

EMPATHY AND SECOND LANGUAGE LEARNING¹

Alexander Z. Guiora, Robert C. L. Brannon, Cecelia Y. Dull²
The University of Michigan

The study reported here represents the culmination of the development and application of a research strategy whereby hypotheses generated in the clinical circumstance are transposed to a behavioral realm where more rigorous, reproducible, reliable, and valid experimentation is feasible (Guiora 1970). Our concern has chiefly been to apply this strategy to the concept of empathy, choosing as the transposed realm of behavior, language, in particular, authenticity of pronunciation of a second language. In a series of studies we investigated the hypothesis that empathy plays a significant role in the ability to authentically pronounce a second language. The measure which proved to be most successful in predicting authenticity of pronunciation was the Micro-Momentary Expression (MME) test as modified by us. The present study confirms the original hypothesis that empathy as measured by the MME is positively related to the ability to authentically pronounce a second language. Essentially the MME measure coupled with the Verbal Mental Reasoning test of intelligence and a simple but apparently effective measure of motivation provide, we believe, a major contribution to the prediction of pronunciation ability. Adding the Army Language Aptitude Test as a linguistic measure, the combined instruments constitute a powerful predictive battery.

Research on second language learning abilities, developing quite naturally out of practical concerns which arise in the classroom situation, has frequently ignored the role of the more subtle psychological processes involved. Viewing second language learning in a real life context, however, reminds us that for people who geographically exchange one culture for another, the task of learning a second language poses a challenge to the integrity of basic iden-

¹The research reported here was performed pursuant to Contract No. DAHC 15 70 C 0239 with the Department of Defense, Defense Language Institute, awarded to the senior author. A complete report of this research, including tables and a detailed description of instruments is available on request from: Defense Language Institute Headquarters, R & S Division, Anacostia Naval Annex, Washington, D.C. The reference is: *Final Report*, Contract No. DAHC 15 70 C 0239, *The Role of Personality Variables in Second Language Behavior*, Alexander Z. Guiora, Project Director, 1971.

²In a very real sense the research reported here is the result of a collective effort. While the authors, listed on the top of the page, carried the study through from its inception to its reporting, the following have all contributed, at different times and in different ways to this endeavor; to all of them we owe a debt of gratitude: Benjamin Beit-Hallahmi, Robert K. Bolin, John C. Catford, James Dew, Ronna Hoy, Neil Kalter, Susumu Nagara, and Thomas Scovel. Robert C. L. Brannon is now at Brooklyn College, City University of New York.

tifications. To engage in learning a second language is to step into a new world. This act of extending the self so as to take on a new identity is, we believe, an important factor in second language learning.

The psychological demands of the language learning situation, in relation to the individual's identity, depend in part upon the particular aspect or aspects of language learning which is emphasized. We would suggest that among the four major aspects of language behavior: reading, writing, understanding, and speaking, psychologically the most demanding is speaking. Speaking a foreign language entails the radical operation of learning and manipulating a new grammar and at the extreme limits of proficiency, modifying one of the basic modes of identification by the self and others, the way we sound.

We would say then that of the skills involved in proficiency in speaking a foreign language, authenticity of pronunciation is psychologically the most critical. As has been stated elsewhere, (Guiora et al. 1969) "language behavior is a unique and complex attribute of man, not only in the evolutionary sense, but in the developmental psychological history of each individual. Language behavior arises and evolves within the context of a more general psychological growth. It is reasonable to speculate that even certain structural aspects of language are in part shaped by and express the broader personality context from which they have emerged." Thus it is only by regarding pronunciation ability within this context that we may begin to understand not only the course of its development but the processes involved in that development.

To mediate the postulated process, a new construct, *language ego*, has been introduced by Guiora (1972). In a manner similar to the concept of body ego, language ego too is conceived as a maturation concept and refers to a self-representation with physical outlines and firm boundaries. Grammar and syntax are the solid structures on which speech hangs, lexis the flesh that gives it body, and pronunciation its very core. Thus pronunciation is the most salient aspect of the language ego, the hardest to penetrate (to acquire in a new language), the most difficult to lose (in one's own).

The permeability of the language ego boundaries, specifically the flexibility of the pronunciation boundaries is developmentally and genetically (in the psychoanalytic sense) determined. That is to say, pronunciation permeability will correspond to stages in the development of the ego; in the early formative stages of general ego development greater flexibility is allowed. Thus a child can assimilate native-like speech in any language. Once ego development is concluded, flexibility will be sharply restricted forever.

In summary, it would appear that second language learning in all of its aspects demands that the individual, to a certain extent, take on a new identity. The last step for the completion of this transformation is pronunciation. Since pronunciation appears to be the feature of language behavior most resistant to change, we are led to suspect that it is probably the most critical to the individual's identity.

Individual differences in the ability to approximate native-like pronunciation should reflect individual differences in the flexibility of psychic processes or, as we have chosen to conceptualize this ability, empathic capacity.³ Going beyond this basic hypothesis we propose that empathic capacity is related not only to pronunciation ability, but also in yet to be determined ways, to the overall capacity to acquire a second language, a new system of communication.

