# Faces in a Crowd: The Individual Learner in Multisection Courses 

Carol A. Klee Editor

# Copyright 1994 by Heinle \& Heinle. No parts of this publication may be reproduced or transmitted in any form or by any means electronic, or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. <br> Manufactured in the United States of America. 

Heinle \& Heinle Publishers is a division of Wadsworth, Inc.
ISBN 0-8384-63673
$\begin{array}{lllllllll}10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2\end{array}$

# Weakest and Strongest Learners in Intensive Language Training: A Study of Extremes 

## Madeline Ebrman

Foreign Service Institute

## Introduction

Many language training programs seek to select students. For example, they may seek the fastest learners for accelerated training or the weakest for remedial training or even for exclusion from training programs. In addition, even in the more usual situation-classes with a range of performance levels-it is often helpful to understand what characterizes the weakest and strongest learners in order to maximize their strengths and minimize their weaknesses. This study looks at the weakest and strongest $2-4$ percent of a large group of adult students in a federal language training program to see how they can be characterized in terms of a wide range of individual difference variables often addressed in the literature of second language acquisition. ' Although the subjects were not university students, nearly all were at least college graduates and thus may in some ways be similar to university students, especially those in majors that lead to careers in foreign affairs (e.g., international relations or foreign language and culture).

The students described here are drawn from a much larger sample in use for a multivariate study in progress that is examining the effects of individual differences on student achievement in intensive language training at the Foreign Service Institute (FSI), the training branch of the U.S. Department of State. Variable categories in the study include language learning aptitude, demographics, preferred learning strategies, motivation and anxiety, and personality factors.

## Review of Literature

A number of individual difference categories are proving of interest in second language classrooms. These include language aptitude, age, sex, motivation, anxiety, self-esteem, tolerance of ambiguity, risk taking, language learning strategies, and language learning styles. Among the last named are often included personality type factors. Much detailed information on individual differences is available in the work of Ehrman (1990 1994), Ehrman and Oxford (1990 and in press), Galbraith and Gardner (1988), Oxford (1992), Oxford and Ehrman (1993, in press), and Skehan (1989). Although these variable categories are treated separately below, most of the cired findings indicate complex interrelationships among them.

## Aptitude

Several tests have been devised to attempt to measure language learning aptitude; these have been of particular interest to government agencies and others who want to select students for training. The oldest of these is the Modern Language Aptitude Test (MLAT; Carroll and Sapon 1959; research summarized in Carroll 1990); it is still used by a number of universities and by at least one government agency. The U.S. Department of Defense agencies have developed their own language aptitude measures; their validity appears to be comparable to the MLAT, according to a personal communication (1993) from James Child a testing expert in the Department of Defense. Another such test commonly used for adolescents and children is the Pimsleur Language Aptitude Battery (described in Skehan 1989). Although some have suggested that such measures-developed for audiolingual language teaching-are incompatible with less structured, highly communicative language teaching approaches (see Parry and Stansfield 1990), at least one study shows that in fact the validity of the MLAT remains essentially unchanged since it was developed, despite substantial changes in training methodology (Ehrman and Oxford, in press).

## Age

Most studies find that age is related to language learning success, generally finding that younger learners do better with respect to critical periods, type of input, and certain affective and sociocultural features (Oxford 1992; Schleppegrell 1987; Singleton 1989). However, one study of the same sample examined for this research showed that for a group of adults between roughly twenty-five and forty-five, the correlation of younger age
with speaking and reading proficiency was only about .30-enough to be worthy of mention, but far from sufficient to be a selection factor for training (Ehrman and Oxford, in press).

## Sex

Many studies in various parts of the world have found some advantages for females over males for use of conscious language learning strategies (Oxford, in press) and for listening skills (Larsen-Freeman and Long 1991). Oxford, Nyikos, and Ehrman (1988) cite a number of works that found that females tend to perform better than males in language classrooms. However, the correlations from the current FSI study (Ehrman and Oxford, in press) show no significant relationship between sex and end-oftraining proficiency in a wide range of languages. (This finding suggests that sex differences found so far may be related to moderating variables such as education level, career choice, or interests, since the sample is generally homogeneous by sex on these variables.)

## Motivation

Countless publications have addressed motivation and have found it to be a key variable in language learning success. Most readers are likely to be familiar with the distinction between instrumental (career-oriented) motivation and integrative (cultural-merging) motivation for language learning (Au 1988; Gardner 1985a; Horwitz 1990; Oxford and Shearin 1994). Additional investigation has added more variables to be considered in evaluating motivation, including need for achievement, fear of failure and possibly even fear of success in the language classroom, self-efficacy, and attribution of locus of control to personal effort versus outside factors like fate or society (Oxford and Shearin 1994). These authors note that motivation is related to expectancy of success and the value students place on such success, and that goals must be clear, challenging, and reachable, with feedback on goal achievement; in addition, mode of instruction also affects motivation (e.g., mastery learning may be more motivating than norm-referenced learning). The field of industrial-organizational psychology distinguishes between motivation due to extrinsic rewards and the intrinsic motivation that is activated by skilled teaching; the latter is thought to be more powerful than external rewards (Beck 1990). Extrinsic motivation may be similar to Gardner's (1985a) instrumental motivation, while intrinsic motivation may well encompass much of Gardner's integrative motivation as well as other motivations such as a desire to feel self-efficacy

## Anxiety

Language anxiety appears in a variety of forms-avoidance, face-saving humor, physical activity, psychosomatic symptoms-and has been shown to be deleterious among university students (Horwitz and Young 1991). Indeed, Horwitz (1990) suggests that in the language learning environment all anxiety is likely to be debilitating because language learning is such a complex and emotionally involved process. Others (e.g., Brown 1987) suggest that some degree of anxiety can actually be helpful for language learning. The correlation findings for the population reported on in this study confirm the latter view, at least for high-functioning adults (Ehrman and Oxford, in press).

## Tolerance of Ambiguity and Risk Taking

Ely (1989) describes tolerance of ambiguity as acceptance of confusing situations and lack of clear lines of demarcation. Ehrman (1993) describes it in Piagetian terms as the ability not only to let new information in, but to hold contradictions in mind while they are integrated into new conceptual frameworks. She operationalizes tolerance of ambiguity through the concept of ego boundaries, readdressing Giora's (1972) concept of "language ego," showing a complex relationship between ego boundaries and learning success. Other studies have found that tolerance of moderate levels of ambiguity is related to persistence in language learning (Chapelle 1983; Naiman, Fröhlich, and Stern 1975) and frequency of use of certain learning strategies (Ehrman and Oxford 1989, 1990; Ely 1989). Risk taking is linked to tolerance of ambiguity, in that tolerance of ambiguity appears to lead to willingness to take some risks in language learning; and risk taking is an essential for progress (Beebe 1983; Brown 1987; Ely 1986; Stevick 1976).

## Language Learning Strategies

There appears to be agreement among a number of authors that language learning strategies can be described as specific behaviors or techniques used by learners to increase their language growth (Cohen 1990; O'Malley and Chamot 1990; Oxford 1990; Wenden 1991; Wenden and Rubin 1987); they are selected in the light of the task, the goals, the curriculum, and the individual's personality and stage of learning (Ehrman 1989; Ehrman and Oxford 1990; Oxford 1990; Skehan 1989). The ability to use the right strategies at the right time appears to be more important than sheer number of strategies (Vann and Abraham 1989); hence language learning
research has focused on cognitive and metacognitive learning strategies such as language practice or hypothesis formation (cognitive) or planning and evaluating one's study (metacognitive) (Oxford 1990).

## Language Learning Styles

Language learning styles are the general approaches students use to learn or that affect their response to variations in curriculum or teaching technique (Ehrman 1990; Oxford 1992; Oxford, Ehrman, and Lavine 1991). One common distinction is analytic versus global (Schmeck 1988): analytic students tend to like to work within clear categories and to analyze components of language, whereas global students are likely to prefer conversation to rule learning and practice (Oxford, Ehrman, and Lavine 1991; Schmeck 1988). Another common category is sensory channel preference for learning activities; the usual distinctions are visual, auditory, and kinesthetic. Reid (1987) showed differences for ESL students' sensory preferences by national origin-for example, Asian learners tended to prefer visual input.

Within the category of learning style come personality preference variables. One common measure of such variables has a very rich history of use in educational settings: the Myers-Briggs Type Indicator (MBTI; Lawrence 1984). Its four dimensions are extraversion-introversion, sensing-intuition, thinking-feeling, and judging-perceiving (Myers and McCaulley 1985). Although it has seemed natural to speculate that the best learners would be extraverts because of their willingness to speak out and interact (Brown 1987), findings by Busch (1982) failed to confirm this hypothesis, and some recent research (Ehrman 1989; Ehrman and Oxford 1990, in press) indicates that introverts are not disadvantaged, at least in the small, long-term intensive classes characteristic of FSI. Intuitive learners, who tend to prefer abstract, random, future-oriented learning, seek generalizations and meaning; they are often bored by concrete, step-by-step learning. On the other hand, sensing learners find facts intrinsically interesting, may be less interested than intuitives in underlying principles, and tend to prefer concrete, sequential learning. Thinking-oriented students prefer logical and impersonal processing, whereas feeling-oriented learners want to make most forms of learning personal and grounded in relationships. Judgers seek closure, product, and a clear external structure; perceivers may resist external structure, be less oriented to meeting requirements, and may respond favorably to the relatively unstructured aspects of communicative methodology (Ehrman and Oxford 1989, 1990).

