

Assessing Language Acquisition and Intercultural Sensitivity Development in Relation to Study Abroad Program Design

Lilli Engle

American University Center of Provence

John Engle

American University Center of Provence, Université du Sud

Study abroad professionals have attempted in recent years to bring a statistical objectivity to the evaluation of the American study abroad experience. The complexity of international education is such that it is far from easy to move towards significant, objectively measurable, and comparable outcomes. What follows is the preliminary examination of one attempt to generate and interpret meaningful statistical assessment of the study abroad experience, within the context of specifically defined study abroad program types. We will examine the data thus far generated, suggest its limitations, and appeal for a continued gathering of information. We will suggest a structured, coordinated, profession-wide assessment effort that will, we hope, gradually reveal a useful correlation between study abroad learning and the input of program variables such as duration, housing, experiential work and on-site mentoring. If, as a profession, study abroad is to invest in outcomes assessment, it would be sensible for such efforts to utilize profession-wide definitions and standards.

The study described here has taken place over eight semesters at the American University Center of Provence (AUCP), a small, independent, semester or full-year study abroad program for students with a high intermediate/advanced level of prior French study. At AUCP an institutional commitment to outcomes assessment was born of two established overseas educational goals, both of which are consciously reflected in the design of its program. Through certain aspects of its program design, AUCP hoped to create circumstances that would allow students to achieve optimal results in the acquisition of foreign language (French), and in the development of cross-cultural sensitivity and skills. Having put in place these learning and experiential strategies, the next logical step was to gather evidence to verify whether or not they had worked.

Involvement in varied forms of cultural exchange and foreign study program direction has allowed the authors to observe the American student experience abroad with close attention for almost two decades. For years, our evaluation of individual and group student experiences, and program strengths and weaknesses, relied upon self-reported, subjective methods: a written end-of-term student questionnaire; anecdotal evidence and conversation with current and past program participants; and the impressions of on-site administration, faculty, and host families. In a world that values quantified measures, however, the limitations of these methods and their credibility became increasingly clear. We turned to objective assessment tools to complement our subjective impressions and, perhaps, to call such subjective impressions creatively into question.

As stated above, the purpose of this article is to present the early findings of this assessment, and examine its varied implications with regard to the relationship between program design and the resulting student educational and cultural experience. This is not a comparative study, but a concrete example of assessing the correlation between specific program design and quantifiable results obtained via standardized testing tools. Meaningful comparative analysis will become possible only when assessment efforts are more widespread and when raw data is gathered and examined within the context of precise aspects of program design. We hope to encourage our colleagues to take similar steps, and hope that sharing our assessment results will encourage others in the profession to look more closely at existing distinctions in study abroad program design with greater specificity. We hope here to provide a practical yet comprehensive framework for the clarity of future comparative studies. Ideally, the ultimate outcome of our collective assessment efforts will be the significant improvement of the educational and cultural experience of students abroad.

AUCP Program Design

The creation and development of AUCP's program have taken place against a background of changing practices in international education. In our own observations of these tendencies, we have recognized three distinct trends. First, we have noted a decreasing level of pre-departure foreign language competence. This may well be the result of institutional pressure to increase study abroad numbers, at the expense of emphasis on pre-departure requirements, and, ultimately, foreign language acquisition. Next, we've observed a trend toward shorter program duration. Like the change in language requirements, the shrinking of program duration seems to be responding to the desire to send more students abroad, to make study abroad as widely accessible as possible. We have also observed, in program structure and design, an increased "comfort zone" for students abroad. Instead of placing students consistently in a challenging

cultural and linguistic interface with their host culture, programs have tended to create conditions enhancing the comfort and responding to the expectations of students, both culturally and materially.