Our previous research has encouraged us to believe that empathic capacity could be operationalized and thus subjected to quantitative analysis and that there was indeed evidence of a relationship between empathic capacity and the ability to approximate native-like pronunciation of a foreign language (Guiora et al. 1967, Taylor et al. 1969). The present study created the conditions to partially replicate and greatly expand the original research.

REVIEW OF SECOND LANGUAGE LEARNING AND EMPATHY

Second Language Learning

The obvious point of departure in the search for predictors of second language achievement is the concept of language aptitude. Carroll (1963) has stated that language aptitude consists of four abilities: phonetic coding, grammatical sensitivity, rote memorization and inductive language learning ability. Other researchers have taken a broader approach to the notion of aptitude and have included as predictors such nonlinguistic variables as age, motivation or interest, as well as personality traits. Of particular interest in the present study were the factors of intelligence, motivation and personality variables.

Intelligence has figured as a factor in many studies of language learning. Carroll (1963) has stressed that intelligence alone cannot account for second language learning. He further emphasized (Carroll

³"Empathy is a process of comprehending in which a temporary fusion of self-object boundaries, as in the earliest pattern of object relation, permits an immediate emotional apprehension of the affective experience of another, this sensing being used by the cognitive functions to gain understanding of the other" (Guiora 1965: 780-782).

1965) that many of the studies which show high correlation between intelligence and language learning base the criteria of language success on the traditional grammar-translation model which is heavily dependent on cognitive skills (Carroll 1965). Gardner and Lambert (1965) in a factor analysis of a battery of language aptitude and intelligence tests found the factors of language aptitude and achievement independent of the intelligence factor. One further point is that with regard to pronunciation, the Modern Language Association (1967) has cited evidence that pronunciation ability is unrelated to intelligence.

On the other hand, Pimsleur, Mosberg and Morrison (1962) presented a number of studies supporting the relationship of grades in language courses and various tests of intelligence. They reported correlations from .21 to .65 when considering the relationship between general intelligence and second language learning. In addition, Smith and Berger (1968) found intelligence related to grammar, vocabulary and reading skills.

The conclusion to be drawn from these studies is that intelligence is indeed positively related to second language learning but the extent of that relationship is rather low and given to variation. As Carroll (1965) has pointed out, cutting points for intellectual ability would give some assurance of success, but wide variations remain in the ability of those of equally high intelligence.

Pimsleur (1968) has included motivation as one of three components of the ability to learn foreign languages. He pointed out, however, that motivation or interest is extremely unstable. It is not clear that high motivation toward learning a foreign language will remain high nor that initial lack of motivation cannot be stimulated. Obviously important situational variables could presumably alter either high or low motivation. In a review of the educative factors in language learning Pimsleur, Mosberg and Morrison (1962) concluded that interest in language in general or in the particular languages being studied correlate positively with achievement. This finding, while modest, is a consistent one.

In a study of under achievement in foreign language learning, Pimsleur, Sundland and McIntyre (1963) found that intelligence and motivation are the two main determiners of success in foreign language learning. They attempted to differentiate integrative motivation based on an interest in assimilating the new language from instrumental motivation based on the desire to succeed in an academic endeavor.

Carroll (1963) has added an important note, however, regarding controlled research situations. He stated that when all subjects are exposed to the same amount of material, motivation is not really

relevant. Motivation is more important when students are free to determine how long they will persevere.

There has been some research and conjecture on personality differences between superior and inferior language learners. In factor analytic research of language aptitude and performance, Gardner and Lambert (1965) found personality related measures associated with language aptitude. In one study a factor was discovered which consisted of loadings on authoritarianism, anomie, ethnocentrism, and preference for America. Individuals who scored high on these measures tended to have poor language performance with course grades and standardized tests as criteria of performance. In another study, a factor was isolated which was loaded by measures of high parental discipline and a low reported evaluation of the subject's mother. Again this was found to be related to lower achievement in language learning. It appears that interest and motivational factors toward the language group of the language being studied are very important in defining the course of language learning. (Anisfeld and Lambert, 1961).

Many teachers of a second language point to the subjective experience of feeling strange and uncomfortable in producing new sounds and making awkward new mouth movements (Pike 1947). This awkwardness may be a much greater problem for some, extreme enough to create feelings of loss of identity. Several theorists point to underlying contempt for a new and strange language and suggest that the inability to speak authentically is often the result of a resistance (Nida 1957-1958, Stengel 1939). The feeling of awkwardness in speaking may create for some a fear of failure which prevents them from learning to speak (Nida 1957-1958). Being in a position of learning the new language may lead to reexperiencing the situation of the helpless state of infancy which awakens early anxieties associated with that defenseless state (Stengel 1939).

Empathy

The personality variable most relevant to the present study is that of empathy. While social perceptiveness or sensitivity, the ability to empathize with the thoughts and feelings of others (or to "put yourself in their shoes") has often been hailed as a critical ability for social existence, it has proven very difficult to devise satisfactory ways to measure this ability. A number of different approaches to quantitative measurement has been proposed in previous research, with varying degrees of apparent success.