Although learners are likely to learn best initially when at least some attention is paid to their stylistic "comfort zone," thoge who can eventually
use strategies not necessarily related to their preferred styles appear to be at an advantage; indeed, some level of versatility seems to be prerequisite to effective learning at some times (Ehrman 1989; Ehrman and Oxford 1990; Oxford and Ehrman, in press; Oxford, Ehrman, and Lavine 1991).

## Methodology

## Sample

The 15 very weak and the 27 very strong FSI students whose characteristics are addressed here were drawn from a subsample of 770 FSI students who had completed training. The subsample in turn was taken from a complete sample to date of over 1,200 students entering intensive long-term language training. For the 770 -person subsample, the mean age was thirry-nine (SD 9) and the educational level averaged between B.A. and M.A. The sample was 55 percent male; the median number of languages previously studied was two. Most of the group ( 71 percent) were from the Department of State, while 10 percent were from the Department of Defense, 8 percent were from the U.S. Information Agency, 7 percent were from the Agency for International Development, and the rest ( 4 percent) were from other government agencies like the Department of Agriculture, the Department of Commerce, and the Drug Enforcement Agency. Of the total sample, 83 percent were employees, and 17 percent were spouses or college-age children. Of the dependents, 84 percent were female. English was the native language of 99 percent of the students. These students spent an average of twenty weeks in full-time intensive training, with a range of eight to forty-four weeks. Slightly less than one-third each were learning Spanish ( 29.6 percent) or French ( 28.5 percent). In order of numbers of students, the other thirty-two languages studied were Italian, Portuguese, Chinese, Arabic, Russian, Thai, Turkish, Hebrew, German, Dutch, Urdu, Indonesian, Burmese, Polish, Romanian, Serbo-Croatian, Greek, Japanese, Korean, Czech, Danish, Hindi, Bengali, Afrikaans, Finnish, Norwegian, Bulgarian, Cantonese, Lao, Swahili, Swedish, and Tagalog. Median end-of-training scores were S-2 R-2. (Scores are discussed in "Measures of Student Language Proficiency," below.)

## Filter Variables

The extreme student groups were designated by two dichotomous filter variables, one for the weak students versus all others, and one for the strong
students versus all others. Both variables were determined through a combination of difficulty of language category, number of weeks of study, and end-of-training speaking score (EOTS). Language categories are established on the basis of relative difficulty for English speakers to learn; normal maximum training lengths differ among categories as follows: Category 1 (Western European languages), twenty-four weeks; Category 2 (Swahili, Indonesian, Malay), thirty-six weeks; Category 3 (all others except Category 4), forty-four weeks; Category 4 (Arabic, Chinese, Japanese, Korean), eighty-eight weeks. There were so few students studying Category 2 languages in the sample that this category was excluded in this study. EOTS is the score from the FSI end-of-training interactive proficiency test, which is described below under "Instrumentation."

## Weak Student Variable

Cases were selected if start-of-training speaking score was $\mathrm{S}-0$ or $\mathrm{S}-0+$ (beginners) and end-of-training speaking proficiency score (EOTS) < 1 , and:

Category of Language Number of Weeks >

1

$$
3
$$

4

16
30
36

## Strong Student Variable

Cases were selected if start-of-training speaking score was $\mathrm{S}-0$ or $\mathrm{S}-\mathrm{O}_{+}$ (beginners) and:

Category of Language Number of Weeks End-of-Training Speaking
$4<26 \quad 2$
$>20$
$<18$
$>36$
< 24
$>36$
$3+$
3
$3+$
3
$2+$

## Numbers of Strong and Weak Students

The numbers of extremely weak and strong subjects in the results cited below vary because not every participating student completed every instrument. Thus both total numbers of those for whom end-of-training data and any given instrument are available vary, as do the numbers of extreme
(strong or weak) students whose scores are available for each instrument. Some of the $N$ 's are low but are included because they are at least suggestive and appear to be consistent with trends indicated for other instruments.

The very strong students represent about 3.5-4 percent of their various subsamples; the weak students represent about 2-3 percent of their respective subsamples. The difference in proportions is attributable to two factors. The first is an artifact of the cutting scores for number of weeks of training and level of end-of-training speaking proficiency that were used in the formulas given above for selecting members of each category. The second reflects the reality that a certain amount of screening of students takes place before they are ever sent to training: students with a poor track record in language learning or with poor MLAT scores may never be selected for training in the first place.

## Individual Difference Variables

## Data Collection

Data collection was done through questionnaires. Students were asked to complete a biographical data form and between one and seven apritude, learning strategies, and learning styles instruments, based on a randomsampling procedure. ${ }^{2}$ No measures were repeated. Instruments are described below under the heading "Instrumentation."

All students were asked to take the MBTI in either the short version (Form G) or its longer version (Form J, with the Type Differentiation Indicator [TDI] scoring system) at the beginning of their training. Many of them also completed one or more of the following: the Strategy Inventory for Language Learning (SILL, $N=262$ ), the Hartmann Boundary Questionnaire (HBQ, $N=233$ ), the National Association of Secondary Schools Principals Learning Styles Profile (LSP, $N=276$ ), the Modern Language Aptitude Test (MLAT, $N=282$ ), and the Affective Survey ( $N=163$ ). Many of those who had completed their training by July $1993(N=770)$ also had end-of-training proficiency ratings in speaking and reading, as well as faculty ratings on overall effectiveness as learners, effort, and other factors. The number of students with complete end-of-training data differ from instrument to instrument.

Almost all students took the MBTI, and many had MLAT scores on record when they entered training. The other instruments were administered on a random-sample basis. That is, student identification codes were selected at random to choose students who would receive one, two, or three
questionnaires in addition to the MBTI. The number for the MBTI (TDI) is much larger than the others because it was also administered at the beginning of training to all French and Spanish students entering the classes beginning about each month from October 1991 through September 1992.

## Instrumentation

The Affective Survey. The Affective Survey is a 114 -item instrument developed by Madeline Ehrman and Rebecca Oxford (1991) based on the general ideas and in some instances adapted items from a variety of surveys and scales by Gardner (1985b), Campbell (1987), Horwitz (1985), Horwitz, Horwitz, and Cope (1986), and others. The authors recognized that no single survey or scale covered all the important affective (emotional and motivational) areas related to language learning success. The Affective Survey contains three parts: motivation (extrinsic, intrinsic, desire to use the language, and effort), beliefs about self as a language learner, and anxiety (as related to public performance, language use with native speakers, making errors, comprehension, self-esteem, competition, tests, outcomes, and general comfort-discomfort with language learning). The Affective Survey also has the option of a "negativity scale," which indicates to what degree a person agrees with negatively worded items about motivation and anxiety. The Cronbach alpha internal consistency reliability for the Affective Survey is .74, and the standardized item alpha is .82 .

The Hartmann Boundary Questionnaire. (HBQ; Harmann 1991). The HBQ was developed for research with sleep disorders and nightmares, using a psychoanalytic theoretical base. It is intended to examine the degree to which individuals separate aspects of their mental, interpersonal, and external experience through "thick" or "thin" psychological boundaries. Its 146 items address the following dimensions: sleep/dreams/wakefulness, unusual experiences, boundaries among thoughts/feelings/moods, impressions of childhood/adolescence/adulthood, interpersonal distance/openness/closeness, physical and emotional sensitivity, preference for neatness, preference for clear lines, opinions about children/adolescents/adults, opinions about lines of authority, opinions about boundaries among groups/peoples/nations, opinions about abstract concepts, plus a total score for all twelve of the above scales. Hartmann has found women and younger people to score consistently "thinner" than men and older people. Cronbach alpha reliability for the HBQ is .93 , and theta reliabilities for subscales fall berween .57 and .92 (Hartmann 1991).

