? Should programs be targeted to specific skills where study abroad has real comparative advantage?

To anticipate,

- The other tests cannot substitute for the OPI although all test scores are positively associated (Section 3);
- Many of the factors investigated affect gains, especially in listening and oral proficiency (see Section 4.7 for a detailed summary); and
- Gains on different modalities are positively associated (Section 5).

The results are strong, often striking, but sometimes puzzling and subject to multiple interpretations. In any case they constitute a corpus of empirically well-grounded phenomena that need to be further explained and thereby set the agenda for future research.

2 Subjects, variables, and methods

2.1 Subjects

The analysis in this report is based on data relating to 658 students who studied in four-month ACTR programs at one of eight Institutes in Moscow or Leningrad over the period Spring 1984 (when ACTR started administering OPIs before and after the program) through Spring 1990.[2]

ACTR's general selection criteria are described in **BDG**. Since the students are not randomly selected, the conclusions here cannot be generalized with confidence to (randomly selected) students of Russian who might have the opportunity to study abroad. Nevertheless, to the extent that ACTR's selection criteria are controlled in the analyses — which, for the most part, they are in the regression analyses of gain — the results hold more generally. (Means, simple cross-tabulations, scatter plots, etc., which are not controlled, must, however, be treated as specific to the sample at hand.) By way of orientation to the analysis, it should also be pointed out that a key factor determining language gains, namely what the students do while in the USSR, is not part of the ACTR database, so we cannot explain *why* the results hold. Nevertheless, hold they do. As noted above, it is the task of subsequent research — and in particular of the NFLC Study Abroad Project — to explain them.

2.2 Language measures

The instruments measuring language proficiency in various modalities are described thoroughly in **BDR**. The Oral Proficiency Interview (**OPI**) and proficiency oriented tests of listening and reading developed by ETS in conjunction with experts in the field (**ETS Listening** and **ETS Reading**) were administered just before and at the very end of the program and are the basis of measures of gains. The ACTR **Qualifying Exams**, measuring achievement in grammar and reading, supplement the proficiency measures determining preprogram levels. With regard to the

---

[2]Students over 35, who know more than one slavic language besides Russian, who have had more than one previous immersion experience, who had studied in Russia in the previous semester and are hence in effect in the second half of a ten-month program, or for whom measures of change on any of the three criteria used below, have been dropped from the analysis. Data are available on an additional 182 students attending programs in Fall 1990 and Spring 1991 but is not complete enough to include in this analysis.

MLAT, the three parts of the test — ability to use analytic (MLAT3), synthetic (MLAT4), and memory-based (MLAT5) learning strategies — are analyzed separately, to see exactly what "aptitudes" and strategies might affect gains, as well as in combination as an overall score (MLATSF).

Over the period analyzed here the language data collected by ACTR varied somewhat. As a consequence the analyses are based on different numbers of cases, depending on the variables involved. ACTR started administering the OPI in Spring 1984. The ETS exams were phased in for Spring 1986. The MLATS begin in Fall 1984 and were discontinued in Spring 1990. In Spring 1990 the Qualifying Reading and Grammar tests were combined into one test. For students before Spring 1990, scores on the Qualifying Grammar and Reading tests have been combined into a composite score, comparable to that of Spring 1990, referred to below as **QualGen**. (The combination is weighted by the number of questions on the Reading and Grammar parts and is accordingly dominated by the Grammar part.) Moreover, there is a small amount of missing data on all of the instruments. For example, it is not possible to calculate change scores for 12 students because they were not tested either pre or post. Table 1 presents a summary of what data are available by program date.

Descriptive statistics for the quantitative (pre)program language measures are given in Table 2. The two ETS tests have been normed by ETS to make them comparable to the levels of the OPI (see **BDG** for details).

Frequency distributions for preprogram OPI and the proficiency coded ETSs are given in Table 3. Note that both ETS tests have a ceiling at 3, i.e. students cannot get a rating above this level, and a floor at 1. (On the OPI no one happened to score above 3.) Further, almost all students are rated at 1 on Listening Proficiency (with none below 1), which is questionable in light of the OPI scores. Because of these artifacts I am skeptical about the validity of the norming, and the analysis accordingly focuses on the raw scores which are well spread and unbounded.

2.3 Student characteristics in the ACTR database

Table 4 presents descriptive statistics for the personal characteristics and educational history variables that are used in the explanation of gains. (The Institute of study in Russia is also included.) Three variables (country of birth, class and lab hours) were

examined, but are not used here because in preliminary studies, they added no explanatory power to the other variables in the table. For many of the variables there is a small amount of missing data; these values were imputed, as indicated in the Table, so as to retain the cases in the analysis.[3]

2.4 Resident Director Ratings

As discussed in **BDG,** ACTR Resident Directors at each Institute were asked to rate students on a scale of 1 (lowest) to 5 (highest) on seven criteria, comparable to the kinds of ratings one finds on a recommendation to graduate school (abbreviations used in the regression analysis are in bold), *viz.*

Intellectual motivation (**IntelMot**)
Natural ability to learn (**NatAbil**)
Willingness to use Russian (**UseRuss**)
Cultural adaptability (**CultAdap**)
Take advantage of cultural opportunities (**CultOpp**)
Ability to work in a group (**WrkInGp**)
Leadership potential (**Ldrshp**)

Since these are the only available measures of motivation and attitudes, their effects are explored below. For all practical purposes Resident Directors used only levels 3, 4, and 5, so in the regressions below all seven variables are recoded to 3 or less, 4, and 5. Since rating on the various criteria are strongly associated, a simple sum of the ratings (SRD), designed to capture what they have in common, was also explored as a determinant. Table 1 shows that there are considerable missing data on these variables, although the number of valid cases is still quite large. Resident Directors either rated everybody or nobody in their programs, but there does not seem to be any systematic bias associated with this.[4]

---

[3]Preliminary studies showed that the imputation does not affect the results. Of course missing data on the language measures (dependent and key independent variables) were never imputed, and these cases were accordingly dropped from the relevant analyses.

[4]There is a small amount of missing data on each criterion for students who have most other ratings; this has been imputed to the mode on that criterion.

2.5 Analytical methods

The methods of analysis used in this report follow standard statistical practice. Where prediction and explanation are concerned (as in the analysis of the determinants of gain) regression analysis or discriminant analysis are used. For the most part regressions are estimated by ordinary least squares, but in the analysis of OPI gains (Section 4.5) binary and ordinal probit regressions are also used. Correlations, regressions, and principal components are used to explore the interrelationships between variables when none is causally prior (as in the discussion of the interrelationships among gains in section 5). In the regression analyses conclusions are based on models that control for *all* (available) relevant factors, i.e. the relationships hold over and above what can be explained by other relevant variables. For each analysis results have been carefully checked to see that they are not influenced by outliers or leverage points — which is entirely possible but happens not to be the case for these data.

In several of the analyses I have looked at the data from many points of view to see that the results are stable, i.e. not overly dependent on a specific model or method of analysis. This is particularly relevant to analyses involving the OPI, an ordinal variable which does not accommodate gracefully to standard fully quantitative or qualitative statistical models, whatever its other uses may be. In presenting the analyses I have used a blend or graphical, tabular, and quantitative summaries which I hope will make the main conclusions accessible to readers who are not technically trained in data analysis but at the same time give readers who are interested in the statistical results what they need to know.[5]

3 Interrelationships among preprogram measures

The interrelationships among measures of various language skills, and in particular whether the OPI, which is expensive and time consuming to administer, can be adequately estimated by other language tests, are important practical questions for research on study abroad. If study abroad has any comparative advantage over domestic programs, it should be on speaking, and especially on pragmatic and sociolinguistic skills which are salient in meaningful interactions with native

[5] All of the graphics and much of the regression analysis were done in Data Desk®, although some of the more complex models were computed using Crawtran and BMDP.

speakers. It is not at all clear whether the OPI measures these skills. (See BDG for a detailed discussion of the OPI as a research instrument.) For the purposes of this paper, investigation of how the preprogram language measures relate to one another is an important preliminary to understanding the subsequent analyses of gains, in which they play a key role.

Before looking at the language variables per se, it might be interesting to look at how they all relate to the cheapest and most accessible "measure" — "proxy" would be a better word — of language level, years of study in college (College Russian). Figure 1 tells that story. Panels I-V are boxplots of the five quantitative measures against four levels of College Russian: 2 years or less, 3 years, 4 years, and 5 years or more. In these plots (see the Data Desk® Documentation for exact definitions):

> the boxes show the middle half of the distribution (from the 25th to the 75th percentile, so the box height is the interquartile range);
>
> the lines inside the box indicate the median (with the shaded area around it a 95% confidence interval);
>
> the whiskers enclose most of the distribution; and
>
> the circles and stars indicate outliers.

It is clear that college Russian is strongly related to Reading (Panels I and II), especially using the achievement oriented ACTR Qualifying Exam, although the 2 and below and 3 year groups are not discriminated. (Note that the median increase and that the 95% confidence intervals do not overlap. The variation in the groups, as measured by the size of their boxes with their whiskers, is reasonably comparable, though not constant.) The boxplots as a whole, however, substantially overlap, indicating that there would be considerable error in trying to predict the score from years of study alone (obviously).[6] College Russian is a less good predictor of Listening Proficiency (Panel III) and Grammar (Panel IV). QualGen, a combination of Reading and Grammar, is in between the two. Panel VI shows the relationship

---

[6]Cross tabulation of the proficiency coded ETSL and ETSR show the same positive but relatively weak relationship to years of college Russian; e.g. with 2 years of Russian 13.9% are above 1 on listening, while with 5 or more years (only) 26.6% are above 1; for reading 24.6% are at 2 or above as opposed to 14.2, 10, and 12.8% for 2,3, and 4 years, respectively.

between years of college Russian and OPI scores. Although this table shows a generally positive relationship (with relatively more students in the top left and bottom right than elsewhere — compare row percents down the columns), it is hardly overwhelming. It should be added that the F tests in the ANOVA for the quantitative measures and χ^2 for the OPI are highly significant. Thus, altogether, relationships to college Russian are in the right direction but there is considerable variation around the trends; not surprisingly, years of study is a weak proxy for more direct measures.

Turning now to the measures themselves, Figure 2 gives scatterplots and correlations between the quantitative measures. The correlations are all very high (even for individual-level data), indicating that these variables are highly interrelated. (The correlation of .99 between QualGram and QualGen results from the fact that the former is the major component of the latter.) In a principal components analysis of ETSL, ETSR, and QualGen, aimed at capturing the joint variation of these variables in a single index, the first principal component accounts for 71% of the joint variance (with the other two components nonsignificant), again indicating very strong interrelations. Thus, to some extent the quantitative variables can proxy for one another.

Figure 3 gives boxplots of the five quantitative variables and the first principal component of ETSL, ETSR, and QualGen (a weighted average of these variables) against four levels of OPI scores. We see very clear trends in the medians, especially for the principal component, but also enough overlap in the boxes to make accurate prediction of the OPI problematic. For purposes of comparison with the boxplots, histograms of the distributions of the principal component are given in Figure 4: in general, as the OPI score increases the histogram moves to the right (indicating the positive relationship between the two variables), but there is considerable overlap in the histograms, indicating that any prediction would produce considerable lack of fit.[7]

---

[7]For predictive purposes one would put three cutpoints on the PC axis and predict OPI as a function of the interval in which the PC of a given case falls. Discriminant analysis would choose these points optimally, but it is obvious from the Figure that there is no way to do this well.

A regression of the OPI score on ETSL, ETSR, QualGen, and College Russian, which would be used for prediction of the OPI taking all of the other language measures together, shows that only ETSL and QualGen are significant (with t statistics of 4.78 and 7.29). That is: once ETSL and QualGen are taken into account, ETSR and years of College Russian add nothing to the prediction of OPI score. Regression of OPI on ETSL and QualGen alone produces a respectable R^2 of 27.1 and t statistics of 6.06 and 8.33, respectively. For issues addressed in this section, however, the quality of predictions, not measures of effect, is at issue, and for that the relationship between OPI and ETSL and QualGen must be examined more closely.