Apparently the earliest attempt to measure empathy was Dymond's (1948) proposal to score TAT responses for this ability. These were coded for "richness of description" of each characters'

thoughts and feelings, an index of role-taking activity presumably related to ability to adopt the viewpoint of another. A later study (Dymond, Hughes, & Raabe 1952) extended this measure to children by asking for stories about a series of pictures. Empathy was taken to be an inverse function of the amount of "prodding" it took before a child would describe the thoughts and feelings of the various characters.

Feffer's procedure (1959) used a series of pictures which suggested one obvious story but indicated a very different story when some of the key pictures were removed. Children were rated as to how well they could tell the story which someone seeing only the second sequence would tell.

A rather unique approach to the measurement of empathy in terms of interpersonal predictive accuracy is found in Mahoney's *Literature Empathy Test* (Mahoney 1960). The test contains four selections from fiction portraying markedly different characters. After "getting a feel" for the individuals portrayed, the subjects are asked to predict 20 multiple choice incomplete sentence responses as they think the fictitious characters would. The criterion is the judgments made by 13 psychologists considered "empathic individuals."

Another and decidedly different approach to the measurement of empathy, less influenced by the factor of knowledge of group characteristics than that discussed above, has developed directly out of a concern with the clinical process.

Wallach and Strupp (1960) assessed the empathy of psychotherapists by showing a film of a patient being interviewed, stopping at 28 selected places and asking "what would you do now?" The answers were evaluated by independent raters, who judged whether the therapist "recognized the existence of an emotional problem and communicated effectively with the patient."

Truax (1963) used a similar technique based on tape recordings, drawn from actual therapeutic sessions of four patients who had improved and four who deteriorated. Judges listened to the tapes and then devised "therapeutic communications;" these were examined for "sensitivity to current feelings and communication in effective terms," by raters using criteria Truax provided.

While the clinical research model, utilized in the two studies mentioned above, is conceptually appealing, it is applicable only to well trained therapists and ties empathy to the ability to express such understanding effectively. Our initial investigations into the relationship between empathy and authenticity of pronunciation led us to search for an empathy measure which could be used with naive subjects in a non-clinical setting.

Research in psychotherapy by Haggard and Isaacs (1966) provided an intriguing technique which seems to lend itself to measurement of the ability to be aware of subtle cues of another's affective state. They found that showing motion pictures of patients in psychiatric interviews at slower than normal speeds allowed observers to see facial expressions suggestive of intense feelings which were not apparent at normal film speed.

The use of facial expressions to measure correct identification of emotions is not new to psychology. Research using pictures of faces assuming certain emotional expressions (Feleky 1914, Frois-Whittmann 1930, Guilford 1929-1930, and Ruckmick 1921) produced very little in the way of conclusive results. Researchers generally agreed that knowledge of the situation markedly improves identifying the emotion, that certain expressions are more easily identified than others, and that training significantly increases accurate judgment. Subsequent investigation of problems in the labeling and identifying of affective states (Osgood 1966, Schlosberg 1954, and Tomkins and Izard 1965) shed light on some of the difficulties. The correct labeling of emotions is an all but impossible task due to the differences in meaning associated with such labels and the generic rather than specific affective states which can be differentiated.

In 1967, Guiora, Lane and Bosworth performed a pilot study using a modification of Haggard and Issacs' (1966) technique, referred to as Micro-Momentary Expression (MME) as a potential measure of empathy. The MME consists of silent film clips shown at various speeds during which subjects are asked to indicate each observed change in facial expression. This initial attempt suggested that even in the absence of sensitive equipment for the recording of responses, a significant relationship was found between MME scores and authenticity of pronunciation by 14 teachers of French (.60 Spearman Rho). This result prompted further development of the MME as a measure of empathy and led to a more refined experiment.

In the second experiment, (Taylor, Guiora, Catford, and Lane 1969) it was proposed that the ability to speak a second language authentically or like a native speaker was related to an individual's sensitivity to cues in interpersonal situations—his empathic capacity. A test of empathy, the MME described above, plus additional control tests were administered to a group of 28 subjects. Following these tests, the subjects learned basic conversations in Japanese in four one-hour sessions. Their pronunciation was then rated by native Japanese speakers on general authenticity and on specific phonetic details for five spontaneous sentences and five sentences

repeated after the instructor. The results indicated differences in speaking skills which were related to two clusters of variables representing independent personality characteristics. An empathy dimension was defined as four variables (tolerance to anxiety, intelligence, involvement in emotional experiences, and perception of emotional expression) and was related to correct pronunciation of specific details in spontaneous sentences ($R = +.72$). An intuition dimension consisting of three variables (tolerance to anxiety, intelligence and psychological mindedness) was significantly related to general authenticity of pronunciation of repeated sentences ($R = +.72$). Each cluster was shown to be a significant predictor of skill in second language pronunciation.

EXPERIMENTAL DESIGN

Data Source

The subjects for this study were students at the Defense Language Institute (DLI), both east (DLIEC) and west coast (DLIWC) branches, engaged in an intensive three month course in one of five languages: Japanese, Chinese-Mandarin, Thai, Spanish and Russian. The original sample size based on subjects targeted for complete testing was 594 subjects. The sample for which complete data were available, however, was 411, a reduction of about 30% which is well within the expected limits. Attrition was due to several presumed but essentially uncontrolled sources; that is, for those subjects for whom only partial data was available we are unable to distinguish between those who were absent from part or all of one of the two testing phases due to "flunking out" of the program or being relieved for academic reasons from those who missed testing sessions for other reasons.