The National Association of Secondary Schools Principals' Learning Style Profile. (LSP; Keefe and Monk 1986; Keefe, Monk, Letteri, Languis, and Dunn 1989). The LSP is a 125 -item composite measure composed of many different approaches to measuring learning style. The main subscales are cognitive skills (analytic, spatial, categorization, sequential processing, detail memory, discrimination), perceptual response (i.e., sensory preferences: visual, auditory, emotive/kinesthetic), orientations (persistence, verbal risk taking, manipulative), study time preferences (early morning, late morning, afternoon, evening), and environmental context for learning (verbal vs. spatial, posture, light, temperature, mobility, and grouping). Cronbach's alpha for the subscales ranged from .47 to .76 , with an average of .61 . Test-retest reliabilities were .36 to .82 after ten days and somewhat lower after thirty days. Concurrent validity of the LSP's analytic subscale with the Group Embedded Figures Test was $.39, p<.002$. Concurrent validity of the perceprual response subscales of the LSP with the Edmonds Learning Style Identification Exercise was $.51-.64, p<.002$. Many of the environmental context subscales of the LSP correlated with Dunn and Dunn's Learning Style Inventory, .23-.71, $p<.04-.002$.

The Myers-Briggs Type Indicator. (MBTI; Myers and McCaulley 1985), Form G. This instrument is a 126 -item, forced-choice, normative, self-report questionnaire designed to reveal basic personality preferences on four scales: extroversion-introversion (whether the person obtains energy externally or internally); sensing-intuition (whether the person prefers to take in information in a concrete/sequential or an abstract/random way); thinking-feeling (whether the person likes to make decisions based on objective logic or on subjective values); and judging-perceiving (whether the person prefers rapid closure or a flexible life). Internal consistency split-half reliabilities average .87 , and test-retest reliabilities are .70-. 85 (Myers and McCaulley 1985). Concurrent validity is documented with measures of personality, vocational preference, educational sryle, and management style (.40-.77). Construct validity is supported by many studies of occupational preferences and creativity.

The Type Differentiation Indicator. (TDI; Saunders 1989). The TDI is a scoring system for a longer and more intricate 290 -item form (MBTI, Form J) that provides data on the following subscales for each of the four MBTI dimensions: extraversion-introversion (gregarious-intimate, enthusi-astic-quiet, initiator-receptor, expressive-contained, auditory-visual); sens-ing-intuition (concrete-abstract, realistic-imaginative, pragmatic-intellec-
rual, experiential-theoretical, traditional-original); thinking-feeling (criticalaccepting, tough-tender, questioning-accommodating, reasonable-compassionate, logical-affective); and judging-perceiving (stress avoider-polyactive, systematic-casual, scheduled-spontaneous, planful-open-ended, methodicalemergent). The TDI includes seven additional scales indicating a sense of overall comfort and confidence versus discomfort and anxiery (guardedoptimistic, defiant-compliant, carefree-worried, decisive-ambivalent, intre-pid-inhibited, leader-follower, proactive-distractible), plus a composite of these called "strain." Each of the Comfort-Discomfort subscales also loads on one of the four type dimensions; for example, proactive-distractible is also a judging-perceiving subscale. There are also scales for type-scale consistency and comfort-scale consistency. The reliability of twenty-three of the twenty-seven TDI subscales is greater than .50, an acceptable result given the brevity of the subscales (Saunders 1989).

The Modern Language Aptitude Test. (MLAT; Carroll and Sapon 1959). This is the classic language aptitude test, with 146 items. The manual describes its five parts: I-number learning (memory, auditory alertness); II—phonetic script (association of sounds and symbols), III—spelling clues (English vocabulary, association of sounds and symbols); IV—words in sentences (grammatical structure in English); and V-paired associates (memorizing words). The MLAT was correlated .75 with the Defense Language Aptitude Battery (Peterson and Al-Haik 1976) and . 67 with the Primary Mental Abilities Test (Wesche, Edwards, and Wells 1982)—the latter suggesting a strong general intelligence factor operating in the MLAT. Split-half reliabilities for the MLAT are .92-.97, depending on the grade or age. For college students, validity coefficients are . 18-. 69 for the long form of the MLAT and .21-. 68 for the short form. For adult students in intensive language programs, validity coefficients are $.27-.73$ for the long form and .26-. 69 for the short form (Carroll and Sapon 1959). In this sample, almost all ( 95 percent) of the MLAT scores were current, that is, administered within the last three years. This study used the long form.

Strategy Inventory for Language Learning. (SILL; Oxford 1989). This eighty-item questionnaire was developed between 1985 and the present and has been used with over 5,700 language learners in many countries. The SILL asks students to react to a series of strategy descriptions (e.g., "I make associations between new material and what I already know") in terms of how often they use the strategies (always or almost always, generally, sometimes, generally not, never or almost never). In studies worldwide, the

SILL's reliability using Cronbach's alpha is .93 to .98 , with an average of .95 , and it has been shown to be a valid, significant predictor or correlate of language proficiency and achievement.

End-of-Training Learning Activity Questionnaires. Two questionnaires were distributed at the end of training. One, developed by Lucinda HartGonzalez , Nikolaus Koster, Gisela Gonzales, and Madeline Ehrman at FSI, addresses various activities reported by a "snapshot" of FSI students for study on their own, without a teacher (Hart-Gonzalez 1991). Its thirty-six items ask the student to (1) assess the utility of the activity on a scale of "not useful, somewhat useful, very useful" and also (2) to estimate the approximate number of hours spent weekly on it. The other was developed by Madeline Ehrman and Frederick Jackson at FSI based on knowledge of frequent classroom events and on student end-of-training comments about their language learning experience. Its seventy items ask students to assess on the same Likert-type scale as the self-study questionnaire the utility of selected classroom events in the areas of conversation, pronunciation, grammar study, listening practice, reading practice, vocabulary study, classroom structure, and role of the teacher. Because it has taken a long time to amass a sufficient number of these two questionnaires for analysis (departing students are less likely to turn in questionnaires than those still in training that we can pursue), reliability and validity studies remain to be done. In fact, the present report is a first contribution to evaluation of their validity.

Measures of Student Language Proficiency. At the end of training, FSI students are given proficiency assessments resulting in ratings ranging from 0 to 5 for speaking (including interactive listening comprehension) and for reading. For example, R-3 means reading proficiency level 3 . S-2 means speaking proficiency level 2 . The ratings are equivalent to the ILR/ACTFL/ETS guidelines that originated at FSI and have been developed over the years by government agencies. (These guidelines are detailed by Omaggio Hadley 1989). FSI usually aims at end-of-training proficiency ratings of S-3 R-3 for full-time training, comparable to ILR Professional Proficiency or ACTFL Superior-Level Proficiency. Reliability studies have shown that government agencies have high interrater reliability for proficiency ratings within a given agency, but that the standards are not always the same at every agency; thus raters at different government agencies do not have as high an interrater reliability as raters at the same agency. Proficiency ratings are thus considered reliable indicators of the level of language performance of an individual student within an agency
(Clark 1986). Descriptive statistics for performance in terms of end-oftraining proficiency are provided in Table 2 in Appendix 1. "Plus" scores (indicating, e.g., proficiency between S-2 and S-3) were coded as .5; thus, for example, a score of S-2+ was coded 2.5 .

Faculty Rating Questionnaire. After training was complete, faculty were asked to rate students on how they compared overall with other FSI students known, on observed language learning aptitude, motivation, effort, and observed anxiety. Data were collected by interview in order to get a rich texture of comments as well as quantitative data. In order to achieve reliability, interviewers were trained and asked to follow the format of the questionnaire.

## Self-Report as an Issue in This Study

Self-report is sometimes viewed as suspect because of possible "social desirability response bias" (SDRB), that is, a tendency to answer in a way that would show the respondent to be in some socially acceptable way a "good person." The Affective Survey, HBQ, LSP, MBTI/TDI, and SILL are all self-report instruments and so are subject to questions of SDRB. Through instrument design, range of response in this sample, and precautions taken in the administration of the instruments, an effort was made to hold SDRB to a minimum in this study. (This problem is discussed in greater detail in Ehrman and Oxford, in press.)

## Data-Analysis Procedures

To test for internal consistency within the sample, an 855 -member subsample including the 770 subjects addressed in this investigation was divided randomly into two subsamples. The means were compared for the 135 variables of the entire project, using $t$-tests. With the exception of four variables, there was no difference between the means of the two subsamples; that is, they were essentially alike. The four significantly different means were within the range of chance at the .05 level, indicating that the two subsamples can be considered equivalent.

Each of the two variables derived from the formulas for finding the weakest and strongest learners was used to compare means for the various individual difference variables through a one-way analysis of variance using SPSS for Windows Version 5.0.1, through the "Compare Means" procedure. Results were considered significant at the .05 level; some were designated as near significant if the significance level was between .05 and
.099; and a few were listed as suggestive if their significance level fell between .10 and .17 . These subsignificant results were included because the number of extremely weak and strong students was so small that a moderate increase in the $N$ could well make the results significant and because they tended to pattern with other results that reached significance.

For correlational results referred to in this chapter, the analysis of choice was Spearman's rho on SAS, a correlation statistic usually used for rank-order data. When used with interval or ratio data, Spearman's rho provides a more conservative result than Pearson product-moment correlations. Since some of the measures involved ordered data with uneven intervals, and other measures involved equal-interval data, it seemed preferable to use the most conservative correlation coefficient (Spearman's rho) that could be used consistently with all the data. However, a check to see if there was a difference between Pearson's rand Spearman's rho revealed only very few small differences, all nonsignificant. All tests of correlational significance were two-tailed. Correlations of at least .20 are reported. Though .20 is low, findings at this level are reported so that later research can further examine them with other populations.