Figure 5 presents a scatterplot of ETSL against QualGen with the points corresponding to various levels of OPI distinguished. The quality of prediction depends on the separation of the OPI groups in the plot. It is clear from the top panel that one could differentiate 0+ and below from 2 and above — not a great accomplishment — by predicting 2 and above for any point up and to the right of the diagonal line and 0/0+ down and to the left, although even here a few mistakes would be made. When 1's and 1+'s are added to the picture it is clear that, although there is some tendency for the 1+'s to be up and to the right and 1's to be down and to the right, there is no line that could be drawn that would not lead to many errors of classification. Were the two panels to be superimposed, again there would trends but it would not be possible to separate 0+ and below from 1 or 1+ from 2 and above, and there would be many errors comparing 0+ and below with 1+ or 1 with 2 and above. Looking at the matter more quantitatively, optimal assignment of students to the four OPI levels in Figure 5 on the basis of a discriminant analysis, using ETSL and QualGen as predictors (ETSR and CollRuss adding no information), leads to correct classification 40.2 percent of the time (77.8 percent correct for 0/0+, 36.5 percent for 1's, 20.2 percent for 1+'s, and 67.3 percent for 2 and up), with many errors in the middle groups. Thus, although there is a clear relationship between the variables, and a 60 percent error rate might be acceptable as a crude first cut, one could not substitute measures on other modalities for the OPI for most purposes.

4 Predictors of gain

4.1 Measures of gain

For the quantitative variables measuring listening and reading skills (ETSL and ETSR), gain is intuitively defined as the difference between preprogram and postprogram scores, and the assessment of factors affecting gain can be accomplished

with ordinary least squares regression.[8] On both Listening and Reading Proficiency, gain is very strongly, negatively related to preprogram level, with correlations of -.522 for Listening and and -.344 for Reading, i.e. the higher the initial level the less the gain. This phenomenon is consistent with a "normal" s-shaped learning curve, since most people are beyond initial levels. (In the case of ETS Reading, the correlation is probably attenuated somewhat by the ceiling on the test.) As a consequence of these strong relationships looking at the effects of other variables (such as Gender, previous immersion, MLATS, etc.) makes sense only with preprogram levels controlled, i.e. in terms of what is *not* explained by preprogram level.

For the OPI the situation is considerably more complicated because of the nature of the scale. The OPI score is an ordinal variable, and as such each level should be thought of as a grouping of scores on an underlying unobserved scale of proficiency on which variation is more continuous. Grouping loses information — in effect introducing a measurement error — in that students with quite different (unobserved) proficiency levels could be given the same OPI score. In general, grouping attenuates relationships.[9] Gain on the OPI (i.e. the difference between scores pre and post), also ordinal, has the same difficulties, difficulties which it inherits from its components. Table 5 shows the relationship between preprogram and postprogram OPI scores. The main diagonal — the cells enclosed in boxes — represents no gain; losses are below the diagonal, gains above; moving to the left of the main diagonal in any row indicates a loss; and moving to the right one (two, three, ...) column(s) indicates a gain of a half (one, one and a half, ...) point(s). As with Listening and Reading, and for much the same reasons, the association is

---

[8]For the ACTR data there is reason to consider ETSL2, rather than the gain, as a criterion because the two ETS tests of listening proficiency are not of the same difficulty (see **BDG**). As a technical matter in OLS regression it does not matter whether the dependent variable is taken as gain itself or the post program score, as both yield the same results, i.e. the same residual sums of squares, coefficients, and t statistics — which in this section are our primary concern. More complex latent variable models, which are not warranted here, would use ETSL2 as dependent. Gain is fine for ETSR, for which the two tests are equivalent.

[9]Allowing for measurement error in coming up with the OPI score introduces another level of complexity that I do not consider here.

strongly negative, with the probability of gaining sharply reduced as initial level increases.

In my view, because the number of levels one could gain is so limited, especially as initial levels increase, there is no fully satisfactory way of quantifying gain using the OPI. Instead, analyses with a number of different definitions and procedures for controlling for initial levels were run in order to get a handle on the effects: results that are not sensitive to the definition and procedure are likely significant, whereas results that hold for only one procedure may well be artifacts. Two gain criteria are used in the Tables of Section 4.5: a simple no increase vs. increase (coded 0/1); and a three-level variable, no gain or loss, a gain of one level, and a gain of two or more levels (coded 0/1/2).[10] Of course, for both definitions of gain preprogram level must be controlled in the analysis. An adjustment of the raw half points gained, reflecting the difficulty of gaining as a function of the preprogram level (see BDG), was also explored in some detail but found to be essentially equivalent to the 0/1 and 0/1/2 criteria.

4.2 Regression strategy

The following three sections present the results of analyses of factors affecting gains on each of the three criteria of language proficiency defined in the ACTR database. The results derive from a series of regression analyses in which it is possible to exercise sufficient control to estimate the effects on gain of individual characteristics, previous educational history, and language abilities over and above any relationships that these variables might have with one another. The focus is thus on the estimated coefficients[11] and their statistical significance (in contrast to the analyses of the interrelationships among the language measures in the previous Section and among the gains themselves in the next Section, where the focus was on goodness of fit).

---

[10] A four-level criterion (0, 1, 2, 3 or more half points) was also explored but turned out to be equivalent to the three-level criterion.

[11] The coefficients should be interpreted as the effects of the corresponding variable on gain after the effects of all other variables in the equation have been removed. That is, they measure effects that cannot be accounted for by the other variables, although the effects may be accounted for by variables *not* in the equation.

The regression strategy in each section is the same. I start with a baseline of student background characteristics, as listed in Table 4, and the preprogram level of the criterion in question. Possible effects of undergraduate college, program date, Institute of study, highest degree, and undergraduate major are explored by examining their relationship to residuals from the baseline. With the notable exceptions of undergraduate major and program date for oral proficiency, these variables turn out to be generally nonsignificant and are not discussed further. The language measures (MLATs, Qualifying Exam, and preprogram scores on other criteria) are then added to the baseline, separately and in combination, to gauge their effects. The different sets of regressions are based on different numbers of observations because of the missing instrument pattern specified in Table 1, so R^2 across sets cannot be directly compared. There is always sufficient information in a set to evaluate the significance of coefficients. Having identified the factors that affect the gain, nonsignificant variables are eliminated to produce a "good" model in which the coefficients and their Standard errors (and hence the t statistics) are estimated with the greatest possible precision.

As noted, regressions of gains in ETS Listening and Reading are estimated by ordinary least squares (OLS). The methods of choice for estimating effects on gains on the OPI, a qualitative variable, are probit and ordinal probit regression, depending on whether the criterion is 0/1 (no gain/gain) or 0/1/2 (no gain/one step/two or more steps).[12] Logit and ordinal logit models could also be used but they yield almost exactly the same results as the probits. Indeed, as will be seen in Tables 17-20, OLS (which is used to get initial values for the probit estimations) gives essentially the same levels of significance as the probits (and logits) themselves. (Coefficients are not directly comparable because of normalizations in the probits, but the ratios of coefficients can be compared.) That is: probits, logits and OLS lead to the same qualitative conclusions. Accordingly, when variables are being screened

---

[12]Probit regressions have been chosen because of the continuity of the underlying standard normal theory model estimated by OLS, but the equivalent logits have equally good rationale. In the standard model (used for ETSL and ETSR) the gain y is modeled as $y = x'\beta + \varepsilon$, where x is a vector of independent variables, β the coefficients, and ε a normal error. In a probit model an observed variable $y^* = x'\beta + \varepsilon$; if $y^* < 0$, the observed $y = 0$, else if $y^* \geq 0$, $y = 1$. The ordinal probit also specifies $y^* = x'\beta + \varepsilon$, but there is a second cutpoint (μ, a parameter to be estimated) in addition to 0, which defines three intervals for y^*, with the observed value of y.determined by the interval in which y^* falls.

for effects I simply report the OLS results as they are easier to interpret for many readers. For the "good" models both OLS and Probit (ordinal probit) coefficients and t's are presented.

4.3 Predictors of gain: Listening Proficiency

Tables 7 - 11 contain the results of the analysis of factors related to gains in Listening Proficiency. The variables in Table 7 constitute the baseline against which other factors are assessed. As noted above, the preprogram Listening Proficiency level is a strong determinant of gain, no matter what is controlled, and must be included in any analysis. *Over and above what can be accounted for by preprogram level and all of the other variables in the equation,* several individual characteristics which relate to the hypotheses (questions) put forward in Section 1 have significant effects judging from the t statistics:[13]

Gender: men gain more than women;

Age: younger people gain more than older;

HS Russian: people who have had Russian in high school gain less than people who haven't;

NonSlav: people who know other languages gain more than people who don't; and

PrevImm: students with a previous in-country immersion experience gain more than those without.

The implications of the Gender effect are discussed at length in BDG. Knowledge of other languages and a previous immersion, and age, are as one would expect. The lack of effect of college Russian is probably the result of controlling for preprogram levels, with which it is correlated. The negative effect of having had Russian in high school is frankly puzzling: stories could be told, but in my view it is probably an artifact of its interrelations with the other variables in the equation.[14]

---

[13] t statistics > 1.65 are significant at the .05 level (one tailed); t's > 1.96 are significant at the .025 level; t's >. 2.33 are significant at the .01 level; and t's > 3 are very highly significant (a > .001).

[14] Knowledge of another Slavic language is rare in the sample and included here only for exploratory purposes.

Model 2 in Table 7 also shows that

> Preprogram Reading Proficiency level is very strongly related to gains in Listening.

One possibility is that Reading Proficiency is picking up the divergence between "true" listening proficiency and the test score (i.e. that an underestimate of preprogram listening proficiency by the ETS test shows up as an apparent gain explained here by the reading test), but this is unlikely in view of the strength of the effect. Rather it seems that students who are better readers are in a better position to acquire listening skills. It is also worth noting that the R^2 in these equations is quite respectable for individual-level data.

Screening for Major, highest Degree, Undergraduate college, Institute, and Program date shows that none of these variables is significant and that they do not affect the relationships in the Table.

Table 8 shows three models adding the different parts of the MLAT to the baseline.[15] From the Model 3 it is clear that MLAT3 (use of analytic strategies) is highly significant, MLAT4 (use of synthetic strategies) is significant, and that MLAT5 (use of memory-based strategies) is nonsignificant. (A separate test shows that the total MLAT score, MLATSF, cannot substitute for its individual components.) Comparing Model 5 and Model 4, the MLAT3 effect remains very strong, but preprogram Reading Proficiency seems to account for much of the MLAT4 effect. Reading Proficiency could be a mechanism for MLAT4 effects or simply a correlate explaining its apparent effect. I shall return to changes in the baseline effects presently.

Tables 9 and 10 explore the effects of the ACTR Qualifying Exams. In the Model 6 of Table 9 QualGen seems to have an effect but this is accounted for in Model 7 by ETS Reading (which is not surprising because the two exams have somewhat overlapping content). Table 10 shows that it is the Reading component of the Qualifying Exam, not the Grammar, that matters (Model 8), and that the the ETS exam dominates the Qualifying Exam as a predictor of Listening gain (Model 9).

---

[15] Note that Table 8 is based on over 100 fewer cases than Table 7 so the R^2's cannot be compared with an F test. They are, however, quite comparable in magnitude as they should be if the selection on having an MLAT is random with respect to the effects.

Table 11 presents a "good" model for the predictors of gains in Listening Proficiency, with the nonsignificant variables College Russian and knowledge of another Slavic language dropped from the baseline and the highly significant predictors MLAT3 and Reading Proficiency included, as in Table 8. The high R^2 indicates that this is an excellent model indeed.

4.4 Predictors of gain: Reading Proficiency

Tables 12 - 16 contain the results of the analysis of factors related to gains in Reading Proficiency. As Table 12 shows, factors in the baseline are generally nonsignificant with only knowledge of other (nonSlavic) languages having a clear effect. Preprogram Listening Proficiency has a positive effect over and above Reading Proficiency but it is not nearly as strong as the effect of Reading on Listening. One should bear in mind that study abroad is not designed to teach reading, and that there is no apparent comparative advantage in learning to read abroad as opposed to at home. In this light the lack of effects is not surprising.