Data analysis was based on the pooling of all subjects either as a single group or by language. No attempt was made to distinguish between DLIEC and DLIWC students, the assumption being that both were samples from the same population.

Materials

Micro-Momentary Expression Test (MME). This is a test of ability to attend and be receptive to interpersonal cues of affective states. It consists of three short film segments showing a young woman from the shoulders up, engaged in conversation. The segments are projected first at normal speech (24 f.p.s.), then at extremely slow speed (4 f.p.s.). Subjects viewing the film are each equipped with a push-button type switch, connected via cable to a

central recording unit. Recording of responses is made by an Esterline-Angus Chart Drive Operation Recorder. Subjects are instructed to press the button-switch each time a change in facial expression of the projected image is perceived (no more than this is stated, to allow individually operationalized measures of what constitutes a "change in facial expression"). Five scores were developed based on MME responses to single and combined film segments at both normal and slow speed. Referring to each of the three segments as A, B and C, and normal and slow (S) speeds, the following scoring system was used: $A + B + C$ (24 f.p.s.); C (24 f.p.s.); $SA + SB + SC + (4 \text{ f.p.s.})$; SC (4 f.p.s.); $\text{Total} = A + B + C + SA + SB + SC$.

Thematic Apperception Test (TAT). This is a measure of empathic activity as reflected in response production to three standard projective test stimuli presented in booklet form. Instructions to subjects were the standard instructions commonly employed for the Thematic Apperception Test: "Tell what might be going on in the picture, and what the person is probably thinking and feeling." Responses to the measure were scored by a scheme specific to this research—the "richness" of description of each set of responses was quantitatively coded on a 1-5 scale by at least two professional psychologists.

Photographic Perception Test (PPT). This is a measure of empathic activity, designed to extend and supplement the Thematic Apperception Test. Three finely-detailed photographs were provided in booklet form. Instructions were to "write down everything you can" about the person's personality—including ideas, feelings, and emotions. Responses were scored for richness of description by at least two professional psychologists. For each of the three pictures, three scores were derived from the subject's written responses based on the number of words, number of psychological ideas, and total number of ideas contained in the description. Three additional scores were developed by summing the three scores described above across all three pictures.

Literature Empathy Test (LET). This measure is conceived as a test of empathic accuracy, based on criterion scores obtained by the combined judgements of a specially constructed panel of experts. Subjects are asked to read a short passage from English literature; in the present case a selection from Schulberg's *Crowd Pleaser*. In a series of twenty questions following the passage, the subject is asked to answer as the individual portrayed in the passage would do if he were answering the same questions. Criterion answers have been established for this measure by submitting the task to thirteen trained clinicians. Only items on which this panel

reached unanimous or near-unanimous agreement were retained in the final test. Scores are based upon the number of correct answers for certain specified questions plus a constant ten.

Personal Rigidity (Authoritarianism). This is a series of measures designed to tap stable personality factors relating to interpersonal sensitivity. Specific items in the measures are largely taken from the Adorno et al. (1950) scales of Authoritarianism. A six-point scale of intensity of agreement, ranging from "strongly agree" to "strongly disagree", is provided for each item. In addition to these responses, subjects were allowed to write in "can't decide" beside an item for which they could reach no decision, in order to minimize meaningless responses.

Verbal Mental Reasoning (VMR). This test is a general measure of ability in manipulating verbal concepts, specifically, verbal analogies. It consists of forty items, for each of which two choices are required in order to achieve a correct answer.

General Information. A short questionnaire was developed in order to obtain demographic, motivational, and previous language experience information about the subjects. It was felt that such information was critical in terms of the nature of the sample. In addition, it made possible the testing of hypotheses concerning the role of such variables in authenticity of pronunciation.

Authenticity of Pronunciation (AP). This is a measure of authenticity of second language pronunciation, as judged by a panel of three trained experts in the language being studied by the student. A script composed of ten words and ten sentences in each desired language was prepared, each word and sentence having from one to three predetermined points of pronunciation difficulty. A high-quality tape recording was made as a native speaker of this language enunciated each of the sentences in a natural style. In the testing phrase, subjects were asked to repeat each sentence, reproducing the words exactly as they sounded. (A printed copy of the script was available during this procedure in order to minimize effects of short-term memory and lexical and grammatical familiarity.) A two-track recording was made during this test, with the stimuli recorded on one track and the subject's responses recorded on the other. In the evaluation phase the tapes were replayed by three experts in each language, working independently. Each predetermined critical point of pronunciation was evaluated on a fixed scale of 1-3 as follows: 1. Poor - strongly marked with foreign accent or almost unintelligible; 2. Fair - passable and not likely to cause any misunderstanding though not completely native-like; 3. Good - quite native-like or close to native-like pronunciation. Ratings of the three judges were summed. Interjudge reli-

bility of these ratings ranged from .72 to .91 with a mean value of .82.