These and other developments have come to pass against the backdrop of a globalized, U.S.-dominated world economy, the increasing prevalence of English as a nearly universal *lingua franca*, the daily reality of instant electronic communication, and the near omnipresence of certain realities and symbols of American pop culture. For most overseas study participants, the prospect of a genuinely challenging, creatively unsettling cultural and linguistic experience abroad has receded yearly. While the authentic experience of profound cultural and linguistic difference remains a living reality for the motivated and prepared, our point is simply that it has become increasingly easy for young American sojourners in foreign countries to live in a comfortably superficial and ultimately unchallenging relationship with their host cultures. The result abroad is less often a climate of challenge than one of comfort.

Certain components of the AUCP were designed as intentional responses to the above tendencies. This is a small, independent immersion program for advanced French learners (students with at least two years' university preparation, or the ability to take all classes in the target language). Student participants come from numerous public and private American colleges and universities, the vast majority reputed for their high admission standards. All courses are taught in French, either in-house by French faculty or at the local university (one or two classes maximum per semester). In addition, students sign a binding, enforced pledge to use only the target language in all curricular and extra-curricular circumstances. A relatively standard curriculum of French language, history, civilization, literature, and film includes a required semester-long course in French Cultural Patterns. Coursework for the latter is linked to a mandatory program of extracurricular activities designed to aid the cultural integration process. These include weekly "language partner" linguistic exchange with a local French learner of English, a personal-interest activity enjoyed within the French community (club, sports, lessons, etc.), and at least two hours of required weekly community service. All students are housed individually in *demipension* home stays with non-English speaking local families. The fall semester group typically averages 35 students, about one quarter of whom enroll for the full academic year. Spring semester characteristically enrolls more students, with up to 45 students attending.

The leading program components here—consistent use of French, coursework, required intercultural contact, guided cultural reflection, individual housing—are intended to combine to form a synchronized, harmonious whole. Working together in a coherent synergy, program components reinforce linguistic and cultural benefits, the total ideally equaling more than the sum of its parts. Linking the classroom and the

street, direct experience and abstract understanding, the program engages students in both academic and experiential learning and keeps them fruitfully occupied and focused on their linguistic and cross-cultural goals.

The central program component is the core course, French Cultural Patterns. This continuing cultural orientation extends for fifteen weeks the traditional short, front-loaded orientation session. Intended to bring to light the dynamic relationship between hidden cultural values and assumptions and the visible characteristics of culture and society, this required course addresses the central concepts of cultural awareness (e.g., time, space, high- and low-text context communication, etc.) as well as the concrete particulars of daily life as they occur, with their adversarial tensions and rewards. Experiential learning components such as the individual home stay and required community service provide students with lived situations, conflicts, misunderstandings, fears which become rich topics of discussion and collective analysis.

In the interest of future comparative studies, a clear breakdown of the AUCP program into its various components will enable eventual comparison with other programs of similar design. Outcomes of study abroad programs must be linked with the specifics of the program input to draw viable conclusions from assessment efforts. The nature and emphases of most study abroad program types can be determined by eight key components.¹ Those of the AUCP break down as follows:

1. *Length of student sojourn*: one semester or academic year
2. *Entry target language competence*: high intermediate?advanced
3. *Required language use (in class and out)*: target language
4. *Faculty*: on-site with local professors; supplementary direct university enrollment
5. *Coursework*: advanced language study; current social issues, literature, art history, etc.
6. *Mentoring, or guided cultural reflection*: intensive one-week arrival orientation program, followed by required semester-long, credit-bearing course in cross-cultural communication.
7. *Experiential learning initiatives*: required community service, personal interest activity (club, team, lessons, etc.), "language exchange" with local student
8. *Housing*: individual integration home stay

Measurable Outcomes

In the early years of AUCP program set-up and development, our evaluation methods were largely subjective. Students completed a descriptive end-of-semester questionnaire for each individual course and for the program. Indicators of program strengths or weaknesses from the questionnaire responses were supplemented by observations, discussions with former students and letters from alumni, and the impressions of on-site personnel and host families. Such assessment was largely personal—a student's or staff member's feeling that things went more or less well in relation to the student's personal goals or to the institution's mission. The only comparative dimensions were confined either to an internal administrative sense of the evolution of program quality semester to semester, or to an anecdotal evaluation of the program's accomplishment of its mission vis-à-vis that of other programs with which the staff had prior experience. It was certainly possible and satisfying to perceive clear linguistic progress on the part of students, note that relatively few left on regular weekend trips, and be told of participants' perception that their experience was distinctly different from that of other American students they met. Such assessment, however, remained difficult to seize concretely, to evaluate meaningfully, and, finally, to communicate to cooperating institutions, prospective students, and program participants themselves.