Table 13 shows the effects of the MLATs. Although it seems from Model 3 that MLAT3 and MLAT4 are individually significant, an F test comparing Model 3 and Model 4 shows that the total score, MLATSF, contains all of the information in its components and can substitute for them. In Model 5 the MLAT effect is not explained by preprogram Listening Proficiency; the MLAT does, however, account for the nonSlavic effect, and Gender (men gaining more than women) almost reaches significance.

Tables 14 and 15 examine the effects of the ACTR Qualifying Exams. In Table 14 QualGen is a significant predictor of Reading, as it was with Listening Proficiency, even with preprogram Listening controlled. From Table 15 it is clear that the Reading component is significant, but that the Grammar component is not (Model 8). Inclusion of preprogram Listening Proficiency does not explain the QualGen effect (Model 9). Indeed in Model 2 of Table 15 the Listening Proficiency effect is to a considerable extent accounted for by the reading exam.

Two "good" models for gains in Reading Proficiency, with nonsignificant baseline variables dropped, is presented in Table 16. (There are two models here because of the diffent sample sizes; I have included QualRead instead of QualGen because the QualGen effect is entirely QualRead and it is a more relevant measure of initial levels.) The good models are lacking in predictors, with only the MLATS and

QualRead highly significant. Preprogram Listening Proficiency is barely significant and Gender is still nonsignificant. Of course preprogram ETS Reading Proficiency is very strongly negatively related to gains, with the ceiling no doubt playing a role. I also think that the QualRead effect is picking up the difference between "true" reading proficiency and the preprogram test score, so that the QualRead effect has little substantive meaning. I reiterate that study abroad is not oriented toward reading.

4.5 Predictors of Gain: Oral Proficiency

For the analysis of gains in oral proficiency, as measured by the OPI, a somewhat different regression strategy from that of the previous two sections was employed. First, as discussed in Sections 4.1 and 4.2, the ordinal but nevertheless qualitative nature of the scale requires examination of several criteria of gain. Two criteria as defined above are used: 0/1 (no gain/gain) and 0/1/2 (no gain/gain of 1/gain of 2 or more). Second, different regression models are required for the different criteria, Probit for 0/1 and Ordinal Probit for 0/1/2. OLS results are also presented for comparability with previous sections. Third, it turns out that college major and program date have significant effects and must be added to the baseline (referred to as the "expanded baseline" below). Bear in mind that a given variable may not necessarily have the same effect on the two criteria. As a general rule a variable that discriminates only between gains of 1 and 2 will be significant for 0/1/2 but nonsignificant for 0/1. By contrast a variable that discriminates between 0 and 2 will be significant for 0/1 but not for 0/1/2.

Table 17 presents a model of gain in oral proficiency using the 0/1 criterion. Independent variables include the expanded baseline and preprogram Listening and Reading Proficiency scores, estimated by both OLS and Probit regressions.[16] First of all, note that the t statistics are essentially the same and that the Probit coefficients

---

[16]Models with just the expanded baseline but not the ETS Proficiency score were also run. The only difference in results is that knowledge of nonslavic languages is significantly positive without the ETS scores: thus knowledge of other languages affects preprogram levels and operates through these intervening variables, rather than having a direct effect of its own, e.g. through learning strategies.

are about 3 to 4 times the size of the OLS.[17] Note that in the expanded baseline only Program date and High School Russian are significant. Students with High School Russian are *more* likely to gain than those without (in contrast to gains in Listening Proficiency); it remains to be explained. I do not see any obvious explanation for the effect of Program date, i.e. that students in recent programs are less likely to gain than students in earlier programs; it cannot be explained by selection criteria related to any of the other variables in the study, e.g., that students are less well qualified in 1990 than in 1984, since this is controlled. Preprogram proficiency on other language modalities, as in the previous analyses, positively affects gains, but anomalously, only Reading, and not Listening Proficiency is significant.

Table 18 presents the results of the effects of same independent variables, this time for the 0/1/2 criterion. Note again the equivalence of the OLS and Ordinal Probit results. High School Russian maintains its significance. Knowledge of nonSlavic languages is almost significant. Undergraduate Major is significant on this criterion. The sign indicates that, in contrast to Russian and areas studies major, if students in the humanities, social and natural sciences gain at all they are likely to gain considerably. Preprogram Listening Proficiency is significant on this criterion and in the right direction.

Analyses of the effects of the MLATs and ACTR Qualifying Exams, not reported in tables here, show that the MLATs (the parts or the total score) are completely insignificant as predictors of gains in oral proficiency on either criterion. The Qualifying Exams, with Reading carrying most of the weight, are significant predictors of OPI gains, as they were for the other modalities.

Tables 19 and 20 present "good" models, i.e. with nonsignificant variables on both criteria dropped, for 0/1 and 0/1/2 respectively. Note the positive QualGen effect. Results in Tables 19 and 20 are generally consistent with Tables 17 and 18, with the exception that preprogram Listening Proficiency loses explanatory power. To get some idea of goodness of fit of these models, Figure 6 shows predicted values from the "good" model for 0/1/2 in the OLS regression (equivalent to the Ordinal Probit). Although the histograms move to the right with increasing gains (as they should in

[17]This reflects the scaling of the Probit model, as does the constant term, which should not be interpreted.

a well-fitting model), the separation is hardly dramatic. Remember, however, that it is the coefficients and their significance, not goodness of fit, that are at issue in this section.

As noted in Section 4.1 a vexing problem in the analysis of OPI gains is how to adequately control for initial levels. A direct method of control is to hold initial level constant by running the analysis separately for different initial groups.[18] Students with preprogram OPIs of 1 and 1+ are numerous enough to analyze. Results for initial OPIs of 1 are generally consistent with those of the whole sample, which is not surprising in that they constitute the large majority of students. On the 0/1 criterion in the expanded baseline, only Program date is significant. Slavic languages is borderline but negative ($t = -1.6$) and is probably not worth interpreting. ETS Reading and QualGen are highly significant, but ETS Listening is not significant. With the 0/1/2 criterion, Major and ETS Listening are significant, as in the whole sample.

The situation is more interesting with people who start at 1+. Only the 0/1, which represents a significant gain to 2 or above, is used because of the small sample size. Table 21 gives the results from the OLS regression. Gender is highly significant, with men more likely to reach 2 than women; this accounts for the incipient Gender effects in the whole sample and has important implications (see **BDG** for discussion). Both knowledge of Slavic and nonSlavic languages have positive effects, indicating the possible effect of previous language learning experience (again see **BDG**). QualGen (the numbers are too small to separate Reading and Grammar) is very significant, while MLATs and proficiency in Listening and Reading are not. All in all the fit is quite good, especially since R^2 tends to be reduced with qualitative dependent variables.

In summary, analysis of gains in oral proficiency yields interesting and intriguing results which should certainly stimulate further research.

4.6 Resident Director ratings

---

[18] Against this strategy is the fact that the numbers in each group are perforce smaller than the whole and that it is difficult to estimate common structure. With the ACTR data, however, the numbers in the subgroups are still much larger than those in other studies in the literature.

As discussed in Section 2.4 (*q.v.*) the ACTR database contains ratings by the Resident Directors at the Institutes in which the students lived and studied on seven attitudinal, motivational, and behavioral factors, all of which are arguably important correlates of gains. An exploratory analysis of the effects of these variables is presented here. The Resident Director (RD) ratings are kept separate from the main body of this Section partly because of the large amount of missing data, and partly because the judgments could have been made on the basis of observed gains, thus undermining the basic assumptions of the regressions.[19] Nevertheless, there does not seem to be any systematic bias in the missing data, and the results are so interesting that they are worth investigating in future research.

Gains in all three language modalities are examined. With the OPI it turned out that the three-level criterion (0/1/2) and the two-level criterion (0/1) produced the same results, so only the former is reported. In the regressions Major and PgmDate were added to the baseline variables to try to catch any selection bias in the missing data, and in any case they should be included as controls because they are sometimes significant in the RD rating subsamples. The regression strategy used was, first, to examine the effects of all seven RD rating variables controlling for the expanded baseline; second, to eliminate the nonsignificant ones to get good estimates of the effects of the significant ones; third, to then control for other language tests and MLATs, as appropriate to the dependent variable; and, finally, to check whether, with the language controlled in the equation, any nonsignificant RD ratings should be added back into the equation (none did). Adding all of the RD ratings to produce a single index was also examined but in every case this proved inferior to the original rating criteria.

Table 22 presents the results for gains in ETS Listening Proficiency. When only the baseline is controlled (Model 1) Natural Ability and Leadership are clearly significant and Cultural Adaptability is borderline. The negative sign for Leadership means that the greater the rated leadership potential the *less* the gain in Listening. One might speculate that this has something to do with how students spend their time with other Americans and with Russians. With the language measures controlled, Natural Ability loses its significance. This is not surprising since the

---

[19]The RD rating would then be endogenous variables and the OLS regressions subject to simultaneous equation bias.

MLAT is supposed to be measuring natural ability too.[20] In the expanded baseline Major is significant (with Russian and area studies majors gaining less), as are Gender and Age.

Table 23 presents the results of the analysis for changes in ETS Reading Proficiency. Three different variables are significant here: Willingness to Use Russian, Taking Advantage of Cultural Opportunities, and Ability to Work in a Group. Interestingly, only CultOpp has a positive sign. Ability to Work in a Group may operate like Leadership (however that works); and Willingness to Use Russian may indicate that students who are not willing to speak are reading, and vice versa. When the language variables affecting Reading gains are controlled, UseRuss loses its significance somewhat. Note that in this subsample nothing in the baseline except Major is significant.

Finally, Table 24 presents the results for gains on the OPI using the three-level criterion (0/1/2). Four RD rating variables are significant, Natural Ability, Cultural Adaptability, and Ability to Work in a Group very much so. Leadership potential is significant without the language controls and insignificant with them, and probably should be ignored. It is interesting to note that adding the MLATs to the equation does not account for the rating on Natural Ability; on the contrary, natural ability maintains its strong predictive power and MLATS are completely nonsignificant. Either Natural Ability is picking up factors related to gains which are not captured by the language measures, or the judgment itself is based on observed gains in oral proficiency and hence should not be included in the equation. Again, only further research will tell. As for the baseline, only Major and Program Date are significant in the subsample with RD rating controlled.

4.7 Summary: Factors Affecting Gains

Table 25 summarizes the analysis of factors determining gains on all three modalities so that the effect of each variable across the board can be easily seen. With regard to the questions raised in Section 1 the following comments are in order:

---

[20] In an analysis of the relationship between rated natural ability and the MLATs the rating is a significant predictor of MLAT4 ($p = .002$ in the ANOVA), but not significant for MLAT3 and MLAT5, so synthetic strategies seem to weigh heavily with the RDs.

Gender	Men are more likely than women to gain on Listening, and to go from 1+ to 2 and up on the OPI. Needs study based on data on experiences abroad.
Age	Younger students gain on Listening.
HighSchool Russian	Negatively related to Listening, positively related to OPI; may be an artifact of the correlations among other variables.
College Russian	Not significant; accounted for by preprogram language measures.
Slavic Languages	Significant for OPI 1+ to 2 and above; not significant for Listening unless MLATS included.
NonSlavic Languages	Significant for OPI and significant for Listening and Reading when MLATs not included: the more other languages known the more the gain.
Previous Immersion	Positive for Listening but not significant for OPI.
Major	Humanities, social science, and science majors gain more, if they gain at all, on the OPI than do Russian language and literature or area studies majors.
Program Date	Negatively related to OPI for reasons that are not clear.
MLATs	MLAT3 (analytic) and MLAT4 (synthetic) are good predictors of Listening and Reading; MLAT5 (memory) is not predictive. MLATs *do not* predict OPI.
Qualifying Exams	Qualifying Reading and the exam as a whole predict all modalities, but the Grammar part by itself does not add anything to Reading.
Preprogram Level on Modality	On all modalities the higher the initial level the less likely a gain; this is a function of the learning curve and the nature of the scale.
Preprogram Reading Proficiency	Like the Qualifying Exam, higher preprogram Reading Proficiency seems to facilitate gains on the OPI and Listening; Reading may be picking up measurement errors, or the ceiling may be coming into play, but this is not likely.