Procedure

Phase I, administered during the first week (approximately) of the students' attendance of DLI included the following measures: MME, TAT, PPT, LET, Personal Rigidity, VMR, and the General Questionnaire.

Phase II, administered approximately 3 months after arrival at DLI, consisted of the AP measure. The AP was administered in a language laboratory, in small groups consisting of class sections. The pre-recorded tests (one for each of the five languages) were played simultaneously to each student in the section, through ear phones in each booth. Tape recorders in each booth recorded the stimulus sounds from the master tape and the responses made by the student.

RESULTS⁴

Reliability of Instruments

Most of the psychological instruments employed in the present research are refinements of instruments that have been used in previous work. Because of the unique characteristics of the DLI student population, however, and because of the aim of this study to develop a test battery possessing predictive validity rather than merely qualitative findings, special interest was attached to the reliability of the various measures. Wherever possible, this was assessed in more than one way in the current study.

In general the reliability of the measures employed were acceptable. Although test-retest figures are lower in every case than coefficient alpha for internal homogeneity, the time span between testings was three months. Thus the obtained figures in the range of .61 - .85 are actually rather impressive. The internal homogeneity of test variance at any testing period ranged from .69 to .92, so that these figures represent upper limits to the attainable validity of the individual measures. In view of the complexity of most of these measures and the necessity for group administration in large rooms under slightly different circumstances at different test dates, this range is about as high as we believe we could have expected to attain.

⁴For a complete description of statistics and tabular results, consult the *Final Report*.

Demographic, Language Experience, and Motivational Factors

In addition to the theoretically based measures of variables potentially related to second-language aptitude discussed above, other factors known or hypothesized to be relevant to this ability were examined.

Prior familiarity or experience with the assigned language, and with all other languages other than English, was determined for all subjects in this research. This included speaking, reading, and understanding proficiency with each language just prior to attending DLI, amount of formal instruction (if any) in each language, experiences with all languages while living outside the continental United States, and other sources of contact or familiarity with any foreign language. After some preliminary screening on these variables all those subjects who knew more than a few words or phrases in the assigned language prior to reporting to DLI were eliminated from data analysis, and are not presented in data reported in the remainder of this section.

Motivation for learning is a construct which presents special difficulties for measurement in almost every context. It was our expectation that motivation "to do well enough to get by" would be reasonably high but largely independent of motivation "to excel in learning the assigned language." In addition it was anticipated that changes in both dimensions of motivation would frequently occur during the relatively short but intensive period which students spend at DLI.

One approach to measurement of general motivation which we attempted was a set of essentially straightforward questions which asked the student to describe "as accurately and honestly as possible" his degree of interest in learning a foreign language. This battery of questions actually resulted in a quite negligible predictive relationship with language proficiency. A social-desirability index interspersed among these questions indicated that most subjects were inclined to give the socially-approved answer—such as that they were extremely enthusiastic about learning a foreign language. This tendency may have accounted for the lack of predictive validity for this form of measure.

A second and considerably more successful method of assessing effective motivation was based on configural scoring of responses to two desirability-balanced items. The wording of these questions was as follows: (A) If you could be doing any of the following things during the next few months, which would you most like to do.⁵ (B) Now suppose that you had to spend the next few months learning

⁵The choices for this question were: teach in a high school, learn to fly a light airplane, win \$25 in a contest, drive a jeep for a General, learn a foreign language, or be stationed in Canada.

some foreign language but you could select any language in the world to learn. Which language would you choose? Subjects were asked to indicate their first three choices in response to each question.

The intent of Question (A) was to provide the student with a less abstract and more specific set of alternatives to learning a foreign language, and in Question (B) to learning the particular foreign language they had been assigned. This use of specific alternatives appears to have succeeded to some extent in overcoming the tendency to exaggerate interest in the assignment; although most students included the assigned alternative among their three choices, less than 30% indicated that learning the particular assigned foreign language was what they would most like to be doing "during the next few months."

Questions (A) and (B) were combined by means of configural scoring rather than simple additive combination. (For a complete discussion of the theory and method of configural scoring, see Meehl 1950, and the *Final Report* of this study).

Prediction of Performance in General (Across Languages)

Five languages—Japanese, Chinese-Mandarin, Thai, Spanish, and Russian—were included in the present study. Previous theoretical analysis of second language aptitude have suggested that this ability is relatively nonspecific (e.g. Carroll 1963), and the most widely used existing measures of second language ability (the Modern Language Aptitude Test (MLAT), and the Army Language Aptitude Test (ALAT) are general tests. Therefore our first step in analysis was to determine the degree to which the measures we employed could be used to predict performance in all five languages taken together.

A multiple regression coefficient $R = .323$ ($N=401$) was obtained based upon six measures: MME, Personal Rigidity, LET, VMR, Configural Motivation, and Grades. A comparison with ALAT scores which were available for approximately half of the sample indicated that these measures were no more successful than the ALAT ($r = .36$) in predicting pronunciation performance.

In the course of data analysis it became apparent that the low over-all prediction was due to cancelling relationships among the five different languages which made up our sample. We therefore proceeded to re-analyze the results by each of the five languages individually.