It was necessary to find appropriate testing instruments to measure both linguistic progress and the development of cultural sensitivity. The criteria for choice of such instruments were clear. Standardized, independently produced, with results impartially calculated by respected experts in the field, they should provide a reliable, subtle guide to progress in these domains and be widely available for eventual comparative use by similar institutions. As they would be administered twice to one-semester students (and three times to full-year participants), they should be relatively easy and quick to administer, and should not be inordinately expensive.

At the time of this writing, some 257 participants in AUCP's program have completed various forms of standardized pre- and post-semester testing measuring both foreign language acquisition and the development of cross-cultural sensitivity and related cultural skills. For the measurement of language acquisition, we chose the *Test d'Évaluation de Français (TEF)*. To measure objectively the acquisition of cultural skills, we turned to the Intercultural Development Inventory (IDI). The next sections present each testing instrument, its use at AUCP, and a brief presentation and preliminary analysis of results.

Outcomes Assessment in Language Learning

The Testing Instrument

The *Test d'Évaluation de Français (TEF)* is a diagnostic exam developed in 1998 by the *Chambre de Commerce et d'Industrie de Paris*. The only French language testing instrument referenced by the immigration services of the Canadian federal government, the *TEF* is used for, among other purposes, for evaluating the French level of international students applying for admittance to numerous *grandes écoles*, the most prestigious of French institutions of higher learning. It is used in more than 120 testing centers of the *Alliance Française* worldwide, and in nearly fifty universities. The following comments concern the three obligatory elements of the *TEF*. An “optional” supplementary exam, devoted to written and oral expression, also exists.

The *TEF* is administered according to strict procedures, but it can be administered at the time and place desired by program organizers. Unlike the Simulated Oral Proficiency Interview (SOPI), the *TEF* places a strong emphasis upon written and oral comprehension as well as precise understanding of vocabulary and language structures. The *TEF* instructions to the test candidate are in French, a subtle reinforcement of AUCP's emphasis upon linguistic and cultural immersion. The test itself is conceived, published, graded, and scored in the host country by a French organization whose mission is to help target individuals best able to function professionally or academically in France or the francophone world.

Targeting three major areas of language competence and attributing a numerical score and level rating to each, the *TEF* provides a nuanced profile of each candidate's overall linguistic skills. A manageable two hours and ten minutes in length, it is composed of 150 multiple-choice questions divided into three sections: 1) Written Comprehension, 2) Oral Comprehension, and 3) Vocabulary and Structure. Like the language assessment tools developed by Educational Testing Service such as the Test of English as a Foreign Language (TOEFL) or the Test of English for International Communication (TOEIC), various versions of the test exist. One semester program participants take the *TEF* during the first week and then again during the last week of their academic term abroad; full-year participants take the test three times: at the beginning and end of the fall semester, then again at the end of the spring term.

TEF Scoring

In their test results, candidates receive a raw overall quantitative score relative to 900 possible points. In addition, the overall total is broken down to reflect performance in each of the three sections of the test. Out of the 900 point total, the three

categories are weighted as follows: Written Comprehension: 300, Oral Comprehension: 360, and Vocabulary/Structure: 240. Both the overall total numerical score as well as those earned in each of the three categories are accompanied by interpretative labels: “Elementary,” “Intermediate,” or “Advanced”; each is then broken down into “high” and “low” to provide a six-level rating scale whose labels and standards have been conceived to correspond with those used by the European Council.