Preprogram Listening Proficiency	Does not seem to be very significant for the other criteria.
RD Ratings	Several of these variables are significant for each modality; subsequent research should explain these effects and include more measures of this type; basis of rating of natural ability should be explored.

5 Interrelationships between gains in different modalities

In this final substantive section the interrelationships among gains on the different modalities are discussed. Having examined gains on each modality separately as functions of their determinants, we now ask how the gains relate to one another. The bearing of the results on issues of counselling and program design is discussed in **BDG** and will not be repeated here. The facts need to be established, however. As in the regression analysis control is essential. The fundamental question is: after the determinants of gain isolated in Section 4 have been taken into account, do students who gain on one skill tend to gain on the others? From a statistical point of view the correct approach is a multivariate analysis, but preliminary studies show that little would be added to the conclusions that can be drawn from the more informal plots and regressions presented below.

Now the raw gains in ETS Listening (ΔETSL), ETS Reading (ΔETSR), and the OPI using the 3-level criterion 0/1/2 (ΔOPI), are positively correlated but not highly so. The correlation between gain ΔETSL and ΔETSR is .242, and the correlations between the ETS's and ΔOPI are .065 and .100, respectively; it should be remembered that correlations are attenuated (shrunk toward zero) when one or both of the variables are qualitative. The results of the previous sections show that the initial levels of all of the variables are correlated, and that the gains themselves are affected by initial levels on other modalities. Since initial levels are part of gain *by definition* correlations between gains could simply reflect initial levels. For this reason we must look at the residuals of gains after initial levels and any other pertinent variables have been removed. (This is what a multivariate analysis of covariance amounts to.) Thus, the analysis of this section addresses what is *not* explained about gains by the variables in the previous one.

Figure 7 displays in graphical form the relationships among the residuals of gains (called resΔETSL, resΔETSR, and resΔOPI) with initial levels of all three modalities removed. resΔETSR is plotted against resΔETSL for three bands of resΔOPI with the coordinate axes as guidelines. The ΔETSL and ΔETSR residuals in all three panels are positively correlated, and, comparing the panels, the high positive ΔOPI residuals are generally up and to the right of the low negative ΔOPI residuals, with the residuals around zero somewhere in between (although this may be somewhat hard to see), altogether indicating that gains on all modalities are positively interrelated. The scatter of the plots, especially for the moderate resΔOPIs, indicates that there is still considerable variance in gain to be explained.

In quantitative terms the correlation between gain resΔETSL and resΔETSR is .363, and the correlations between the ETS's and resΔOPI are .135 and .161, respectively, all higher than their raw counterparts. Looking at the joint distribution of the three residuals, all of the regressions of one on the other two are highly significant ($a < .001$). In a principal components analysis of the correlation matrix, the first component, a positively weighted average of the gains, has variance 1.456 which accounts for slightly less than half (48.5%) of the joint variance; i.e. the gains have a significant component in common. Neither of the two other components is significant (their variances are less than the residual gains individually). The components contrast the ETS exams with the OPI and the ETS exams with each other (ignoring the OPI); these are the patterns that might be further explored. In two separate analyses designed to check the assumptions above, one adding the expanded baseline to the calculation of the residuals, and the other examining preprogram OPI = 1 only, nearly identical numerical results were obtained:

> Gains in Listening, Reading, and Oral Proficiency tend to be associated: students who gain on one are likely to gain in the others.
>
> Nevertheless, the patterns of gain are varied, with many students gaining on one and not the other modalities.

6 Next steps

In this paper I have presented a comprehensive analysis of the data collected by ACTR over a period of many years on students studying Russian in what was then the Soviet Union. Three broad issues concerning study abroad, of interest to researchers and policymakers alike, were addressed using this unique and extensive database:

> How are preprogram skills and abilities interrelated among students like those who apply for and are accepted into programs like ACTR's;
>
> What are the determinants of gain on the three language modalities measured — oral, listening, and reading proficiency; and
>
> How are gains related to one another.

Each issue has strong implications for program selection, individualized counselling, and program design. The results are summarized at the ends of Sections 3 and 5, and for the determinants of gain, in Section 4.7, and need not be repeated here. Suffice to say that they establish in broad statistical terms the main outlines of the phenomena that now have to be further studied by close investigation of what students actually do during their sojourns abroad, linking their activities and experiences to the gains that they make. This is the rationale of the NFLC study of Language Learning during Study Abroad, which builds on the ACTR data, and which will be presented in subsequent papers in this series.

Table 1
Administration, Missing Data, and Valid Cases for Various Instruments, by Program Date

Program	Total Obs	Number Missing Ch'ng OPI	Ch'ng ETSL	Ch'ng ETSR	ACTR Qual†	MLAT	RD Rating
Spring '84	19	0	*	*	1	*	19
Fall '84	18	0	*	*	2	1	0
Spring '85	21	0	*	*	5	0	21
Fall '85	13	0	*	*	0	0	0
Spring '86	19	0	0	0	0	0	0
Fall '86	20	0	3	2	1	0	20
Spring '87	42	1	0	1	2	4	0
Fall '87	36	1	1	1	0	2	17
Spring '88	49	2	1	1	4	0	1
Fall '88	100	2	2	2	4	4	31
Spring '89	100	1	2	2	12	3	45
Fall '89	119	2	6	6	23	4	8
Spring '90	102	3	9	9	6	*	0
Valid Cases	658	646	563	563	598	519	496

* indicates that the instrument was not administered.

†The numbers refer to the composite score, QualGen. For analyses in which the qualifying grammar and reading tests are used as separate variables (thus using students before Spring 1990 only), numbers in the table give missing data for one or the other or both, but the base numbers are essentially the same.

Table 2
Descriptive Statistics for Pre Program Language Measures†

Variable	Median	Mean	s.d.	Range
ETS Listening (pre)	21	22.1	8.8	1-49
ETS Reading (pre)	19	19.7	8.1	2-45
Qualifying Grammar	642	636.2	141.3	156-987
Qualifying Reading	655	646.5	177.7	88-985
Qualifying General	640	630.0	144.6	187-975
MLAT3	32	32.8	8.9	6-50
MLAT4	32	31.2	5.6	10-44
MLAT5	23	21.1	4.1	4-24
MLAT SF (total)	86	85.2	13.0	29-115

† Bolded letters give abbreviations of variable names that appears in the analysis results below.

Table 3
Frequency Distributions for Preprogram Ratings on OPI and ETS Listening and Reading

Test	0	0+	1	1+	2	2+	3	Total
OPI	4	49	378	134	60	21	4	660
(%)	0.1	7.4	57.3	20.3	9.1	3.2	0.1	100
ETSL			482	63	18	7	8	578
(%)			83.4	10.9	3.1	1.2	1.4	100
ETSR			116	257	111	41	52	577
(%)			20.1	44.5	19.2	7.1	9.2	100

Table 4
Descriptive Statistics for Student Characteristics

Variable (label in bold)	Summary statistics	Number missing	Treatment of missing data
Age	median = 21; 75% bet. 20 and 22; range 17-33	0	
Gender	58% women	0	
Country of Birth	explored in preliminary runs but not used here	0	
Undergraduate College	explored as a background factor	3	
Major	Russian(59%), AreaSt(16%), Humanit's(12%),Other(13%)	12	imputed Russian
Degree	40% still ugrad; 49% BA; 11% MA or PhD	14	imputed as BA
High School Russian	75% none; a scattering above; recoded to 0/1	15	imputed as 0
College Russian	mode = 3: for frequencies see tables below	15	imputed as 3
Class hours	explored in preliminary runs but not used here	15	
Lab hours	explored in preliminary runs but not used here	15	
Slavic Languages	0/1 variable; 5% know one	19	imputed as 0
nonSlavic Languages	0 (11%); 1 (46%); 2 (31%); 3+ (12%)	13	imputed as 1
Previous USSR Immersions	0/1 variable; 25% have a previous immersion	15	imputed as 0
Institute	programs were held at 8 Institutes		

Table 5
OPI Scores Pre and Post
(count/row percent)

		Post OPI							
		0+	1	1+	2	2+	3	3+	Total
Pre OPI	0	1	3	0	0	0	0	0	4
		25.	75.	0	0	0	0	0	100
	0+	1	30	16	1	1	0	0	49
		2.04	61.2	32.7	2.04	2.04	0	0	100
	1	0	83	203	71	16	1	0	374
		0	22.2	54.3	19.0	4.28	0.27	0	100
	1+	0	8	48	48	29	1	0	134
		0	5.97	35.8	35.8	21.6	0.75	0	100
	2	0	0	7	33	18	2	0	60
		0	0	11.7	55.	30.	3.33	0	100
	2+	0	0	0	5	6	10	0	21
		0	0	0	23.8	28.6	47.6	0	100
	3	0	0	0	0	0	3	1	4
		0	0	0	0	0	75.	25.	100
	Total	2	124	274	158	70	17	1	646
		0.310	19.2	42.4	24.5	10.8	2.63	0.15	100

Table 6
Relationship of Years of College Russian to Preprogram OPI Score
(count/row percent)

		Pre OPI				
		0/0+	1	1+	2 and above	total
College Russian	2 or less	19	84	28	22	153
		12.4	54.9	18.3	14.4	100
	3	22	178	43	27	270
		8.15	65.9	15.9	10.0	100
	4	7	81	41	19	148
		4.73	54.7	27.7	12.8	100
	5 or more	3	28	18	16	65
		4.62	43.1	27.7	24.6	100
	total	51	371	130	84	636
		8.02	58.3	20.4	13.2	100

Table 7
Relationship of Student Characteristics and Preprogram Reading Proficiency to Gain in Listening Proficiency

	Model 1		Model 2	
Variable	Coeff	t	Coeff	t
Constant	26.7270	10.40	26.4273	11.10
ETSL1	-0.45396	-14.40	-0.69307	-18.30
Gender	1.10094	2.00	1.08075	2.13
Age	-0.48017	-4.03	-0.53897	-4.89
HSRuss	-0.84511	-1.27	-1.21239	-1.97
CollRuss	0.16610	0.51	-0.04947	-0.16
SlavicL	-0.32619	-0.28	-0.41620	-0.39
nonSlav	0.89972	2.72	0.74801	2.44
PrevImm	1.15666	1.76	1.15560	1.91
ETSR1			0.40028	9.81

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
30.9	29.9	554	.412	.403	552

Table 8
Relationships of MLATs and Reading Proficiency to Gain in Listening Proficiency

	Model 3		Model 4		Model 5	
Variable	Coeff	t	Coeff	t	Coeff	t
Constant	18.8100	4.84	17.7883	5.13	19.8915	6.07
ETSL1	-0.50394	-13.90	-0.50403	-13.90	-0.70387	-16.40
Gender	1.01521	1.69	1.06181	1.78	1.11071	1.98
Age	-0.37002	-2.72	-0.36184	-2.67	-0.45978	-3.59
HSRuss	-0.98562	-1.38	-0.96373	-1.35	-1.09234	-1.63
CollRuss	0.05962	0.17	0.07319	0.20	-0.06447	-0.19
SlavicL	-2.20085	-1.61	-2.20787	-1.62	-1.99665	-1.56
nonSlav	0.41200	1.14	0.39827	1.11	0.32503	0.96
PrevImm	1.38472	2.02	1.36855	2.00	1.45107	2.25
MLAT3	0.17259	4.83	0.16950	4.80	0.13175	3.92
MLAT4	0.10344	1.81	0.09895	1.75	0.06759	1.26
MLAT5	-0.04693	-0.59				
ETSR1					0.35853	7.64

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.357	.341	442	.356	.342	443	.432	.418	441

Table 9
Relationship of ACTR Qualifiying Exams and Reading Proficiency to Gain in Listening Proficiency (1)