Specific Language Performance Results

Separate analysis for Chinese-Mandarin, Thai, Japanese, Spanish,

and Russian were carried out. This involved the calculation of product-moment correlations and gammas and a (step-wise) multiple regression coefficient based on the best five predictor measures for each language. (The limitation to 5 measures was imposed for two reasons: (1) relatively short predictive instruments were desired as the product of this research; (2) as the number of measures included in a multiple regression coefficient increases, the possibility of chance factors playing a significant role also increases, with the likely effect that replicated results would be lower than expected.) For each language the relationship of ALAT scores from students' records to language performance in our study is discussed for purposes of comparison.

For Chinese-Mandarin, Thai, Japanese, and Spanish the results indicate that the battery of measures developed in this research have considerable validity as predictors of second language performance as measured in this study. The respective multiple R 's range from .59 to .69 and perhaps more importantly, in each case they exceed the predictive validity of the ALAT by at least twenty points. For Russian, on the other hand, all correlations in general are quite low and in particular the correlations with the empathy measures.

It is, however, not sufficient to look merely at the multiple R 's. In order to more fully understand the meaning and significance of the resulting predictive batteries, it is essential to examine more closely the correlations, by language, of which the multiple correlation coefficients are composed.

Empathy Measures and Authenticity of Pronunciation

Among the five highest correlations for each of the five languages, at least two of the variables are measures of empathy: for both Spanish and Russian two of the five correlations are based on empathy measures; for Chinese-Mandarin, Thai and Japanese three of the five predictor variables are based on empathy measures. The MME is common to all three languages. The results indicate that in general our empathy measures are among the highest predictors to AP for all languages. It is also apparent that certain measures are more consistent in predicting to AP than others.

The Personal Rigidity measure is the least successful of the empathy measures in terms of frequency of inclusion in the predictive batteries. It is found only in the Thai battery. Nevertheless the magnitude of the correlation ($r = .41$) substantially exceeds that between ALAT and AP ($r = .21$) for the Thai group.

The LET is a better predictor of AP for Chinese-Mandarin ($r = .44$) and Japanese ($r = .35$) than the ALAT ($r = .36$ and $r = .32$ respectively for Chinese-Mandarin and Japanese) although it does not appear at all among the five highest predictors for Russian, Spanish, or Thai.

The PPT measure also occurs in the predictive batteries of only two of the five languages, namely Chinese-Mandarin ($r = .57$) and Thai ($r = .30$). While both PPT scores are based on the number of psychological ideas count, for Chinese-Mandarin the correlation is based on total psychological ideas for all three pictures. For Thai the correlation is based upon total psychological ideas for picture #3 only. Although there is no reason to think that these refined scores are measuring totally different things, nevertheless the correlation for Thai is positive and the correlation for Chinese-Mandarin is negative.

Looking at the TAT measure we find that it occurs in three out of the five predictive batteries. For Russian ($r = .15$) and Spanish ($r = -.22$) the correlation is based upon scores on TAT #1; for Japanese ($r = .62$) the correlation is based upon scores on TAT #3. Despite the fact that TAT correlations for Russian and Spanish are based upon the same type of scoring scheme, the sign of the correlations differ with Spanish negative and Japanese and Russian positive.

Of all the empathy measures, the MME appears to be the most consistent. It is among the five highest predictors for all five languages and with the exception of the Russian group, it exceeds the magnitude of the ALAT/AP correlation ranging from .49 to .62. There are, however, certain apparent inconsistencies with regard to the direction of the correlation. For Spanish ($r = .49$), Russian ($r = .16$) and Chinese-Mandarin ($r = -.35$) correlations are based on equivalent scores. The correlations for Russian and Spanish are positive; while for Chinese-Mandarin it is negative. For Thai ($r = -.54$), and Japanese ($r = .42$), the MME correlations are positive for Japanese and negative for Thai.

Verbal Mental Reasoning (VMR) and Authenticity of Pronunciation (AP)

The VMR test, the intelligence measure employed in this study, figures in three of the five predictive batteries: Spanish ($r = .23$), Chinese-Mandarin ($r = .30$) and Russian ($r = .19$). By way of comparison, the correlation of VMR/AP for Thai is $r = .19$ which falls in the range of the other correlations but which does not fall in the top five predictors for Thai. For the Japanese group, however, the correlations between VMR and AP is 0.

Differences of Student Populations in the Five Language Groups

If student populations at DLI vary radically in important demographic, motivational, or psychological factors, it would not be surprising to find highly different trends among the measures that could be useful in predicting their language performance. In this case the differences observed in the previous section would be essentially artifacts of the language-assignment process at DLI, with no substantive meaning for the understanding of second-language acquisition.

To determine if this was the case we examined the distribution of the following factors available within our data for each language population: age, rank, education, number of years in military service, self-reported grades in school, birth order, locale of childhood, size of home community, verbal-mental reasoning ability, and reported motivation to learn the assigned language.