Upon completion of the exam, the *TEF Attestation de Résultats* provides the candidate with a personalized certificate combining a number of helpful and revealing indicators: the cumulative raw score out of the 900 possible points, raw scores and rating in each of the three test categories, as well as a customized individual commentary. Figure 1 shows pre-semester *TEF* results of a fairly typical incoming AUCP program participant.

Figure 1: Typical Test d'Evaluation de Français (TEF) score of incoming American participant

Total Score: 420 / 900	
Written Comprehension : 166 / 300	Level 3 – Low Intermediate
Can identify a document relating to a familiar area. Understands abstract elements of a document from everyday life. Understands the relationships between different elements of a document. Understands simple and clearly structured sentences.	
Oral Comprehension: 186 / 360	Level 3 – Low Intermediate
Understands both concrete and abstract descriptions. Grasps the essential elements of a message relating to a familiar area. Understands a simple and clearly expressed opinion. Understands the concrete facts of a lengthy message. Distinguishes sounds in a satisfactory manner.	
Vocabulary and structure: 68 / 240	Level 2 – High Elementary
Possesses a limited vocabulary. Can manage simple structures and some complex structures. Limited ability in finding grammatical errors.	

TEF Results and the Principle of Achievable Progress

The objective results upon which further evaluation and interpretation of linguistic achievement are made are not, strictly speaking, based upon raw numerical achievement (i.e., the overall numerical score of the second test minus that of the first). Rather, we have chosen to highlight achievable progress (AP), defined as to what extent each student bridges the gap between his or her entry-level linguistic competence and the goal of French fluency, quantified by the *TEF* as a global score of 900 points. The acquisition of linguistic skills measured as a percentage of a student's

achievable progress, we feel, better reflects the degree of individual achievement. This emphasis most notably neutralizes the “advantage” of low-entry-level learners, who have much more comparatively to learn, and may thus potentially make raw point advancement superior to that those students with a higher entry level. Such an approach has the possible further advantage of enabling eventual comparison between the results here generated and those generated in another program using a similar, if not identical, assessment instrument.

The *TEF* has been administered to all students on the program on a *pre-* and *post-* basis for, at the time of this writing, eight semesters. Two hundred and twenty-five students completing one semester of study and 32 full-year students have thus far been tested. Figures 2 and 3 present a brief summary of the results.

Figure 2: Average incoming TEF scores of 257 AUCP students

	Raw Score	Corresponding TEF Level	Average Achievable Progress
Total Score	440 / 900	Level 3 (Low Intermediate)	460 point gap
Written Comprehension	151 / 300	Level 3 (Low Intermediate)	149 point gap
Oral Comprehension	183 / 360	Level 3 (Low Intermediate)	177 point gap
Vocabulary/Structure	106 / 240	Level 2 (High Elementary)	134 point gap

Figure 3: Average end of semester TEF scores of 257 AUCP students

	Raw Score	Corresponding TEF Level	Average Raw Point Increase	Average Percent of Achievable Progress Attained
Total Score	608 / 900	Level 4 (Intermediate)	168	37%
Written Comprehension	210 / 300	Level 4 (Intermediate)	59	38%
Oral Comprehension	258 / 360	Level 4 (Intermediate)	75	42%
Vocabulary/ Structure	140 / 240	Level 4 (Intermediate)	34	25%

Preliminary Analysis

Discrepancies in Language Level Perception

Before examining the results, it is worthwhile to point out a discrepancy between the American sense of linguistic competence and that of Europeans. The home institutions with which we work, most of them leading colleges and universities, generally assume that their students have had a strong preparation in French before they depart, and that they will have excellent French skills when they return. Applying the *TEF* scale, however, we see that their arrival scores and on-site progress place them solidly in that test's Low Intermediate to High Intermediate range. Such a discrepancy is understandable in the light of the significantly heavier language preparation in European secondary schools and universities, and a cultural imperative for practical foreign language competence, based in geography, history, and economics. For most Europeans foreign language is perceived as a necessary tool—and less as a nicety or cultural embellishment, as it often is in the United States. The *TEF* scoring scale reflects this cultural difference.