For all statistical tests reported in this study, the acceptable significance level was set at $p<.05$ level.

## Results

This report of results is organized by a set of questions about each group of students. Each question is answered first for the weakest students, then for the strongest. The questions are:

What are they like (demographically and in personality)?
How do they feel (motivation and anxiety)?
What are their abilities?
What do they do to learn (strategies)?
How do their teachers rate them?
Details of number of subjects, means and standard deviations for the criterion group and the remainder of the sample, F-statistics, degrees of freedom, and significance levels are provided in Table 1 (Weakest Students) and Table 2 (Strongest Students) in Appendix 1. These data are therefore not ordinarily provided in the following text, unless required to make a special point.

## What Are They Like?

## Weakest Students

These students have significantly less previous language learning experience than other students, in terms of number of languages previously studied. There are no gender differences at the significant, near-significant, or suggestive levels.

On the HBQ, while most of the differences were not statistically significant, the general direction was for the weakest students to show thicker ego boundaries on every one of the HBQ categories and on both HBQ factors. Those results reaching significance include Factor II (external boundaries), preference for sharp edges in images, neat and orderly surroundings, clear lines of authority in organization, and total HBQ score. All but the total score (which includes both internal and external items) are in the external boundary group.

On the MBTI, the weakest students show a preference for taking in information through sensing: these people prefer practical, sequential, factoriented learning, with little need to make inferences or design aspects of their own training. On the TDI scoring system, which provides subscales for each of the main MBTI scales plus seven comfort-discomfort scales, weakest students significantly report themselves as pragmatic (vs. intellectual).

## Strongest Students

These students show highly significant superiority in education level, in number of languages previously studied, and for highest speaking and reading scores in previously learned languages, relative to those who are not in this group. They also tend to be quite a bit younger (by about six years) than all other students and markedly younger than the weakest students (by nearly ten years). As in the case of the weakest learners, there is no difference with respect to gender.

The HBQ does not distinguish the strongest learners as clearly as it marks off the weakest ones. Only one HBQ category, a preference against neatness and order in the external world (thin), characterizes the strongest group. In contrast with the weakest students, who reported thicker ego boundaries on every category, whether significant or not, the strongest learners were much less consistent; there is no apparent pattern to the categories that have higher (thin) and lower (thick) means for this group relative to all the others.

On the MBTI, the only significant categories for strongest students were on two TDI subscales, where they reported themselves as more imaginative (vs. realistic) and more emergent (vs. methodical). The former is a sensing-intuition subscale, and the students reported themselves on the intuitive pole; the latter is a judging-perceiving subscale, and the students reported themselves on the perceiving pole.

Thus we see that weaker students are less experienced language learners; have thicker ego boundaries, especially with respect to the outer world; and prefer sensing perception to gather information. In contrast, the strongest students tend to be younger than other students, are advantaged with respect to previous learning, reject neatness and order in their surroundings, and report themselves as more imaginative (intuition) and emergent (perceiving) on the MBTI. There are no gender differences.

## How Do They Feel?

## Weakest Students

There are no significant or near-significant results for the weakest students on the Affective Survey. Faculty ratings for observed motivation and anxiety do not distinguish them either.

## Strongest Students

These students report lower intrinsic and overall motivation levels. On the other hand, they tend to endorse significantly fewer negatively phrased items, both in general and with respect to items probing various forms of anxiety. A combination of motivation and anxiety totals that is interpreted as indicating overall emotional arousal level is significantly lower for the strongest students than for all others. Teacher reports do not distinguish them by observed anxiety level, but they are viewed by their teachers as having been significantly more extrinsically motivated than their classmates.

## What Are Their Abilities?

## Weakest Students

On the main measure of language ability in this study, the MLAT, the weakest students performed strikingly worse than their classmates and did so to a high level of significance on all the MLAT subscales as well as on its total and the scaled Index Score. The most distinguishing score was the Index (T-score), which represents the entire performance of the student on
the MLAT; the weakest students did worse than other students to a highly significant degree. Of the subscales, those on which these students did particularly poorly were Parts III and IV. Part III tests both English vocabulary level and the ability to sort peculiar sound-related spellings. Part IV tests sensitivity to English structural nuances at the sentence level. In addition, the weakest students were significantly less adept at simultaneous visual processing on the LSP, showing less ability to match a stimulus to a whole image of which it is a part.

## Strongest Students

Although the MLAT separated the strongest students from the rest, it did so less clearly than it did for the weakest students. Again, the Index appears to distinguish the strongest students, as it did the weakest, but Parts IV (English sentence structure) and V (rote memorization) provided no discrimination for this group, and Parts I, II, and III had a weaker effect (measured both by significance level and by F-statistic) for the strongest students than for the weakest ones. None of the LSP ability scales were distinctive for the strongest students.

## What Do They Do to Learn?

## Weakest Students

Too few students in this group turned in the SILL and the end-of-training questionnaires on self-study and classroom activities for meaningful analysis.

## Strongest Students

There is more information available about how the strongest students learned, and even more about what they believed not helpful. On the SILL, they reported significantly more use of techniques to enhance memory. None of the results on the end-of-training questionnaire about selfstudy activities were significant.

An additional end-of-training questionnaire asked about the relative utility of various activities that take place in the classroom. The strongest students reported significantly more utility to constant correction by the teacher of their pronunciation (and conversely significantly less usefulness to teacher's withholding correction in the interests of communication). Otherwise they considered a variety of activities less useful: group work with other students, a regular routine to lessons, learning grammar patterns in context, translation into English when reading, and target-language word games.

## How Do Their Teachers Rate Them?

## Weakest Students

The weakest students were rated substantially lower on both an overall scale that compares them with other FSI students the faculty member has known and on a scale of observed language learning aptitude. Where the median percentile ratings for the whole sample are 60 and 58 , respectively, the weakest students have median scores of 34 and 32 , respectively.

## Strongest Students

Conversely, the strongest students were rated as both better students overall and in terms of observed aptitude. Their median ratings were 92 and 87 , respectively.

## Summary of Results

Table 3 in Appendix 2 summarizes the findings of this study. We see that the weakest learners relative to all other learners in the sample appear to be characterized by fewer resources (previous language learning experience and awareness of use of learning strategies), lower tolerance of ambiguity (as manifested in the HBQ and the MBTI), and much lower cognitive aptitude. The strongest learners relative to all orhers seem to be characterized by more resources from the beginning of their training, including higher level of education, more language learning experience, and a tendency to use learning strategies associated with independence. They tend to display signs of more tolerance of ambiguity, reveal greater tested cognitive aptitude, and may exhibit more emotional stability under the stress of intensive language training.

## Discussion

## Student Characteristics

The lack of previous experience characterizing weakest students and the presence of previous experience reported by the strongest suggest a "Matthew Effect" ("to those who have more shall be given"). However, the direction of causality is difficult to ascertain. Certainly, it is likely that those who have learned more languages before and to a higher level of proficiency have learned how to learn. On the other hand, it is also common for those who find learning easy to gravitate to more learning of the same
kind; foreign languages and international affairs should be no exception, so ability may also play a role from the beginning for the strongest students' involvement with language learning.

Most other studies have found that it is better to be younger, but such findings receive only limited confirmation here. Much more interesting is the fact that the mean age of the strongest $4-5$ percent of learners is thirtyfour, much older than those normally thought to be likely to learn languages. While not everyone can be a superstar, it is interesting to note further that the mean age of the entire sample, almost all of whom succeed in learning foreign languages to a high degree of proficiency in a rather short time, is thirty-nine, and roughly two-thirds of these generally successful students fall between the ages of thirty and forty-eight.

This study confirms other findings from the Language Learning Profiles Project that there is no gender difference in learning success (e.g., see Ehrman 1993, 1994). I interpret them as further supporting the statement made above in the literature review that the gender differences found in other populations may be more the result of moderating variables like education level, interests, career choice, socialization, and so on, than anything inherent to males or females (see also Meunier, this volume).

The relationship between thickness of ego boundaries on the HBQ and membership in weakest or strongest student groups is consistent with the correlational findings that are cited in Ehrman (1993) and Ehrman and Oxford (in press). These indicate that thicker ego boundaries are somewhat related to lower success rates for the sample as a whole and suggest a particular importance of external boundaries over internal boundaries in the learning of language and culture. Correlations berween the HBQ categories and speaking and reading proficiency are low-in the 20 s and 30 s . The present findings suggest that any effect of this aspect of personality is greater for the weaker students and attenuated in the middle and top; hence the low correlations. In other words, thick boundaries may be more detrimental than thin ones are helpful. Since thick boundaries can mean that information is never taken into the learner in the first place, and moreover that if it is, such information may not be integrated with other knowledge, this relative effect of thick boundaries is not too surprising (see Ehrman, 1993, for extensive discussion of this question).