Mean scores for seven of these variables were calculated for each of the five language groups. The only group which varied appreciably from the total population (via Sheffé test) is the Thai population, which is somewhat older and more likely to have attained officer rank. Although statistically significant. ($p < .05$) these differences are by no means large in absolute terms, and none of the other "background" factors reveal any significant differences among languages. The Russian group, whose performance was only moderately successfully predicted in the present study, did not differ significantly from the other groups on any factor.

DISCUSSION

The attempt to predict authenticity of pronunciation on the basis of data pooled for all languages was relatively unsuccessful. A more fruitful approach to prediction was provided by separate analyses for each of the five target language groups. An examination of the various measures which make up the predictive batteries indicated that certain measures were more consistent than others, in terms of frequency of inclusion in the test batteries. In addition, measures from the same test occasionally gave rise to correlations with AP which differed in magnitude and, in several instances, sign.

In evaluating the results of our study we are faced with two very important issues: 1) general versus specific language learning ability, an issue with both theoretical and practical implications and 2) small numbers versus large numbers, a critical methodological issue which has recently been discussed by Tversky and

Kahneman (1971). These two issues with regard to the present study are not completely independent.

In comparing the five predictive batteries, it is clear that the empathy measures employed in this study figure importantly in the resulting multiple R 's. Looking at the measures which make up the five batteries, it is apparent that the greatest variation is found in Thai, Chinese-Mandarin and Japanese while for Spanish and Russian, the measures are virtually identical.

It is critical, we believe in interpreting these results to recall the sample size for each of the five language groups. We are extremely reluctant to interpret the differing patterns of measures for Thai ($N = 40$), Chinese-Mandarin ($N = 39$) and Japanese ($N = 22$) as reflecting language specific differences when we consider that we are dealing with samples of the size indicated above. Otherwise we would be led to conclude, for example, that intelligence via the VMR test is a component of pronunciation ability in Russian, Spanish, Chinese-Mandarin, and Thai but does not figure in the ability to authentically pronounce Japanese. Considering what we know of the role of intelligence in virtually all cognitive skills, plus the fact that the results for Japanese were based on a sample size of 22, we can see that such a conclusion is unwarranted. Similarly, the finding that the pattern of empathy measures which predict AP in Thai, Chinese-Mandarin, and Japanese is slightly different for each language does not let us conclude with confidence that these differences are absolute. The results for these languages, however, can be usefully thought of as general trends which may give tentative evidence in support of the findings for the larger language groups such as Russian and Spanish.

In making comparisons between the predictive batteries for Russian and Spanish, we are immediately struck by the fact that the *test* measures included in the batteries are identical. What is most interesting is that if we now include the small language groups in our analysis, a definite pattern emerges in which, with the exceptions noted above, we find the MME, VMR, and Configural Motivation measures as the critical cluster of variables common to all the batteries.

In selecting measures of empathy to include in our study we were concerned with the general inadequacy of existing measures. The measure which we had employed in previous research, the MME, was of particular interest to us. Theoretically, the MME provides the best operationalization of the construct of empathy as we have defined it. We had predicted that MME scores would correlate positively with Authenticity of Pronunciation but the results of an initial study, (Taylor, Guiora, Catford, and Lane 1969)

showed that MME correlated negatively with Authenticity of Pronunciation. At that time we interpreted this unexpected finding in terms of anxiety; that is, a high score on the MME was a reflection of anxiety which was aroused by observing the emotional responses of the woman in the film. A recent study (Garwood, Guiora, and Kalter 1970) which attempted to investigate the hypothesized relationship between MME total responses and anxiety concluded that anxiety as measured by the *Taylor Manifest Anxiety Scale* and the *Mandler-Sarason Test Anxiety Questionnaire* was not related to MME total response score. The results from the present study indicate that MME does in fact show a *positive* relationship to AP thus confirming our original hypothesis. The correlations for Russian, Spanish and Chinese-Mandarin are all based on total scores from segment C at 24 f.p.s. We are unable at this point to account for the fact that this score appears to be a more sensitive predictor of AP than other MME scores including the total response score used in the previous studies. Considering, however, the sample size of the Russian and Spanish groups we must conclude that MME empathy is indeed positively related to AP.

While we have been concerned to find commonalities among the predictive batteries, we are nevertheless faced with a number of important differences. As indicated above, we are reluctant to interpret differences among the small language groups as language specific differences. It is clear, however, that despite the similarity between the Russian and Spanish groups, there is the obvious fact that all correlations for the Russian group are exceedingly low. Considering that the Russian group did not differ significantly from the other groups in terms of various background factors such as age, education, intelligence, etc., we are somewhat at a loss to know how to interpret this finding. That is, we do not know precisely what aspect of the total language situation might be responsible for this result.

REFERENCES

- Adorno, T. W., Else Frenkel-Brunswik, D. J. Levinson, and N. Sanford. 1950. *The Authoritarian Personality*. New York: Harpers.
- Anisfield, M., and W. E. Lambert. 1961. Social and psychological variables in learning Hebrew. *Journal of Abnormal and Social Psychology* 63,524-529.
- Carney L. 1924. Studies of emotional reactions. *Journal of Experimental Psychology* 7. 325-341.
- Carroll, J. B. 1963. Research on teaching foreign languages. In N. L. Gage (ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally.