TEF Scores and Achieved Progress: One Semester

TEF initial testing figures are generally consistent with what observation, prior academic performance, and in-house testing tools indicate about the French language competence of students upon arrival. Similarly, an average one-semester gain of 37% in personal achievable progress confirms the administrative and teaching staff's subjective sense of the generally substantial language-learning on the part of students. Discrepancies between *TEF* quantified results and our perceptions of student progress are rare.

Over eight semesters, the results from the 257 students tested have demonstrated an average one-semester personal gain of 37% in overall personal achievable progress. The eight semester averages from Fall 2000 to Spring 2004 vary as follows: F00: 42% / S01: 34% / F01: 36% / S02: 37% / F02: 39% / S03: 30% / F03: 41% / S04: 39%. Because of the consistency of the overall averages over the years, we have come to consider the attainment of 40% of overall achievable progress as confirmation of what we subjectively perceive to be a "good" semester. The pattern of these overall averages also has provided us with a clear indication of the degree of linguistic progress we can reasonably expect from AUCP students in keeping with elements of its program design.

Consistency in the rate of progress measured on an individual student basis has also been a valuable and encouraging result of our use of the *TEF* assessment tool. As expected, those students with lower levels on arrival tend to make more gain in raw points of progress. But using the percentage of achievable progress indicator, test

results show consistency in the measurement of the average group achievement and that of individual students. In eight semesters of *TEF* testing, 57% of the students achieved personal progress within ten percentage points, higher or lower, of the final program-wide average of 37%. In short, *TEF* results confirm what we subjectively perceive to be true, that the majority of students move forward in their linguistic progress at a fairly consistent, predictable rate.

Linguistic Precision and Language Learning

An interesting characteristic of student *TEF* scores is the consistently higher scores in the Oral and Written Comprehension sections compared to those in the Vocabulary and Structure section. In every semester's testing, incoming *TEF* results, end-of-semester scores, and percentages of gain in achievable progress in the three test sections have confirmed this tendency. Even those students with the highest overall entry scores frequently show a score one or two levels weaker in the Vocabulary and Structure section.

These results have followed what seems to be a pattern of relatively steady decline, notably in competence testing on students' arrival. Compared with the 240 point *TEF* ideal, average entry level scores in this category showed a steady downward pattern over seven consecutive semesters: 127 – 117 – 113 – 102 – 90 – 88 – 83; in the process they have declined from the *TEF*'s Low Intermediate to its High Elementary level. In Spring 2004 this sequence was broken, with an average entry score of 103; while an improvement, this score nevertheless remains within the High Elementary category.

Yet these students have been identified stateside as advanced in their French ability. The test itself may explain this discrepancy. The clear weakness visible here is in the "Vocabulary and Structure" section of the test, the area focusing on precise word definition and meaning, and details of grammar and structure. Unlike those parts of the *TEF* addressing comprehension, questions in Vocabulary and Structure demand verbal precision.

We suggest that the consistently weak scores in this area may be the result of the relatively low value placed upon linguistic precision in the prior language training of our students. Acceptable language learning standards on many home campuses emphasize task-oriented communication over linguistic precision. That a student makes herself understood, albeit approximately—or sincerely tries to—is considered acceptable. Our preliminary *TEF* results show that the same students who arrive abroad with Vocabulary and Structure as their weak areas, make an average of only 25% achievable progress in these areas after one semester of study, as opposed to roughly 40% achievable progress in Written and Oral Comprehension.