In particular, one factor on the HBQ distinguishes both weak and strong students from the rest of the students: relative preference for neatness and order. Perhaps the weakest students' strong need for various kinds of order and clarity among categories, including sharp edges in visuals and clear lines of authority, may hinder their ability to cope with
the inconsistencies and lack of immediate closure entailed by communicative classrooms. In contrast, the active rejection of neatness and order by the strongest learners may suggest that too much prestructuring of material would get in their way. Other educational research findings indicate that this may be the case (Snow 1989).

The MBTI findings for weakest learners suggest a student who wants to learn only what is necessary, without the need to go beneath the surface of the material presented in class. Earlier findings showed that sensing students were most comfortable in structured, well-defined learning situations that required little in the way of making inferences or what-if imagining (Ehrman 1989; Ehrman and Oxford 1990). The subscale results are consistent with this picture: pragmatic people tend to be characterized by the words "executive, applied interests, things" in contrast to intellectual ones: "scholarly, knowledge for its own sake, words". (Saunders 1989, p. 6). Similarly, realistic people are described as "sensible, matter-of-fact" as opposed to "ingenious, enjoys play of imagination" (Saunders 1989, p. 6). Since many sensing students with the same preferences do very well in language training (hence the low correlations in Ehrman and Oxford, in press), it is likely that the key is some kind of moderating variable. Ehrman (1989) and Ehrman and Oxford (1990) suggest that the key is flexibility in adapting to different styles of teaching and cite a case of an extremely successful sensing learner who said that "intuitive-type" activities like round-robin storytelling were hard for him but that he found ways to adapt. Ehrman (1993) attempts to trace the source of the flexibility to the tolerance of ambiguity construct, operationalized at least in part by the HBQ.

Strongest learners show a tendency toward intuition (on one TDI subscale) and toward perceiving (on the MBTI main scale and on one TDI subscale). Intuitive and perceiving students enjoy abstractions, like to work out underlying systems (especially if intuitive and thinking), tend to use learning strategies characterized by the use of meaning, are often responsive to discovery learning procedures, and may experience boredom with routine and thus welcome some unpredictability in their learning experiences (Ehrman and Oxford 1989, 1990). In most work on the MBTI, intuition and perceiving are moderately correlated in the .30 s and .40 s range (Myers and McCaulley 1985); in the present sample they are correlated at .40 (Ehrman 1994), so the appearance of both together as strongest learner characteristics is not surprising (strongest learners prefer perceiving nearly significantly, $F=3.3548, d f=751, p=.0675, N=25$ out of 657 and significantly prefer the emergent [perceiving] pole of the
methodical-emergent subscale). More surprising is the fact that judging does not appear along with sensing for the weakest learners, perhaps because the orderly study habits that judgers tend to prefer (Ehrman 1989) may compensate for some of these weakest learners' cognitive shortcomings. The continuous score means for weakest learners do fall more in the judging direction, at 88 as compared to 93 for all others, though at a nonsignificant level. (MBTI continuous scores below 100 indicate extraversion, sensing, thinking, and judging.)

Other investigations of this sample have revealed a relatively strong correlation between thick ego boundaries and MBTI sensing, thinking, and judging, and conversely berween thin ego boundaries and MBTI intuition, feeling, and perceiving (Ehrman, 1993). These relationships appear in the present data as well: weak students are characterized by thick ego boundaries on a variety of categories and by sensing. In turn, strongest students have thin (external) ego boundaries on one HBQ subscale and subsignificantly tend to prefer perceiving as well as the significantly differentiating intuitive and perceiving poles on two TDI subscales.

## Affective Factors

It is not surprising that the instructors of the strongest students tend to see them as more motivated than their classmates and that these students tend to have a relatively low level of negativity about their language study, though the direction of causality is uncertain for both findings. (In other words, do students succeed because they like what they are doing, or do they like what they are doing because they are succeeding?)

Much more unexpected is the consistently lower degree of motivation across the board reported by the strongest students relative to the rest of the students in the sample. To attempt an explanation for this paradoxicalseeming finding, we can turn to the significantly lower overall affective arousal level reported by these students as well, although it correlates positively with speaking proficiency in general at $.32 p<.0001$ (see Ehrman and Oxford, in press). It is possible that in these results this group of students displays a relative coolness and resistance to the pressures of intensive language training; the strongest students may be those who respond to this normally stressful situation with more than usual emotional stability.

FSI students contrast in one significant area with university students. Horwitz and her colleagues (Horwitz 1990; Horwitz, Horwitz, and Cope 1986; Horwitz and Young 1991) found that the most debilitating form of anxiety for her students was that related to speaking in class. In contrast,
the FSI correlational data (Ehrman and Oxford, in press) show that such anxiety is facilitating (i.e., correlates with higher EOT proficiency). Although the results in this study were nonsignificant, there was a difference in the means berween strongest and other students that was clearly in the direction of more anxiety about speaking in class for the former, consistent with the correlational findings. I have speculated that the difference has to do with what the anxiety is about: FSI students tend to compete for "air time" in class, perhaps driven by the fact that the end-of-training speaking test on which pay and advancement depend consists entirely of oral interaction at an advanced level. University grades often depend less on such evaluations.

## Abilities

As mentioned above in the literature review, the MLAT appears to maintain roughly the same validity coefficient (about .50) for both the audio-lingual training in fashion when it was developed and the largely communicative training in use in most FSI programs today (Ehrman and Oxford, in press). The results reported here indicate that it is especially discriminating for students at the lowest end of the achievement continuum, but that it can also distinguish the best performers as well, though less dramatically. (Note that the weak student mean Index score is 43 vs. 63 for the rest; thus the weakest are about two standard deviations below the FSI mean of 62 . The strong student mean Index score is 68 vs. 61 for the rest, or about two-thirds of a standard deviation above the FSI mean of 62.)

These findings are consistent with the informal observations of language training supervisors at FSI over the years to the effect that the MLAT appears to be more useful at the extremes of ability than in the middle of the range. The importance of Part IV to differentiating the weakest learners is consistent with findings in universities, where students needing language waivers have been shown to be particularly weak on this subscale (Gajar 1987; Ganschow, Sparks, Javorsky, Pohlman, and BishopMarbury 1991). Parts III and IV were the MLAT subscales having the strongest correlations with speaking and reading proficiency in the present sample. Interestingly, both of these subscales have to do with subtleties of English language proficiency in vocabulary and sentence structure, respectively, thus possibly suggesting that some related form of language aptitude may also affect first language proficiency level when opportunity to learn is more or less equal.

Speculation elsewhere suggests that part of what makes the MLAT valid across methodologies is the fact that it in part tests the ability to deal
with unexpected input, part of a tolerance of ambiguity construct (Ehrman, 1993). If in fact tolerance for ambiguity-usually defined as an aspect of personality-is an important contributor to success on the MLAT, we have here an interesting intermixture of personality and aptitude. A link between personality and aptitude, at least as tested by the MLAT, is also reflected in the correlations between the MLAT Index and MBTI intuition and the intuition pole of the TDI subscale pragmaticintellectual, at $.28 p<.0001$ and $.39 p<.0001$, respectively. We have seen that the sensing-intuition scale and the pragmatic-intellectual subscale also play a role in characterizing the extreme groups in the present study.

The relative inferiority of the weakest students on the LSP simultaneous processing variable suggests that these step-by-step, analytic rather than global students may have a disadvantage in processing holistically. This weakness may interfere with their ability to cope with language when it goes beyond linear, discrete-point processing.

## Learning Strategies and Techniques

Unfortunately, too few of the weakest students turned in the end-of-training questionnaires about their learning activities for analysis. On the SILL, university students have tended to be differentiated by strategy use (Oxford and Ehrman, in press). In contrast, the SILL does not appear to distinguish among FSI learners. The absence of significant FSI results on the SILL might suggest that weakest students are not aware of any special patterns of strategy use. They may react with a kind of scattershot approach that tries anything that might work but without clear rationale, as was suggested for less adept university students (Vann and Abraham 1989). The low rate of return of the end-of-training questionnaires may reflect an overall lower level of interest in the program at the end of training. Many students who have found the language learning experience frustrating simply want to complete it as fast as possible and not dwell on it.