- Carroll, J. B. 1965. Contributions of psychological theory and educational research to the teaching of foreign language. *Modern Language Journal* 49, 273-281.
- Dymond, R. A. 1948. A preliminary investigation of the relation of insight and empathy. *Journal of Consulting Psychology* 12, 228-235.
- Dymond, R. A., A. S. Hughes, and V. L. Raabe. 1952. Measuring changes in empathy with age. *Journal of Consulting Psychology* 16, 202-206.
- Feffer, M. H. 1959. The cognition implications of role taking behavior. *Journal of Personality* 27, 152-168.
- Feleky, A. 1914. The expression of the emotions. *Psychological Review* 21, 33-41.
- Frois-Whittmann, J. 1930. The judgment of facial expression. *Journal of Experimental Psychology* 13, 113-151.
- Gardner, R. and W. Lambert. 1965. Language aptitude, intelligence, and second language achievement. *Journal of Educational Psychology* 56, 191-199.
- Garwood, R., A. Z. Guiora, and N. Kalter. 1970. Manifest anxiety and perception of micro-momentary expression. *Comprehensive Psychiatry* 11, 576-580.
- Guilford, J. P. 1929-1930. An experiment in learning to read facial expression. *Journal of Abnormal Psychology* 24, 191-202.
- Guiora, A. Z. 1967. Toward a systematic study of empathy. *Comprehensive Psychiatry* 8, 375-385.
- Guiora, A. Z., H. L. Lane, and L. A. Bosworth. 1967. An exploration of some personality variables in authentic pronunciation of a second language. In H. L. Lane and E. M. Zale (eds.) *Studies in Language and Language Behavior* 4, 261-266.
- Guiora, A. Z., L. Taylor, and M. A. Brandwin. 1969. A contribution to the psychology of second language behavior. *Proceedings of the XVth Congress of Applied Psychology*. 585-588.
- Guiora, A. Z. 1970. Transpositional research in the clinical process. *Comprehensive Psychiatry* 11, 531-538.
- Guiora, A. Z. 1972. Construct validity and transpositional research, toward an empirical study of psychoanalytic concepts. *Comprehensive Psychiatry* 13, 139-150.
- Haggard, L. A. and R. S. Isaacs. 1966. Micro-momentary facial expressions as indicators of ego mechanisms in psychotherapy. In L. A. Gottschalk and A. H. Auerback, (eds.), *Methods of Research in Psychotherapy*. New York: Appleton-Century Crofts.
- Mahoney, S. 1960. The Literature Empathy Test. Unpublished doctoral dissertation, University of Oklahoma. Ann Arbor, Michigan: University Microfilms N. 60-2677.
- Mandel S. 1927. The differentiation of emotional responses in infants. *Journal of Comparative Psychology* 7, 265-270.
- Meehl, P. E. 1950. Configural scoring. *Journal of Consulting Psychology* 14, 165-171.
- Modern Language Association. 1967. Childhood and second language learning. In S. Levinson and W. Kendrick (eds.), *Readings in Foreign Languages for the Elementary School*. Waltham, Massachusetts: Blaisdell Publishing Co., 53-60.
- Nida, E. G. 1957-1958. Some psychological problems in language learning. *Language Learning* 8, 7-15.

- Osgood, C. E. 1966. Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology* 7. 1-30.
- Pike, K. L. 1947. *Phonemics: A Technique for Reducing Languages to Writing*. Ann Arbor, Michigan: University of Michigan Press.
- Pimsleur, P., L. Mosberg, and A. Morrison. 1962. Student factors in foreign language learning: A review of the literature. *Modern Language Journal* 46, 1960-70.
- Pimsleur, P., D. Sundland, and R. McIntyre. 1963. Underachievement in foreign language learning. Final report, U. S. Department of Health, Education and Welfare, Office of Education.
- Pimsleur, P. 1968. Aptitude testing. *Language Learning, Special Issue* No. 3. 73-78.
- Ruckmick, C. A. 1921. A preliminary study of the emotions. *Psychological Monographs* 39.
- Schlossberg, H. 1954. Three dimensions of emotion. *Psychological Review* 61. 81-88.
- Smith, F. and E. Berger. 1968. An assessment of three foreign language learning strategies utilizing three language laboratory systems. U. S. Department of Health, Education, and Welfare, Office of Education, Bureau of Research, Project No. 5-0683.
- Stengel, E. 1939. On learning a new language. *International Journal of Psychoanalysis* 20. 471-479.
- Taylor, L. L., A. Z. Guiora, J. C. Catford, and H. L. Lane. 1969. The role of personality variables in second language behavior. *Comprehensive Psychiatry* 10. 463-474.
- Tomkins, S. S. and C. E. Izard (eds.). 1965. *Affect, Cognition, and Personality: Empirical Studies*. New York: Springer Publishing Co.
- Truax, C. 1963. Effective ingredients in psychotherapy: an approach to unraveling the patient-therapist interaction. *Journal of Counseling Psychology* 10. 256-268.
- Tversky, A. and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76. 105-110.
- Wallach, M. and H. Strupp. 1960. Psychotherapist clinical judgments and attitudes toward patients. *Journal of Consulting Psychology* 24. 316-323.