	Model 6		Model 7	
Variable	Coeff	t	Coeff	t
Constant	24.0230	8.59	25.9608	9.77
ETSL1	-0.52229	-14.50	-0.68913	-17.10
Gender	1.49932	2.67	1.32455	2.49
Age	-0.55702	-4.36	-0.57510	-4.76
HSRuss	-1.21628	-1.83	-1.36478	-2.16
CollRuss	0.06578	0.20	-0.04743	-0.15
SlavicL	-0.93924	-0.80	-0.79314	-0.71
nonSlav	0.90369	2.67	0.79415	2.48
PrevImm	1.00250	1.49	0.93113	1.47
QualGen	0.00931	4.30	0.00252	1.13
ETSR1			0.36188	7.78

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.345	.333	503	.415	.404	501

Table 10
Relationship of ACTR Qualifiying Exams and Reading Proficiency to Gain in Listening Proficiency (2)

	Model 8		Model 9	
Variable	Coeff	t	Coeff	t
Constant	24.7642	7.62	26.0426	9.03
ETSL1	-0.54500	-13.30	-0.69334	-15.50
Gender	1.40468	2.29	1.38609	2.38
Age	-0.53762	-3.72	-0.56004	-4.11
HSRuss	-1.45060	-1.99	-1.21589	-1.76
CollRuss	-0.22612	-0.60	-0.18473	-0.51
SlavicL	-1.98355	-1.49	-1.85711	-1.47
nonSlav	0.86748	2.35	0.69118	1.97
PrevImm	1.23368	1.74	1.13325	1.68
QualGram	0.00260	0.86		
QualRead	0.00749	2.77	0.00269	1.27
ETSR1			0.36051	6.79

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.347	.331	415	.410	.396	415

Table 11
Model for Gain in Listening Proficiency with Only Significant Variables Included

	Model 10	
Variable	**Coeff**	**t**
Constant	22.1062	7.61
ETSL1	-0.69437	-16.50
Gender	1.13240	2.04
Age	-0.48846	-4.25
HSRuss	-1.10406	-1.68
SlavicL	-2.04223	-1.60
PrevImm	1.40194	2.23
ETSR1	0.36425	7.80
MLAT3	0.14506	4.45

R^2	adjR^2	d.f.
.429	.418	444

Table 12
Relationship of Student Characteristics and Preprogram Listening Proficiency to Gain in Reading Proficiency

	Model 1		Model 2	
Variable	Coeff	t	Coeff	t
Constant	12.5547	5.15	11.0640	4.54
ETSR1	-0.27910	-8.57	-0.35672	-8.50
Gender	0.43217	0.82	0.51260	0.98
Age	-0.19102	-1.66	-0.16072	-1.42
HSRuss	-0.54527	-0.85	-0.57423	-0.91
CollRuss	0.11408	0.36	0.02785	0.09
SlavicL	-0.56892	-0.51	-0.67296	-0.61
nonSlav	0.83250	2.61	0.79594	2.53
PrevImm	0.51177	0.82	0.31751	0.51
ETSL1			0.11795	3.03

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.137	.124	554	.149	.135	552

Table 13
Relationships of MLATs and Listening Proficiency to Gain in Reading Proficiency

	Model 3		Model 4		Model 5	
Variable	Coeff	t	Coeff	t	Coeff	t
Constant	-1.68801	-0.45	-0.96856	-0.28	-2.88915	-0.83
ETSR1	-0.35200	-9.18	-0.34481	-9.03	-0.41584	-8.74
Gender	0.68322	1.18	0.79607	1.39	0.89761	1.59
Age	-0.07682	-0.58	-0.09491	-0.72	-0.04609	-0.36
HSRuss	-0.54490	-0.79	-0.46921	-0.68	-0.54661	-0.81
CollRuss	0.25096	0.72	0.30864	0.89	0.175078	0.51
SlavicL	-0.65917	-0.50	-0.55399	-0.42	-0.68180	-0.53
nonSlav	0.24402	0.70	0.23797	0.68	0.18289	0.53
PrevImm	0.28910	0.44	0.22559	0.34	0.11940	0.18
MLAT3	0.11740	3.37				
MLAT4	0.24093	4.43				
MLAT5	0.10103	1.31				
MLAT SF			0.14890	6.37	0.15085	6.55
ETSL1					0.10976	2.57
	R^2	adjR^2	d.f.	R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
	.200	.180	442	.194	.177	444	.209	.191	442

Table 14
Relationship of ACTR Qualifiying Exams and Listening Proficiency to Gain in Reading Proficiency (1)

	Model 6		Model 7	
Variable	Coeff	t	Coeff	t
Constant	9.83357	3.60	9.00223	3.31
ETSR1	-0.37799	-9.31	-0.42342	-8.88
Gender	0.35927	0.65	0.46103	0.85
Age	-0.24618	-1.97	-0.22123	-1.79
HSRuss	-0.68594	-1.05	-0.68997	-1.07
CollRuss	0.00110	0.00	-0.05239	-0.16
SlavicL	-0.77545	-0.67	-0.78929	-0.69
nonSlav	1.06469	3.22	1.01559	3.09
PrevImm	0.45004	0.70	0.31414	0.48
QualGen	0.00980	4.31	0.00892	3.91
ETSL1			0.08330	2.02
	R^2 $adjR^2$ d.f.		R^2 $adjR^2$ d.f.	
	.172 .507 503		.175 .159 501	

Table 15
Relationship of ACTR Qualifiying Exams and Listening Proficiency to Gain in Reading Proficiency (2)

	Model 8		Model 9	
Variable	Coeff	t	Coeff	t
Constant	10.8567	3.44	10.7924	3.64
ETSR1	-0.41715	-8.77	-0.44169	-8.08
Gender	0.54602	0.90	0.68824	1.15
Age	-0.28247	-1.99	-0.26182	-1.86
HSRuss	-0.78915	-1.10	-0.72934	-1.03
CollRuss	-0.08721	-0.23	-0.16013	-0.43
SlavicL.	-1.15052	-0.88	-1.11684	-0.86
nonSlav	0.97873	2.69	0.92909	2.57
PrevImm	0.11990	0.17	0.03361	0.05
QualGram	0.00253	0.85		
QualRead	0.00869	3.14	0.00898	4.12
ETSL1			0.07042	1.53

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.177	.157	415	.172	.153	415

Table 16
Models for Gain in Reading Proficiency with Only Significant Variables Included

	Model 10		Model 11	
Variable	Coeff	t	Coeff	t
Constant	-3.39018	-1.65	-5.51360	-2.47
ETSR1	-0.41367	-8.85	-0.48236	-8.98
Gender	0.85530	1.54	0.78725	1.37
ETSL1	0.10968	2.61	0.07314	1.66
%MLATSF	0.15342	6.91	0.15581	6.73
%QualRead			0.00684	3.40

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.205	.198	448	.225	.215	408

Table 17
Relationship of Student Characteristics, Listening and Reading Proficiency to Gain in Oral Proficiency (0/1 Criterion)

	OLS		Probit	
Variable	Coeff	t	Coeff	t
Constant	152.063	4.30	568.268	4.35
OPI1	-0.24786	-10.83	-0.82580	-9.42
Gender	0.01646	0.46	0.04461	0.35
Age	-0.00053	-0.07	-0.00696	-0.26
HSRuss	0.08839	2.04	0.31025	1.98
CollRuss	-0.00898	-0.43	-0.01209	-0.16
SlavicL	0.06009	0.78	0.23283	0.82
nonSlav	0.02375	1.11	0.07619	1.00
PrevImm	0.02397	0.57	0.06386	0.43
ETSL1	0.00279	0.99	0.01112	1.13
ETSR1	0.00917	3.14	0.03476	3.25
Major	0.00793	0.49	0.03422	0.60
PgmDate	-0.07585	-4.27	-0.28459	-4.33

R^2 (OLS) = .236, d.f = 552

Log Likelihood (Probit) = -275.577

Table 18
Relationship of Student Characteristics, Listening and Reading Proficiency to Gain in Oral Proficiency (0/1/2 Criterion)

	OLS		Ordinal Probit	
Variable	Coeff	t	Coeff	t
Constant	272.934	5.01	509.546	4.98
OPI1	-0.38227	-10.84	-0.73267	-10.09
Gender	0.07262	1.32	0.12317	1.21
Age	0.00289	0.24	0.00390	0.17
HSRuss	0.12306	1.84	0.24553	1.99
CollRuss	-0.03310	-1.02	-0.05773	-0.96
SlavicL	0.06489	0.54	0.09050	0.41
nonSlav	0.04083	1.24	0.07995	1.32
PrevImm	0.01580	0.24	0.03616	0.31
ETSL1	0.00835	1.93	0.01494	1.88
ETSR1	0.01523	3.39	0.02835	3.35
Major	0.04590	1.85	0.09116	1.98
PgmDate	-0.13600	-4.97	-0.25535	-4.96

R^2 (OLS) = .257, d.f = 552

Log Likelihood (Probit) = -508.142

μ = 1.543 (t = 19.34)

Table 19
Model for Gain in Oral Proficiency with Only Significant Variables Included (0/1 Criterion)

	OLS		Probit	
Variable	Coeff	t	Coeff	t
Constant	127.934	3.46	488.006	3.56
OPI1	-0.25861	-10.31	-0.86840	-9.04
HSRuss	0.09324	2.18	0.32370	2.03
nonSlav	0.02616	1.19	0.08816	1.10
ETSL1	0.00235	0.82	0.00873	0.86
ETSR1	0.00763	2.38	0.02924	2.48
Major	0.01324	0.79	0.05490	0.90
PgmDate	-0.06378	-3.43	-0.24452	-3.55
QualGen	0.00031	1.84	0.00122	2.00

R^2 (OLS) = .231, d.f = 508
Log Likelihood (Probit) = -248.633

Table 20
Model for Gain in Oral Proficiency with Only Significant Variables Included (0/1/2 Criterion)

	OLS		Ordinal Probit	
Variable	Coeff	t	Coeff	t
Constant	226.3520	3.95	424.0150	3.94
OPI1	-0.40553	-10.44	-0.76746	-9.72
HSRuss	0.14042	2.12	0.273361	2.23
nonSlav	0.05106	1.49	0.09329	1.48
ETSL1	0.00738	1.66	0.01311	1.59
ETSR1	0.01184	2.39	0.02237	2.40
Major	0.05234	2.01	0.10336	2.13
PgmDate	-0.11265	-3.91	-0.21249	-3.93
QualGen	0.00056	2.17	0.00104	2.16

R^2 (OLS) = .251, d.f = 508
Log Likelihood (Probit) = -465.656 μ = 1.566 (t = 18.69)

Table 21
Model for Gain in Oral Proficiency
Preprogram OPI = 1+, 0/1 Criterion

	OLS	
Variable	Coeff	t
Constant	186.719	3.76
Gender	0.24261	2.84
HSRuss	0.15325	1.72
SlavicL	0.36988	1.99
nonSlav	0.09163	1.96
PgmDate	-0.09420	-3.77
%QualGen	0.00090	2.60

R^2	adjR^2	d.f.
.252	.212	111

Table 22
Relationship of RD Ratings to Gain in Listening Proficiency

	Model 1		Model 2	
Variable	Coeff	t	Coeff	t
Constant	1052.480	1.91	491.5260	0.70
ETSL1	-0.46297	-12.10	-0.62850	-12.70
Gender	1.24255	2.11	1.53555	2.38
Age	-0.51569	-3.84	-0.57062	-3.62
HSRuss	-0.39455	-0.55	-0.68183	-0.89
CollRuss	0.38899	1.11	0.23116	0.59
SlavicL	0.63885	0.51	-2.19129	-1.44
nonSlav	0.69710	1.92	0.25238	0.62
PrevImm	1.02267	1.45	0.94218	1.28
Major	0.59841	2.23	0.84929	2.74
PgmDate	-0.51743	-1.87	-0.23702	-0.67
NatAbil	1.13992	2.16	0.21698	0.37
CultAdap	0.82613	1.62	1.04913	1.86
Ldrshp	-1.57196	-4.30	-1.59241	-4.13
ETSR1			0.26749	4.67
MLAT3			0.07750	2.05
QualGen			0.00538	1.87