The strongest students appear to be significantly characterized by use of techniques related to enhancing memory and making use of instructor feedback about pronunciation. Such learners appear to take a pragmatic approach to making use of teacher feedback and maximizing at least one form of cognitive processing of what they learn. However, strongest students describe themselves as doing less of a wider range of activities than their classmates. Many of these activities are related to study alone (though at the same time, they do not find group work useful). This constellation of characteristics may be consistent with a tendency to introversion (with intuition) on the part of top achievers that appeared in a chi-square analysis of type
tables (Ehrman, 1994). Their rejection of routine lessons is possibly consistent with their preference for perceiving on the MBTI. Correlations for the two end-of-training questionnaires, not yet reported elsewhere, suggest that proficiency in both speaking and reading is related to use of a variety of unstructured input situations, especially those that involve interaction with native speakers, and that such proficiency is negatively linked to techniques that reduce risk (e.g., not moving on until a grammar point is mastered, strict routine in lessons, or making the teacher responsible for one's learning). Thus, these strong students seem to reject a number of the more limiting items that are negatively correlated with proficiency, whereas they appear to derive value from relatively independent internal manipulation of the language (e.g., hypothesis generation, mnemonics, internalizing feedback). As more questionnaires become available from the weakest students in the latest group now in training, it will be interesting to see if they tend to endorse as useful the lower risk techniques that were rejected by the strongest and were negatively correlated with proficiency.

## Teacher Ratings

The differential teacher ratings for overall quality as a student and observed aptitude provide no surprises. It may be worth notice that the strongest students are more clearly rated high than the weakest students are rated low, at least in terms of significance level. Perhaps this is a result of a charitable mind-set on the part of the faculty, who in general found it easier to rate students high than low; indeed, the sample median for overall quality as a student is the 60 th percentile and the 58 th for observed aptitude, suggesting either a volunteer effect in the sample, a lenient view by teachers, or both. (An attempt was made to eliminate volunteer effect to the degree possible by strongly encouraging all entering students in a cohort to participate in the research project, but it could not be eliminated altogether, since students could opt not to participate.)

## Toward a Broadened Definition of

## Language Learning Aptitude

There are a number of ways in which these findings are of interest. First, they seem to support the importance of tolerance of ambiguity as a key to language learning, at least in FSI classrooms. This concept is realized through the HBQ, the MBTI, possibly the LSP simultaneous processing
variable, and preference for or rejection of various learning strategies. Tolerance of ambiguity and the construct of MLAT-tested learning aptitude maintain their differentiating power despite the homogenizing influences of student preselection and strenuous efforts to find every way to help students reach their training goals once they are at FSI. (It is rare that students are dropped from training for poor performance; instead it is more likely that their training will be extended, within limits.)

By adding support to the centrality of tolerance of ambiguity, these two variable types also contribute to a model of learning that relates achievement to personality variables, beginning at the deepest level with the ego boundary distinction between thick and thin, which in turn is generally manifested by the MBTI categories (see Figure 1). These in turn may represent an approach to learning in which a particular track is favored; for example, a preference for MBTI feeling may be related to a learning track that maximizes interpersonal relations. Certain learning strategy categories are likely to be especially characteristic of one of the tracks; for instance, "judging students" often favor a well-organized and scheduled study approach. Most learners use several or all of the four tracks, but many especially favor one or two. We have seen that an apparently cognitive variable, MLAT performance, also appears to have a link to personality through the tolerance of ambiguity construct and the relative success of certain personality types on the MLAT.

These findings clarify correlational findings that were weak (in the .20 range) but appeared to be patterned, especially for the personality variables (Ehrman and Oxford, in press). The fact that some of the features showing up weakly in the correlations, particularly personality variables, are much stronger at the extremes suggests some sort of nonlinear relationship for which correlations are not the best measure.

Yet another important point is the role of tested language learning aptitude. The MLAT certainly differentiates the extreme learners from their classmates, and it is the most powerful of the variables used. It continues to retain its power in programs in which the role of rote learning has been greatly reduced from the time the MLAT was developed; perhaps the ability to manage unfamiliar and contradictory input leads both to success in communicative classrooms and to high scores on the MLAT. The MLAT may be the best of the differentiators in this study because it requires the examinee to cope with the unfamiliar on tasks that at least partially simulate language learning tasks, whereas personality inventories are asking about general life preferences, and strategy inventories do not

Figure 1
A Four-Track Model of Learning


[^0]address how the strategies are used but only whether the student is aware of using them. The significant correlations between the MLAT and the HBQ, though not strong (between .21 and .33), are consistent across HBQ and MLAT subscales (Ehrman, 1993) and with the present findings. Those between the MLAT and the MBTI are particularly telling for the sensing-intuition scale and most of its subscales, where they are significantly correlated with the intuition pole, with a range of .20 to .39. Again these are not strong, but they are consistent (Ehrman, 1994) both with each other and with the results of the present study.

The ability to learn strategically is receiving increasing attention. Important elements of strategic learning include the ability to connect new knowledge with existing knowledge, the ability to organize and elaborate it cognitively rather than simply add it to a single string of data, and to access appropriate learning strategies automatically (Jones, Palincsar, Ogle, and Carr 1987). The links between the MLAT and personality variables suggest a role for the disposition to use one's cognitive resources in ways that go beneath the surface and that establish elaborated knowledge structures. Those who are open to new material, can manage contradictions, treat their perceptions of input as hypotheses to be tested, are interested in meaning, and find ways to link new input with previous knowledge structures are advantaged in managing the highly complex tasks involved in learning a new language and culture rapidly and to a high degree of proficiency.

Thus perhaps we can broaden the usual definition of language learning aptitude to include not only the cognitive skills most clearly tested by the MLAT, but the kinds of personality attributes that predispose learners to tolerate ambiguity, to process data in elaborative ways, and to impose their own structure on what they would otherwise experience as chaos. The weakest students appear to be overwhelmed by the chaos they encounter; the strongest meet it head on, may even embrace it to a degree (MBTI perceiving), and find ways both to assimilate it to their existing schemata and to accommodate their schemata to what they select as the important elements of their experience.

The strongest learners are indeed differentiated by this nexus of personality and tested aptitude, but the weakest ones are much more clearly delineated. Perhaps this is because FSI students, at least, tend to be already selected for ability to use cognitive processing strategies of the sort that lead to success in learning even before they come to language training, so those who use these strategies even more adeptly will stand out from the main group less than those who have not developed these cognitive skills
do. In any event, it may be that although students may favor one or more of the four MBTI tracks in Figure 1, some level of ability to operate on track 2 (cognitive flexibility) is essential to success in FSI classrooms.

Conversely, the role of motivation appears to work in the opposite direction. FSI students-weak, average, and strong-tend to be highly motivated, almost to a fault (Ehrman 1990). The affective advantage the strongest learners show may be a result of an extra ability to manage their feelings, shown in their cooler approach to the task on the Affective Survey.

## Conclusions

We have seen a combination of individual difference factors that appears to differentiate extremely weak and strong learners in intensive language training. There are, however, cautions to be heeded in applying these findings.

Although a capacity to impose personal structure on chaos seems to be important, some of the influence may go the other way: that is, FSI curricula and classroom techniques have evolved over the years to meet the requirements of a certain type of student who tends to achieve well. Students in this sample appear to be quite similar to their colleagues who have entered in the past, so they doubtless benefit from training that has been developed to fit their predecessors. In other words, it may not be only that the set of aptitude attributes is characteristic of all good classroom learners; instead or in addition, FSI classrooms may have at the same time adapted to a certain profile of learner, thus increasing the advantages of those who most tend to have the features of the profile. We do not know yet to what degree the aptitude nexus that emerged from these data would advantage learners in other classroom milieux, let alone language learning in natural settings.

Needless to say, when FSI students are the subject of this kind of report, we must always ask how well we can generalize from them to the students in other classrooms. FSI students are older, better educated, generally high-functioning, and intensely motivated. They are relatively experienced learners and have already shown an interest in other languages and cultures by their very career choices. On the other hand, perhaps the combination of rested aptitude and personality that works well in FSI language training also would make the stronger students into the kind of good learner in high school and college language classes who would be attracted to a career involving other languages and cultures. Thus, the traits described in this study might well also characterize better and weaker language learners elsewhere.

Like most research studies, this report leaves work yet to be done. Statistical tests other than one-way analysis of variance (e.g., multiple ANOVA) could be done. The same study should be undertaken for reading proficiency, though based on other findings for speaking and reading from this sample, the differences can be expected to be relatively minor. The two end-of-training learning activity questionnaires need validation, and a larger number of them will permit analysis of both questionnaires for both learner extremes. Changes in the formulas used to establish the two filter variables (strongest students and weakest students) might change the results. At some point, it would be useful to examine interaction effects as well as main effects.

Perhaps the most important caution is one against overgeneralizing to individuals. The fact that one personality style, for instance, appears to have a statistical advantage in a certain learning situation does not mean that others cannot learn in that situation, and learn well. Statistics address trends, but do not describe an individual's drive, maturity, intelligence, or coping skills. Furthermore, statistics like these may be highly populationand situation-specific, thus adding weight to the need for replication and investigation of external validity.

Nevertheless, this relarively simple study has provided evidence of the value of looking at a variety of individual differences in understanding learning success and failure, at least among FSI adult students. It has led to an effort to expand the definition of language learning aptitude to include personality dispositions that interact with cognitive processing. Such findings will be even more valuable when related to variations in teacher personality and teaching techniques.