Outcomes Assessment in Intercultural Sensitivity

The Testing Instrument

The Intercultural Development Inventory (IDI) is a multiple-choice testing tool developed in 1998 by Mitchell Hammer and Milton Bennett. The IDI draws on Bennett's own Developmental Model of Intercultural Sensitivity (DMIS), and its breakdown into six stages of developmental evolution, from an ethnocentric to an ethno-relative world view, labeled Denial, Defense, Minimization, Acceptance, Adaptation, Integration. Generated by contemporary leaders in the field and correlated for validity with such instruments as the Worldmindedness Scale and Intercultural Anxiety Scale, this assessment instrument is well-reputed, widely used, easily administered, and independently evaluated. The IDI measures an individual's or group's fundamental worldview orientation to cultural difference, and thus the individual or group's capacity for intercultural competence. As a theory-based test, the IDI meets the standard scientific criteria for a valid and reliable psychometric instrument.

The current version of the IDI is composed of fifty statements with which the candidate must choose to either to disagree or agree on a five point scale. The test is completed in twenty minutes and can be scored in-house with a specialized computer program or sent to the distributors for scoring.

AUCP one-semester program participants took the IDI twice: once during the first week of the program and again during the last week of their term abroad. Full-year students took the inventory three times: twice with the fall semester students, then again during the last week of the spring term.

IDI Scoring

The Inventory and the method of scoring reflect the underlying assumption of the DMIS, that as one's experience of cultural difference becomes more complex, one's potential competence in intercultural interactions increases. In the quantification of the overall results, scores can vary between an absolute low of 5 to a perfect 25, when issues with cultural difference in all five of the IDI categories have been fully "resolved." IDI results present the candidate with an individual profile of his or her worldview, broken down into five categories or scales:

1. DD scale: combines the Denial and Defense stages of the DMIS, and indicates a worldview that simplifies and/or polarizes cultural difference.
2. R scale: indicates a worldview that reverses "us" and "them" polarization, where "them" is superior.

3. M scale: refers to the Minimization stage of the DMIS or a worldview that highlights cultural commonality and universal issues.
4. AA scale: combines the Acceptance and Adaptation stages of the DMIS, and indicates a worldview that can comprehend and accommodate complex cultural differences.
5. EM scale: measures “encapsulated marginality,” characterized by feelings of cultural alienation in which one’s worldview incorporates a multicultural identity with confused cultural perspectives.

Supplementary graphs and numerical values bring more nuance to these overall results, examining more closely where potential problems lie. For example, an unresolved issue with Minimization can be shown to be related to a worldview which has a tendency to assume that people from other cultures are basically “like us” (as revealed by the so-called “similarity cluster” of IDI questions) or which has a tendency to apply one’s own cultural values to other cultures (as revealed by the “universalism cluster” of questions).

IDI Results

The IDI has been administered for this assessment study for a total of six semesters to date, to a total 187 one-semester AUCP participants.

The AUCP IDI results confirm that students, who have already taken two years of a foreign language, and have chosen to study abroad in a full-immersion program, will show promising incoming results as to their openness and intercultural sensitivity. Consistently, incoming IDI scores show that, on average, students attending this program arrive with a score of 19, a relatively small 6-point gap separating them from the perfect score of 25.

The question remained as to whether the students, given their pre-disposition and given the program components to which they are exposed, would progress or regress in their cross-cultural sensitivity during their time abroad.

As for the *TEF* scores, we evaluated the IDI results in terms of each individual student’s achievable progress (AP), an approach which is particularly appropriate since the IDI concerns personal development as opposed to absolute knowledge. Over the six semesters of testing, student groups attained on the average 33% of collective achievable progress, with average group scores varying from semester to semester within a range of 20% to 39%. Figure 4 shows the IDI achievable progress results for the 187 students tested.

A close look at the IDI scores shows that an average of 14% of the student population decline in their cross-cultural competence as measured by the IDI. This rather disappointing statistic is counterbalanced by the fact that 52% of students tested made between 30% and 100% of their achievable progress on the IDI scale.