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.331	.311	434	.452	.422	292

Table 23
Relationship of RD Ratings to Gain in Reading Proficiency

	Model 1		Model 2	
Variable	Coeff	t	Coeff	t
Constant	75.5869	0.14	-321.998	-0.44
ETSR1	-0.28920	-7.36	-0.48787	-8.10
Gender	-0.00737	-0.01	0.57296	0.84
Age	-0.20849	-1.55	-0.14821	-0.88
HSRuss	-0.48566	-0.68	-0.56474	-0.70
CollRuss	0.03842	0.11	0.23247	0.55
SlavicL	-0.53588	-0.43	-1.29648	-0.81
nonSlav	0.66837	1.85	0.36800	0.85
PrevImm	1.23812	1.78	0.55470	0.72
Major	0.32040	1.20	0.65072	1.98
PgmDate	-0.03173	-0.11	0.15984	0.43
UseRuss	-1.00226	-1.97	-0.93686	-1.59
CultOpp	1.62336	2.95	1.88126	3.01
WrkInGp	-0.49445	-1.51	-0.73493	-1.91
ETSL1			0.10274	1.97
MLATSF			0.11356	4.03
QualGen			0.00920	3.11

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.166	.141	433	.268	.228	292

Table 24
Relationship of RD Ratings to Gain in Oral Proficiency (0/1/2 Criterion)

	Model 1		Model 2	
Variable	Coeff	t	Coeff	t
Constant	280.476	7.02	170.339	2.86
OPI1	-0.38460	-11.70	-0.47602	-11.10
Gender	0.04218	0.74	0.03457	0.56
Age	0.01256	1.01	0.00335	0.23
HSRuss	0.12837	1.83	0.10827	1.45
CollRuss	-0.00143	-0.04	-0.00632	-0.17
SlavicL	0.09215	0.73	0.10824	0.82
nonSlav	0.02481	0.71	0.03918	1.02
PrevImm	0.01737	0.26	0.03653	0.50
Major	0.04228	1.59	0.05542	1.91
PgmDate	-0.14065	-7.00	-0.08527	-2.85
NatAbil	0.20575	4.03	0.17466	3.05
CultAdap	0.12799	2.55	0.15789	2.86
WrkInGp	-0.17099	-4.09	-0.16705	-3.81
Ldrshp	0.08264	1.97	0.03861	0.84
ETSL1			0.00251	0.52
ETSR1			0.01036	1.92
QualGen			0.00061	2.12

R^2	adjR^2	d.f.	R^2	adjR^2	d.f.
.286	.265	472	.313	.283	393

Table 25
Summary of Factors Affecting Gains on Three Language Modalities

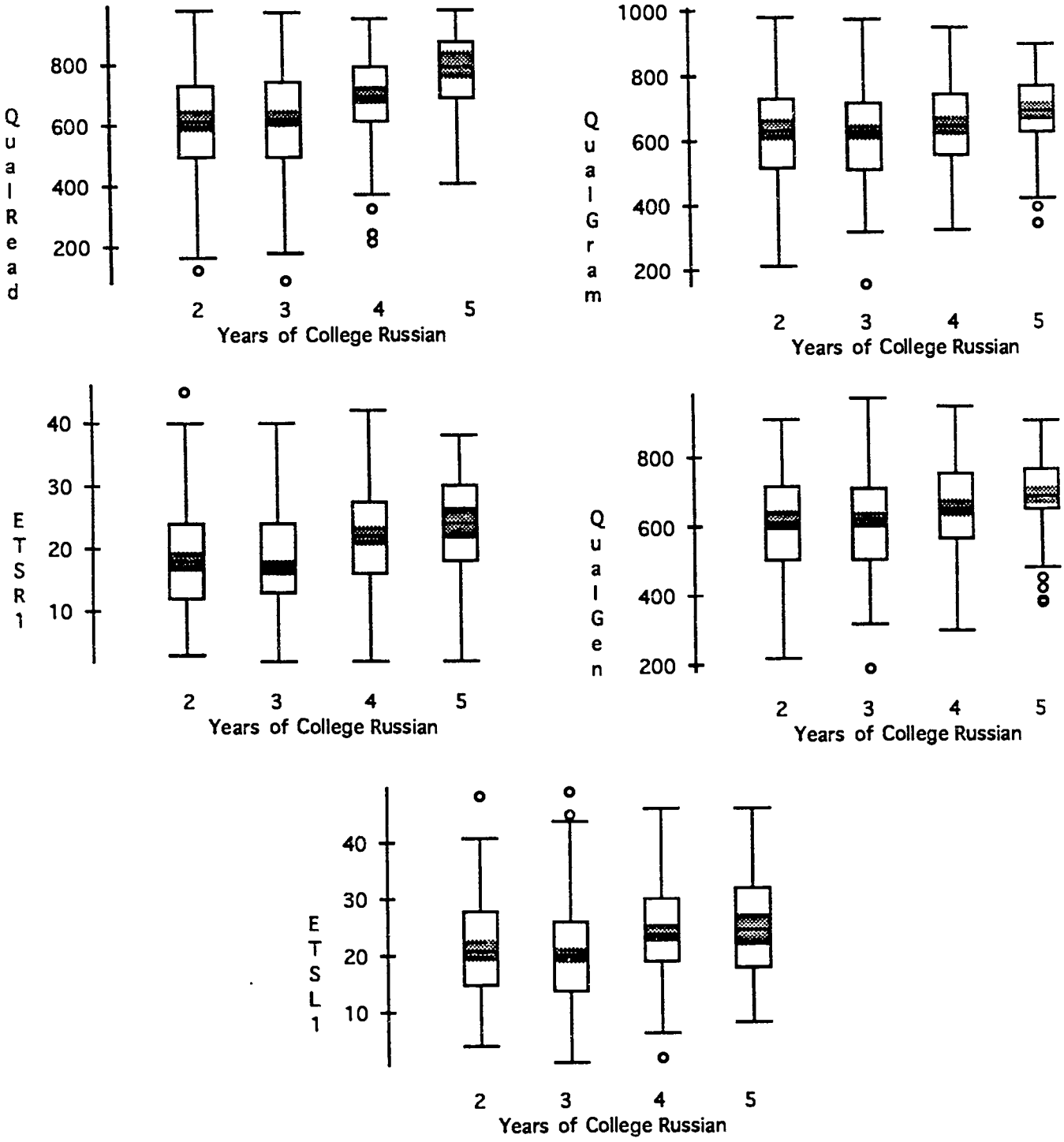
	Gain Modality				
Variable	ETS Listening	ETS Reading	OPI(0/1)	OPI (0/1/2)	OPI 1+ to 2 ↑
Gender	2.04				2.84
Age	-4.25				
HSRuss	-1.68		2.03	2.23	1.72
CollRuss					
SlavicL	-1.60*				1.99
nonSlav	*	*		1.48	1.96
PrevImm	2.23				
Major				2.13	
PgmDate			-3.55	-3.93	
MLAT3	4.45	*			
MLAT4	*	*			
MLAT5					
MLATSF		6.73			
QualGram					
QualRead	*	3.40			
QualGen	*		2.00	2.16	2.6
ETSL1	-16.5	1.66			
ETSR1	7.80	-8.98	2.48	2.40	
OPI1			-9.04	-9.72	
RD Ratings	NatAbl: * UseRus:-1.59 CulAdap:1.86 CulOpp: 3.01 Ldrshp: 4.13 WkGrp:-1.91		same as 0/1/2 criterion	NatAbl: 3.05 CulAdap:2.86 WkGrp: -3.81	not examined

* See text for qualification of comment.

Numbers are t statistics from "good" models.

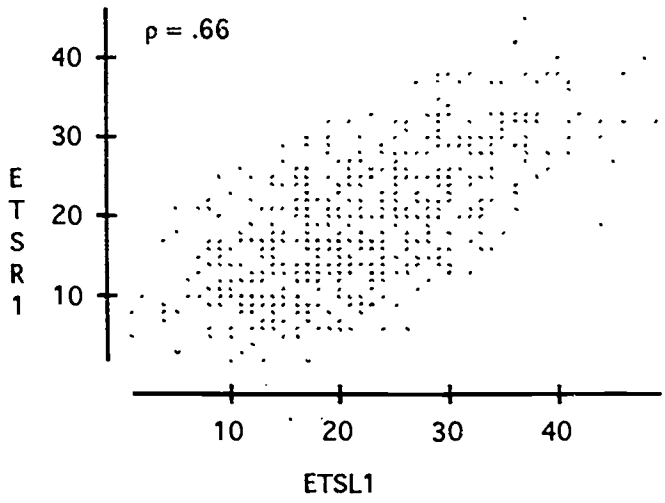
Figure 1

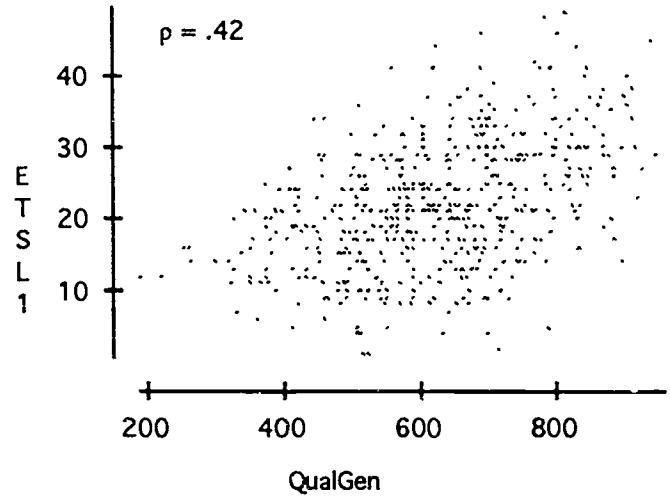
Relationship of Years of College Russian to Language Measures

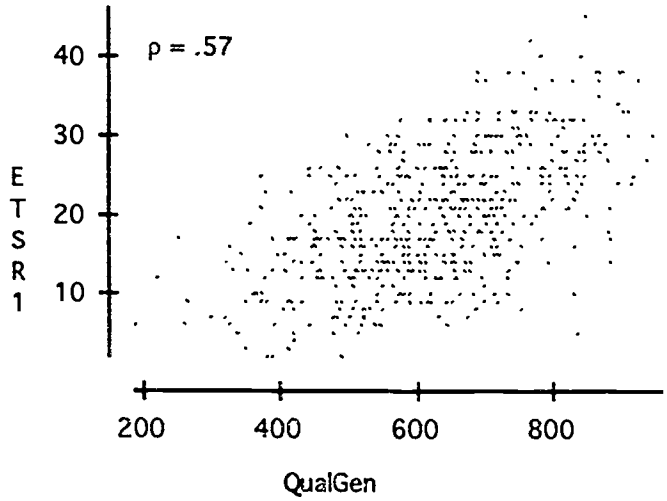


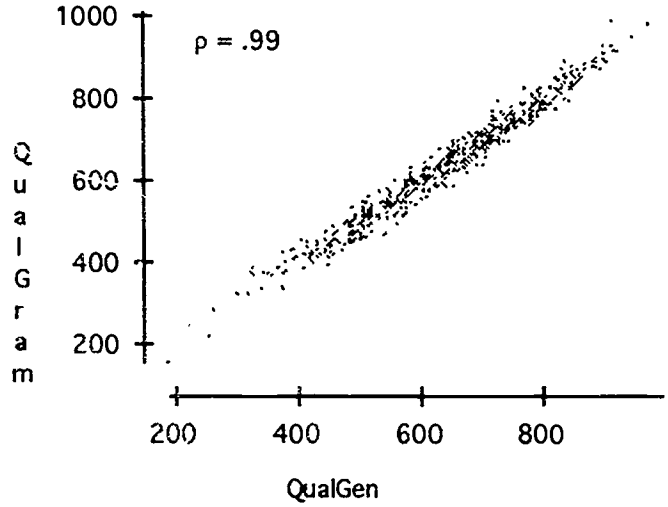
Note: Number of observations on which each boxplot is based differs for each instrument. See Table 1.