## Notes

1. This report owes much to the efforts of the staff of the FSI Language Learning Profiles Project, in particular Lucinda Hart-Gonzalez, Stephanie Lindemann, Gisela Gonzales, and Frederick Jackson, all of whom contributed to it in multiple ways. Dr. Hart-Gonzalez, Ms. Lindemann, and Julie Thornton of the Center for the Advancement of Language Learning made valuable comments on the first draft. The material herein does not represent the policy of the U.S. Department of State: it is the responsibility of the author alone.
2. One group of students ( $N=60$ ) took all the measures. Subsequent groups were given the biographic data questionnaire and the MBTI/TDI
plus from one to three of the SILL, the HBQ, and the LSP. This was done in order to reduce the burden of questionnaires. Although the Affective Survey was given only to the first group in this sample, it has been used for later samples not reported here, and the results confirm the findings that are reported here.

## Works Cited

Au, S. Y. 1988. A Critical Appraisal of Gardner's Socio-Psychological Theory of Second-Language (L2) Learning. Language Learning 38: 75-100.
Beck, R. C. 1990. Motivation: Theories and Principles. 3d ed. Englewood Cliffs, NJ: Prentice-Hall.
Beebe, Leslie M. 1983. Risk-Taking and the Language Learner. In Classroom-Oriented Research in Second Language Acquisition, edited by Herbert W. Seliger and Michael H. Long, 39-65. Rowley, MA: Newbury House.
Brown, H. Douglas. 1987. Principles of Language Learning and Teaching. 2d ed. Englewood Cliffs, NJ: Prentice-Hall.
Busch, Michael. 1982. Introversion-Extraversion and the EFL Proficiency of Japanese Language Students. Language Learning 32: 109-32.
Campbell, Christine. 1987. Survey of Attitudes Specific to the Foreign Language Classroom. Unpublished ms, Defense Language Institute, Monterey, CA.
Carroll, John B. 1990. Cognitive Abilities and Foreign Language Aptitude: Then and Now. In Parry and Stansfield, 11-29.
Carroll, John B., and Stanley M. Sapon. 1959. Modern Language Aptitude Test. New York: Psychological Corporation.
Chapelle, Carol A. 1983. The Relationship berween Ambiguity Tolerance and Success in Acquiring English as a Second Language in Adult Learners. Ph.D. diss., University of Illinois at Champaign-Urbana.
Clark, John. 1986. A Study of the Comparability of Speaking Proficiency across Three Government Language Training Agencies. Washington, DC: Center for Applied Linguistics.
Cohen, Andrew D. 1990. Language Learning: Insights for Learners, Teachers, and Researchers. New York: Newbury House/Harper and Row.

Ehrman, Madeline E. 1989. Ants and Grasshoppers, Badgers and Butterflies: Quantitative and Qualitative Investigation of Adult Language Learning Styles and Strategies. Ann Arbor, MI: University Microfilms International.
___ 1990. Owls and Doves: Cognition, Personality, and Learning Success. In Linguistics, Language Teaching, and Language Acquisition: The Interdependence of Theory, Practice, and Research, edited by J. E. Alatis, 413-37. Washington, DC: Georgetown University Press.
$\qquad$ 1993. Ego Boundaries Revisited: Toward a Model of Personality and Learning. In Strategic Interaction and Language Acquisition: Theory, Practice, and Research, edited by J. E. Alatis, 330-362. Washington, DC: Georgetown University Press
$\qquad$ 1994. The Type Differentiation Indicator and Adult Language Learning Success. Journal of Psychological Types 30: 10-29.
Ehrman, Madeline E., and Rebecca L. Oxford. 1989. Effects of Sex Differences, Career Choice, and Psychological Type on Adults' Language Learning Strategies. Modern Language Journal 73: 1-13.
$\qquad$ 1990. Adult Language Learning Styles and Strategies in an Intensive Training Setting. Modern Language Journal 74: 311-27.
$\qquad$ 1991. Affective Survey. Arlington, VA: Foreign Service Institute. In press. Cognition Plus: Correlates of Language Learning Success. Modern Language Journal.
Ely, Christopher. 1986. An Analysis of Discomfort, RiskTaking, Sociability, and Motivation in the L2 Classroom. Language Learning 36: 1-25.
1989. Tolerance of Ambiguity and Use of Second Language Learning Strategies. Modern Language Journal 22: 437-45.
Gajar, Anna. H. 1987. Foreign Language Learning Disabilities: The Identification of Predictive and Diagnostic Variables. Journal of Learning Disabilities 20: 327-30.
Galbraith, Vicki, and Robert C. Gardner. 1988. Individual Difference Correlates of Second-Language Achievement: An Annotated Bibliography. London, Ontario: University of Western Ontario.
Ganschow, Lenore, Richard L. Sparks, James Javorsky, Jane Pohlman, and Andrea Bishop-Marbury. 1991. Identifying Native Language Difficulties among Foreign Language Learners in College: A "Foreign" Language Disability? Journal of Learning Disabilities 24: 530-41.

Gardner, Robert C. 1985a. Social Psychology and Second Language Learning: The Role of Attitudes and Motivation. London: Edward Arnold.
$\qquad$ 1985b. Motivation Questionnaire. Unpublished ms. produced for use in the Language Skill Change Project, Defense Language Institute, Monterey, CA.
Guiora, Alexander A. 1972. Construct Validity and Transpositional Research: Toward an Empirical Study of Psycho-Analytic Concepts. Comprehension Psychiatry 13: 139-50.
Hart-Gonzalez, Lucinda. 1991. Self-Study Strategies Used by FSI Students. Unpublished report, Foreign Service Institute, Arlington, VA.
Hartmann, Ernest. 1991. Boundaries in the Mind: A New Psychology of Personality. New York: Basic Books.
Horwitz, Elaine K. 1985. Using Student Beliefs about Language Learning and Teaching in the Foreign Language Methods Course. Foreign Language Annals 18: 333-40.
___ 1990. Attending to the Affective Domain in the Foreign Language Classroom. In Shifting the Instructional Focus to the Learner, edited by Sally S. Magnan, 15-33. Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.
Horwitz, Elaine K., Michael B. Horwitz, and J. Cope. 1986. Foreign Language Classroom Anxiety. Modern Language Journal 70: 125-32.
Horwitz, Elaine K., and Dolly J. Young. 1991. Language Learning Anxiety: From Theory and Research to Classroom Implications. Englewood Cliffs, NJ: Prentice-Hall.
Jones, B. F., Annemarie S. Palincsar, Donna M. Ogle, and Eileen G. Carr. 1987. Strategic Teaching and Learning: Cognitive Instruction in the Content Areas. Alexandria, VA: Association for Supervision and Curriculum Development.
Keefe, James W., and John S. Monk. 1986. Learning Style Profile Examiner's Manual. Reston, VA: National Association of Secondary School Principals.
Keefe, James W., and John S. Monk, with Charles A. Letteri, Martin Languis, and Rita Dunn. 1989. Learning Style Profile. Reston, VA: National Association of Secondary School Principals.

Larsen-Freeman, Diane, and Michael Long. 1991. An Introduction to Second Language Acquisition Research. Oxford: Oxford University Press.
Lawrence, Gordon. 1984. A Synthesis of Learning Style Research Involving the MBTl. Journal of Psychological Type 8: 1-15.
Myers, Isabel B., and Mary H. McCaulley. 1985. Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.
Naiman, Neil, Maria Fröhlich, and H. H. Stern. 1975. The Good Language Learner. Toronto: Ontario Institute for Studies in Education.
Omaggio Hadley, Alice. 1993. Teaching Language in Context. 2d ed. Boston: Heinle \& Heinle.
O'Malley, J. Michael, and Anna Uhl Chamot. 1990. Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press. Oxford, Rebecca L. 1989. Strategy Inventory for Language Learning. Alexandria, VA: Oxford Associates.
___ 1990. Language Learning Strategies: What Every Teacher Should Know. New York: Newbury House/Harper and Row.
$\qquad$ . 1992. Who Are Our Students? A Synthesis of Foreign and Second Language Research on Individual Differences. TESL Canada Journal 9: 30-49.
. In press. Gender Differences in Language Learning Styles and Strategies: A Review of Existing Research. Applied Language Learning.
Oxford, Rebecca L., and Madeline E. Ehrman. 1993. Second Language Research on Individual Differences. In Annual Review of Applied Linguistics, edited by William Grabe, 188-205. Cambridge: Cambridge University Press.
. In press. Language Learning Strategies: Correlates and Outcomes for Adult Language Learners System.
Oxford, Rebecca L., Madeline E. Ehrman, and Roberta Z. Lavine. 1991. Style Wars: Teacher-Student Style Conflicts in the Language Classroom. In Challenges in the 1990s for College Foreign Language Programs, AAUSC Issues in Language Program Direction, edited by Sally S. Magnan, 1-25. Boston: Heinle \& Heinle.
Oxford, Rebecca L., Martha Nyikos, and Madeline E. Ehrman. 1988. Vive la différence? Reflections on Sex Differences in Use of Language Learning Strategies. Foreign Language Annals 21: 321-29.