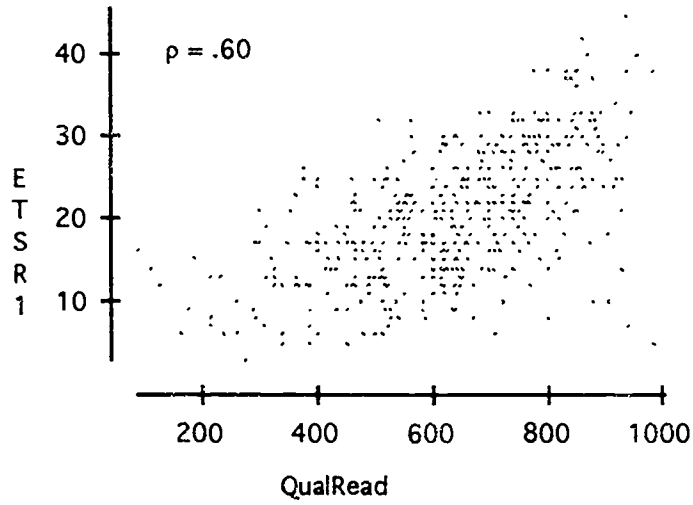
Figure 2
Scatterplots of Qualifying Exams and ETS Proficiency Tests











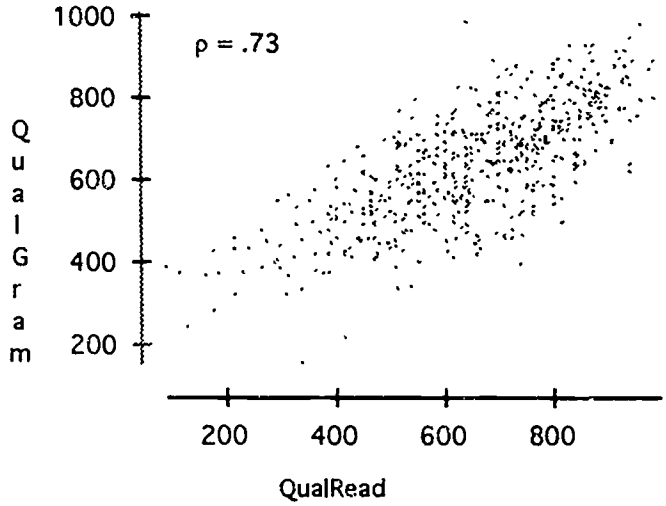
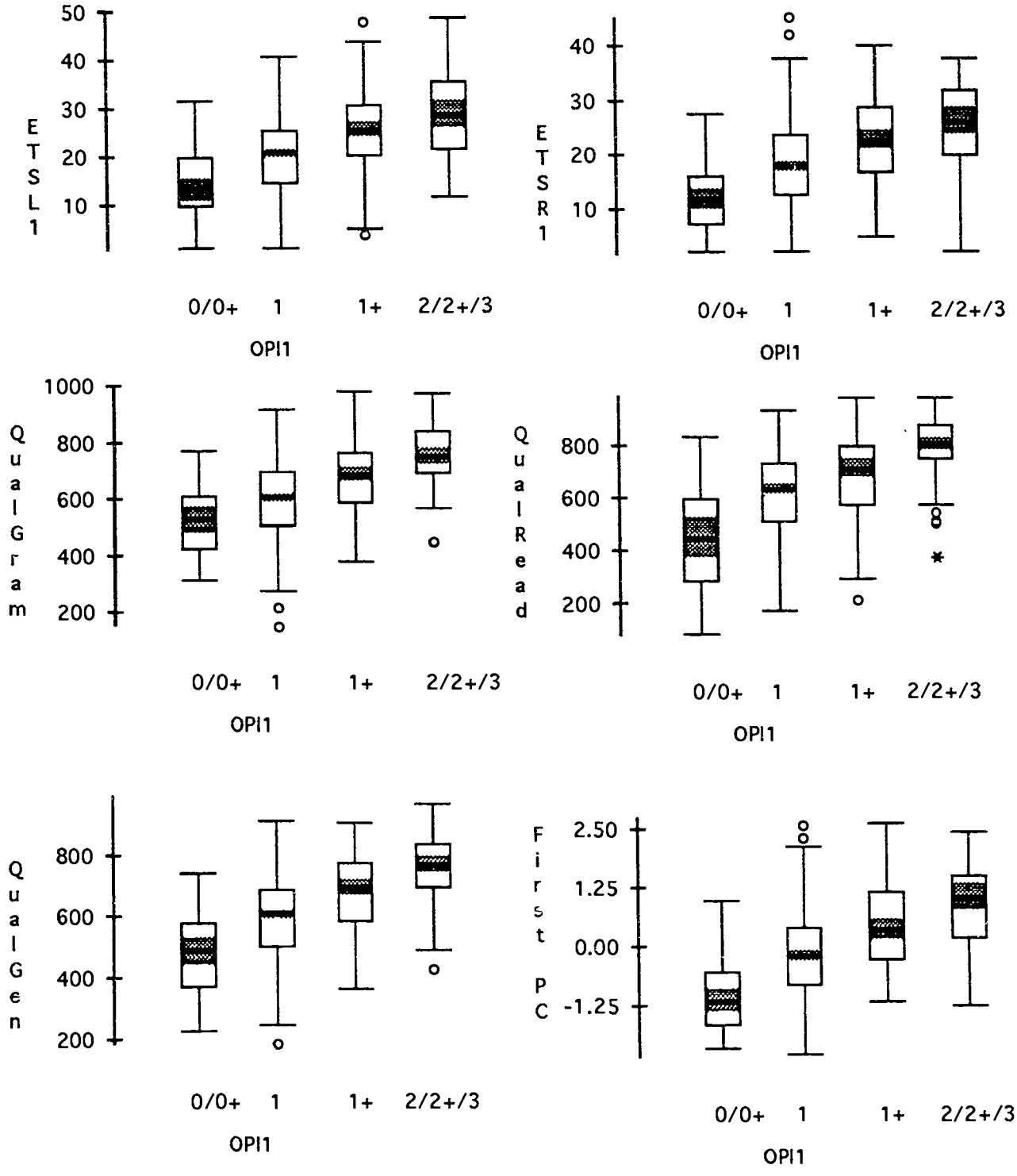
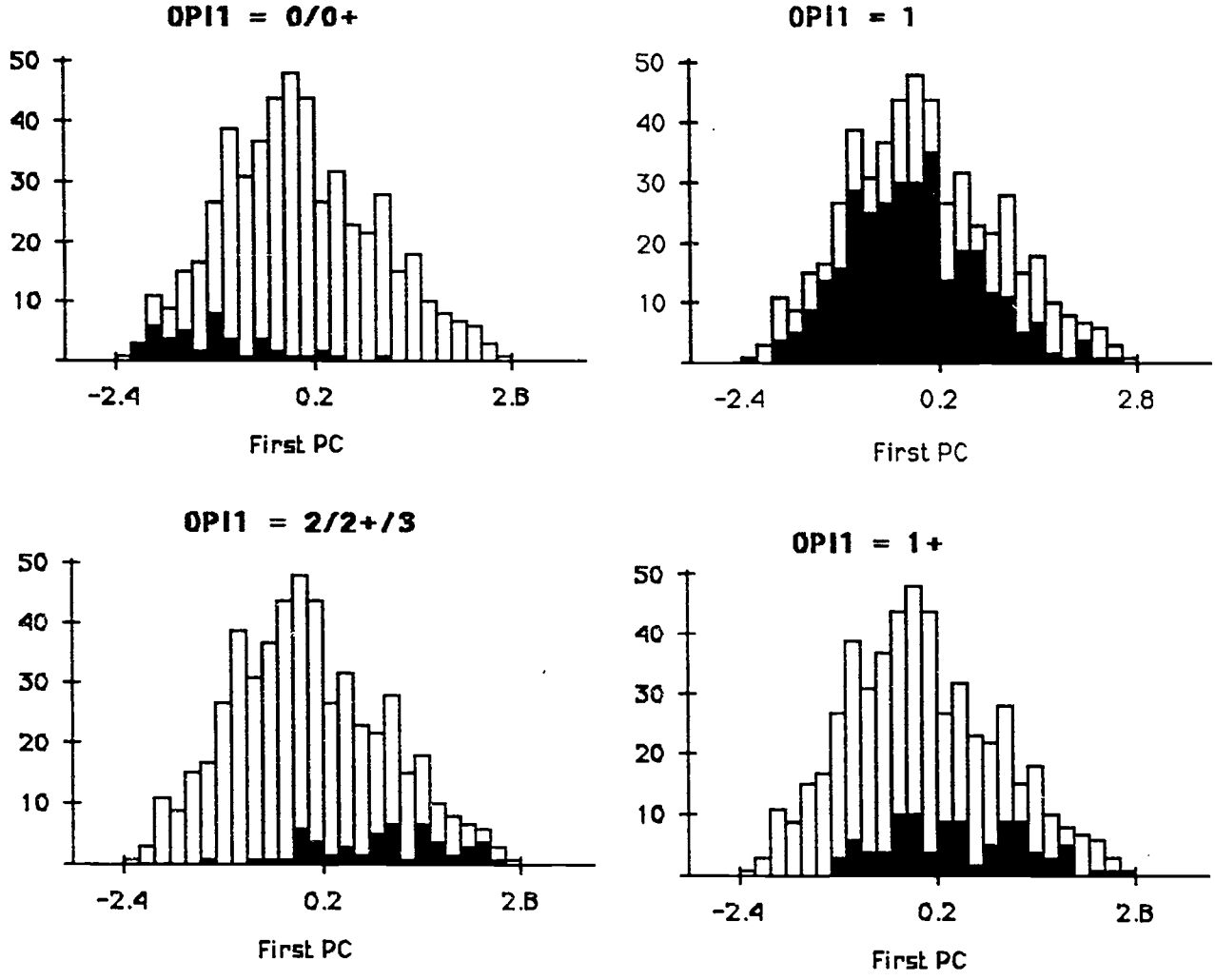


Figure 3
Relationship of OPI to ETS Listening and Reading, Qualifying Exams, and First Principal Component



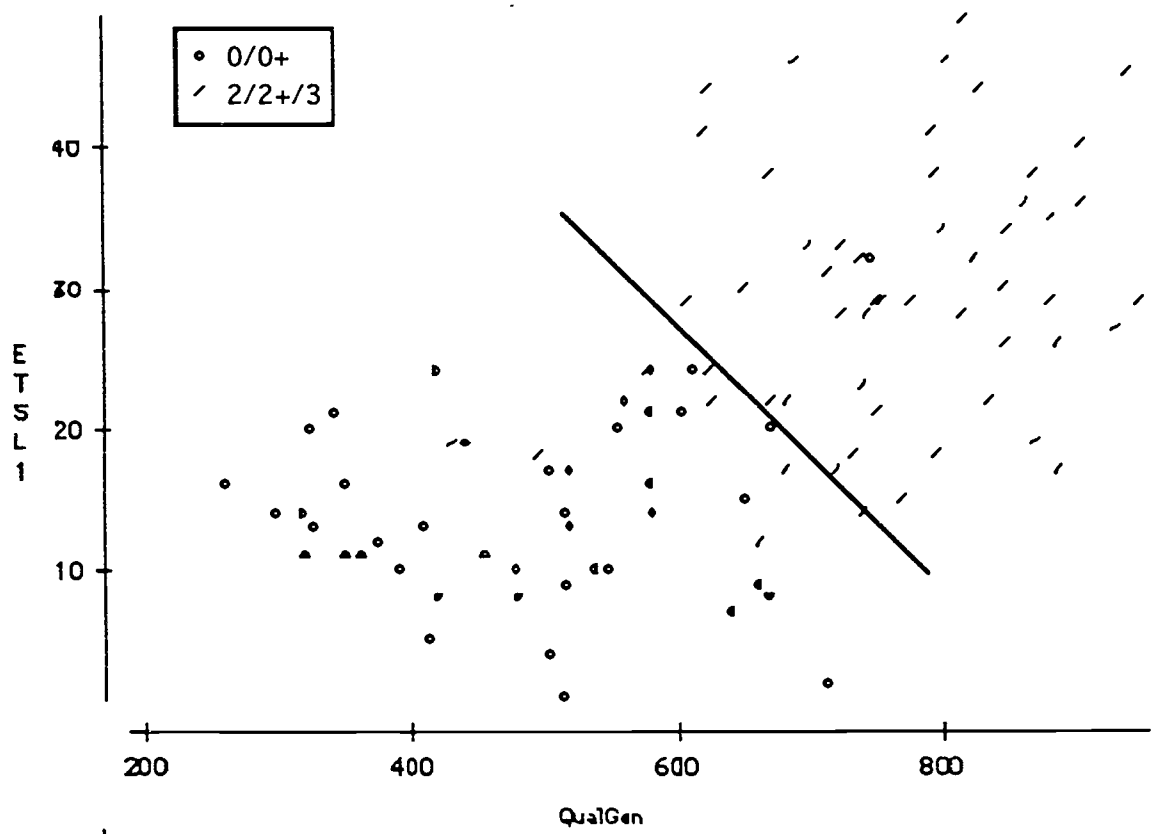
Note: * indicates an extreme outlier

Figure 4
Histograms of OPI1 Groups on First Principal Component of ETSL1, ETSR1, and QualGen



Note: The large histogram in each panel is the whole distribution; the blackened part is the histogram for that group. The worst and the best are quite well discriminated on this one criterion, but the other groups are not well discriminated.

Figure 5
Scatterplot of Pre ETS Listening and QualGen
with OPI1 Levels Distinguished



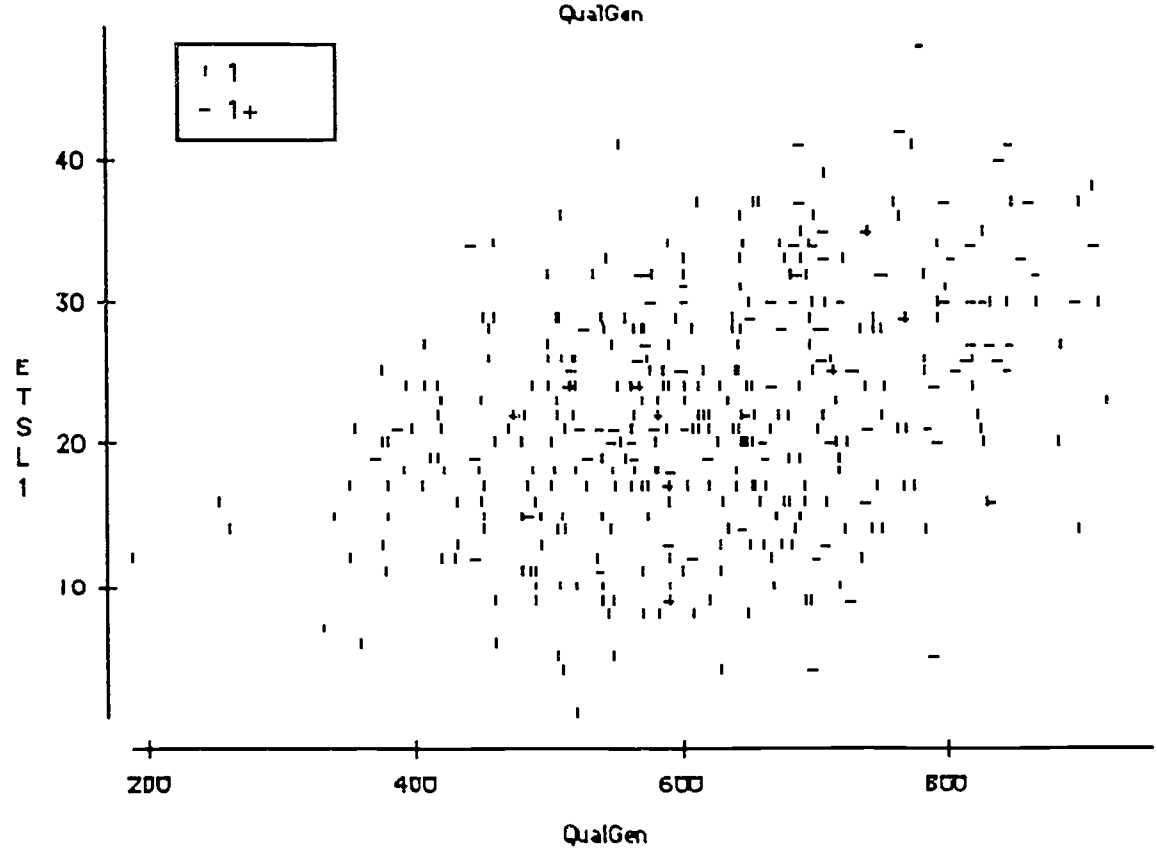
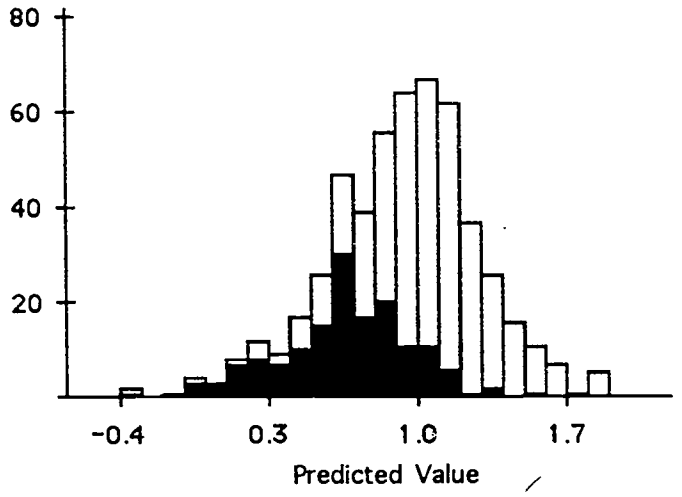
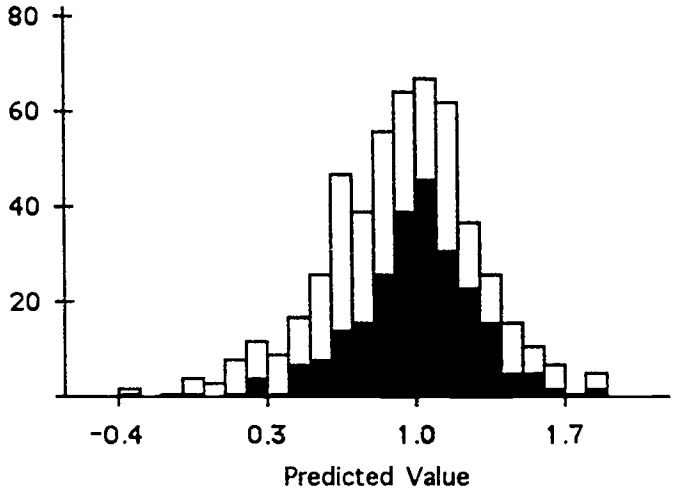


Figure 6
Histograms of Predicted Values from "Good" Equation
for Change in OPI, 0/1/2 Criterion

ΔOPI = 0 or loss



ΔOPI = 1



ΔOPI = 2 and up

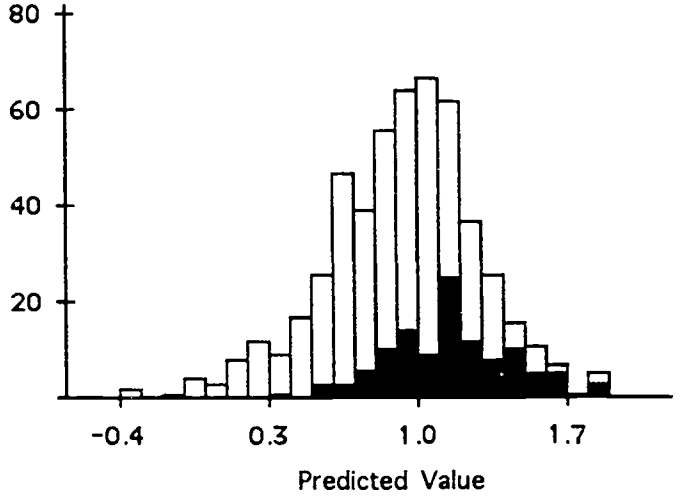
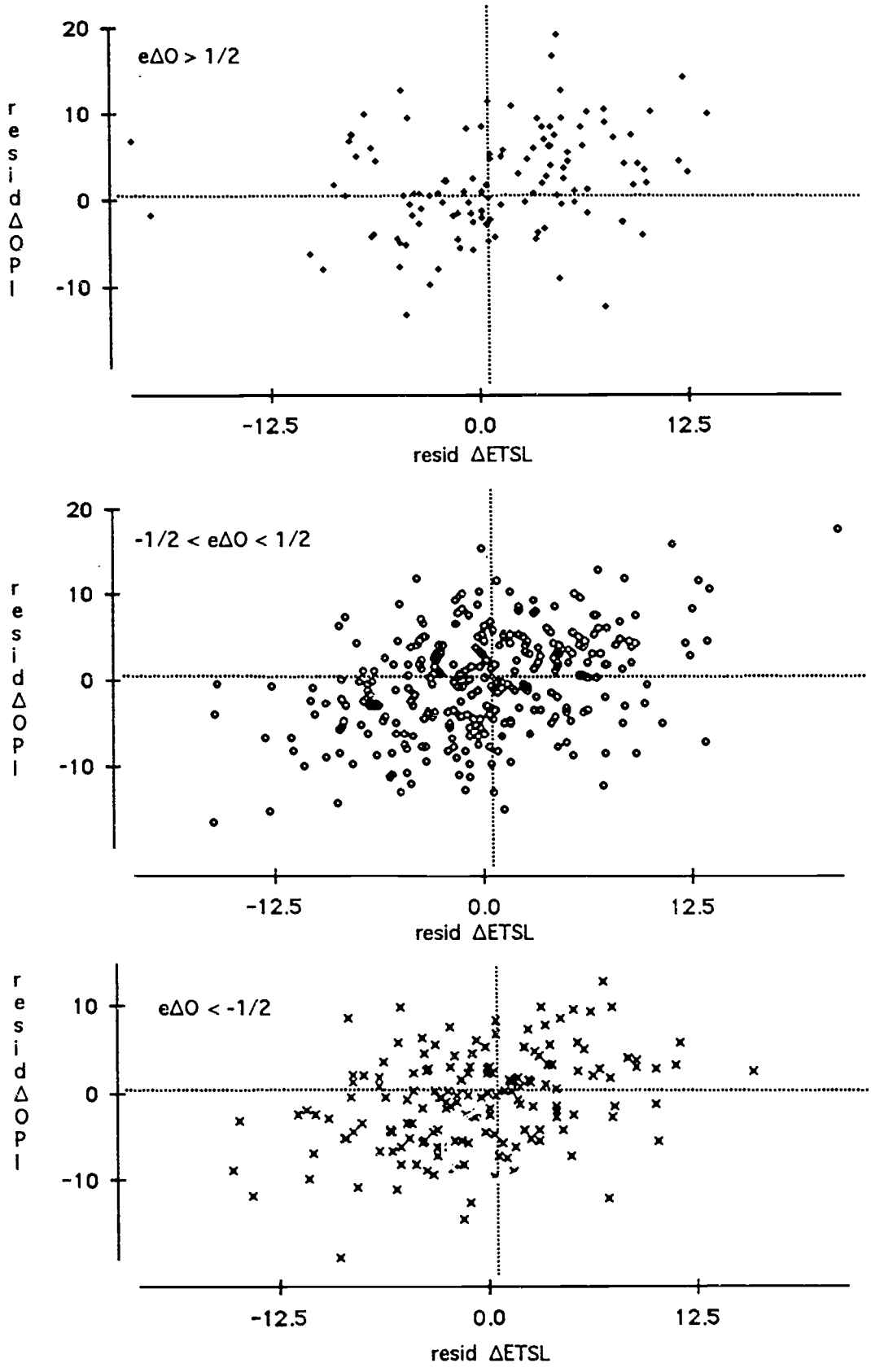


Figure 7
Reading Gain Residuals v. Listening Gain Residuals
for Three Groups of OPI Gain Residuals*



*Residuals computed with preprogram Listening, Reading, and OPI removed.

The National Foreign Language Center
at the Johns Hopkins University
1619 Massachusetts Avenue NW
Suite 400
Washington DC 20036
Telephone 202/667-8100