Oxford, Rebecca L., and Jill Shearin. 1994. Expanding the Theoretical Framework of Language Learning Motivation. Modern Language Journal 77: 12-28.
Parry, Thomas, and Charles W. Stansfield. 1990. Language Aptitude Reconsidered. Englewood Cliffs, NJ: Prentice-Hall.
Petersen, C. R., and A. R. Al-Haik. 1976. The Development of the Defense Language Aptitude Battery (DLAB). Educational and Psychological Measurement 36: 369-80.
Reid, Joy M. 1987. The Learning Style Preferences of ESL Students. TESOL Quarterly 21: 87-111.
Saunders, David. 1989. Type Differentiation Indicator Manual: A Scoring System for Form J of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.
Schleppegrell, Mary. 1987. The Older Language Learner. Washington, DC: ERIC Clearinghouse on Languages and Linguistics.
Schmeck, Ronald. R. 1988. Learning Strategies and Learning Styles. New York: Plenum Press.
Singleton, David. 1989. Language Acquisition: The Age Factor. Clevedon on Avon, UK: Multilingual Matters.
Skehan, Peter. 1989. Individual Differences in Second Language Learning. London: Edward Arnold.
Snow, Richard. E. 1989. Aptitude-Trearment Interaction as a Framework for Research on Individual Differences in Learning. In Learning and Individual Differences, edited by Philip E. Ackerman, Robert J. Sternberg, and Robert Glaser, 13-59. New York: W. H. Freeman.
Stevick, Earl W. 1976. Memory, Meaning, and Method: Some Psychological Perspectives on Language Learning. Rowley, MA: Newbury House.
Vann, Roberta, and Roberta Abraham. 1989. Strategies of Unsuccessful Language Learners. Paper presented at the annual meeting of Teachers of English to Speakers of Other Languages, San Francisco.
Wenden, Anita L. 1991. Learner Strategies for Learner Autonomy: Planning and Implementing Learner Training for Language Learners. Englewood Cliffs, NJ: Prentice-Hall.
Wenden, Anita L., and Joan Rubin. 1987. Learner Strategies in Language Learning. Englewood Cliffs, NJ: Prentice-Hall.
Wesche, Marjorie, Henry Edwards, and Winston Wells. 1982. Foreign Language Aptitude and Intelligence. Applied Psycholinguistics 3: 127-40.

## Appendix 1

Table 1
Data on Weakest Students


Simultaneous visual processing
$199 \quad 4.3 \quad(1.1) \quad 3 \quad 3.0 \quad$ (0) $\quad 4.3550 \quad 200 \quad .04$
Modern Language Aptitude Test

| Part I | 292 | 36.5 | $(9.1)$ | 4 | 24.5 | $(6.5)$ | 6.8524 | 294 | .009 |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | :--- | :--- |
| Part II | 292 | 24.7 | $(4.5)$ | 4 | 18.5 | $(3.5)$ | 7.3634 | 294 | .007 |
| Part III | 292 | 28.3 | $(9.9)$ | 4 | 11.0 | $(8.6)$ | 12.1415 | 294 | .0006 |
| Part IV | 292 | 28.0 | $(7.5)$ | 4 | 15.3 | $(5.3)$ | 11.4289 | 294 | .0008 |
| Part V | 292 | 19.3 | $(5.3)$ | 4 | 11.5 | $(4.7)$ | 8.7868 | 294 | .003 |
| Total Score | 292 | 136.7 | $(27.5)$ | 4 | 80.8 | $(24.6)$ | 16.3881 | 294 | .0001 |
| Index Score | 339 | 62.7 | $(10.5)$ | 6 | 43.2. | $(10.8)$ | 20.5548 | 343 | .0000 |

Strategy Inventory for Language Learning
None
Faculty Ratings
None

Note: End-of-training questionnaires on preferred self-study and classroom activities had too few respondents from the weakest student group to be included here.
Abbreviations used in Table 1: Factor I (HBQ)—Internal boundaries refer to relationships internal to the individual; Factor II (HBQ)—External boundaries refer to relationship to outer us. inner world; HISP—Highest speaking proficiency in previously studied language; MLAT Part I—number learning; Part II—phonemic transcription; Part III—English vocabulary in scrambled spellings; Part IV—sensitivity to English sentence structure; Part V—paired associates: vocabulary learning; Index Score-T-score based on total.

## Table 2

Data on Strongest Students


## Table 2 continued

Data on Strongest Students

| Category | Nonstrongest Students $N$ Mean (SD) | Stronges $N$ Mean | Studen (SD) | $F$ | $d f$ | Significance |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Modern Language Aptitude Test |  |  |  |  |  |  |
| Part I | 22435.0 (9.7) | 1440.5 | (4.9) | 4.4395 | 236 | . 04 |
| Part II | 22424.3 (4.7) | 1427.2 | (2.8) | 5.2765 | 236 | . 02 |
| Part III | 22426.9 (10.2) | 1432.8 | (7.0) | 4.5701 | 236 | . 03 |
| Total Score | 224132.2 (29.6) | 14151.2 | (13.8) | 5.7291 | 236 | . 02 |
| Index Score | 26960.9 (11.2) | 19 68.2. | (5.9) | 7.8286 | 286 | . 006 |

Strategy Inventory for Language Learning
$\begin{array}{lllllllllll}\text { Memory strategies } & 166 & 2.7 & (0.6) & 11 & 3.1 & (0.4) & 6.5273 & 175 & .01\end{array}$

> End-of-Training Self-Study Activities Questionnaire None

## End-of-Training Classroom Activities Questionnaire

| Group work | 63 | 2.0 | $(0.7)$ | 5 | 1.2 | $(0.5)$ | 5.1948 | 66 | .03 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :--- |
| Reg. lesson routine | 74 | 2.3 | $(0.7)$ | 6 | 1.7 | $(0.8)$ | 4.8305 | 78 | .03 |
| Gramm. patt. in context 77 | 2.4 | $(0.7)$ | 7 | 1.7 | $(0.5)$ | 7.1121 | 82 | .009 |  |
| Trans. into Eng. for rdg 77 | 2.5 | $(0.6)$ | 7 | 1.9 | $(0.7)$ | 6.6371 | 82 | .01 |  |
| FL word games | 46 | 2.1 | $(0.87)$ | 2 | 1.0 | $(0)$ | 4.4678 | 46 | .04 |
| Reg. pronunc. correctn 76 | 2.3 | $(0.7)$ | 7 | 2.9 | $(0.4)$ | 4.4039 | 81 | .04 |  |
| Tchr not avoid correctn 78 | 1.7 | $(.07)$ | 7 | 1.0 | $(0)$ | 6.0057 | 83 | .02 |  |

Faculty Ratings
$\begin{array}{lllllllllll}\text { Extrinsic motivation } & 254 & 0.3 & (0.7) & 16 & 1.1 & (1.2) & 16.0680 & 268 & .0001\end{array}$
Abbreviations used in Table 2: Factor I (HBQ)—Internal boundaries refer to relationships internal to the individual; Factor II (HBQ)—External boundaries refer to relationship to outer us. inner world; HIRD—Highest reading proficiency in previously studied language; HISP— Highest speaking proficiency in previously studied language; MLAT Part I—number learning; Part II—phonemic transcription; Part III—English vocabulary in scrambled spellings; Part IV—sensitivity to English sentence structure; Part V—paired associates: vocabulary learning; Index Score-T-score based on Total.

## Appendix 2

Table 3
Summary of Results

| Weakest Students | Strongest Students |
| :--- | :--- |
| Less language learning experience | Younger <br> More language learning experience <br> Higher education level (hence more use <br> of learning strategies?) |
| Thick ego boundaries (inferred as low | Rejection of neatness and order (thin <br> direction) may indicate somewhat higher <br> tolerance of ambiguity) |
| mBTI sensing preference, e.g., <br> mBTI ambity and be manifested in: |  |
| sequential, preorganized, concrete, | MBreferences for imaginative and <br> emergent learning, e.g., more random, <br> unplanned, or ambiguous learning. |

(Tolerance for ambiguity as defined through the HBQ and the MBTI may be less an advantage to the strongest students than its lack is a disadvantage to the weakest.)

Less affectively aroused (motivation, anxiety)
Less negative abour learning
Subsignificant anxiety about classroom
Substantially lower cognitive aptitude (2 SD) as measured on the MLAT.

More cognitive aptitude (ca. .6-. 7 SD). as measured on the MLAT.

Tend to use SILL memory strategies. Less time spent on study activities in general. Use resources (feedback, explanations) but do not want to be routinized; tend to be independent learners and use deep processing of new material.


[^0]:    Reprinted with permission of Georgetown University Press, Ehrman 1993, p. 353.