Figure 4: IDI scores showing percentage of achievable progress attained

Semester	Total students	Decline	0-9% Gain	10-29% Gain	30-49% Gain	50-100% Gain
Fall 2001	32	8	2	4	9	9
Spring 2002	31	3	0	6	9	13
Fall 2002	24	3	3	5	8	5
Spring 2003	33	1	3	5	8	16
Fall 2003	33	5	5	12	10	2
Spring 2004	33	7	2	15	6	3

Preliminary Analysis

We consider such results as largely positive. However, without data from comparable programs available and without a control group of students based at the home campus, these IDI assessment results take on interpretive meaning within the framework of program variables from semester to semester, and by the subjective/objective comparison of individual student profiles.

The IDI is based on the belief that cross-cultural competence is related to the development of a *complex* worldview construct. Competence is thus related to an ability to observe, respect, and adapt to cultural difference. In the past years, we have observed in our American students tendencies which jeopardize progress in developing intercultural sensitivity. Two tendencies most often to limit progress, leading sometimes to outright regression in their intercultural sensitivity. Both can be seen as intellectual coping mechanisms that some students put in place when faced with the disturbing reality of unsettling cultural difference.

Often American students come abroad secure in their cross-cultural skills because they have been culturally conditioned to accept a world of ethnic and racial diversity. The problem arises abroad when the domestically conceived superficial differences of race and ethnic origin have been mistaken for genuine cultural difference. Thus, when abroad, and in contact with true cultural differences, some as simple as “excessive” displays of emotion or a relaxed attitude toward work, some students will judge the action as shocking or backward or inappropriate. Any such tendency to judge other cultures according to a culturally pre-conditioned set of values and assumptions will activate the “us”-“them” polarization which is typical of the less-evolved, ethnocentric Defense stage as defined by the DMIS.

Likewise, when students are protected abroad or, for some personal reason do not wish to focus on cultural difference, the desire for comfort dominates. During their time abroad, these students will reinforce worldviews of Minimization or even Denial as defined by the DMIS, thus reinforcing their ethnocentric worldview and impeding their progress along the IDI scale.

Two factors lead to the clear development of cross-cultural competence in the American student group: as much direct, authentic contact with the host culture as possible, and skillful mentoring which guides, informs, inspires, and stimulates the experiential learning process. Neither factor is easy to implement, and students seeking comfort will fight against the initiative itself. We hope that the extended use of the IDI in study abroad program assessment will bring to light this aspect of the causal relationship between program design input and student outcomes.

TEF and IDI Results: Correlations?

The measurable results thus far yielded by the pre- and post-semester *TEF* and IDI testing of this small sample group hint at encouragingly meaningful patterns yet finally raise more questions than they answer. The results of parallel administration of the *TEF* and the IDI suggest, at first, a striking correlation. The average 37% gain in personal achievable progress in language learning is matched by a 33% rise in IDI scores. Such similar results suggest a close correlation between student progress in these two central areas, one that most overseas study directors would find rewarding.

Closer examination, however, reveals considerable disparity. First, the average IDI score is the aggregate of individual scores that widely vary. The 187 students who sat for the IDI, for example, had scores ranging from a disturbing 53% decline through an unrealistic and nearly as troubling 100% gain in personal achievable progress. This wide variation in individual *IDI* scores suggests that the reaction to cultural difference is a very complex matter being influenced by numerous elements of prior conditioning, both personal and cultural.

A comparison with *TEF* results further highlights the amplitude of the range of IDI scores within a single student group. For both testing instruments, individual student scores were examined for their degree of variance from the overall average rate of achievable progress. For the *TEF*, 57% of the students achieved personal progress within ten percentage points, higher or lower, of the final program-wide average of 37%. Only 24% of those who took the IDI achieved personal progress similarly close to the final average of 33% gain.

We went on to seek correlation on an individual student basis. For this aspect of our inquiry, individual *TEF* scores were compared with the same individuals' corresponding IDI results. Those participants whose IDI scores fell within the range of 10%

above or below the *TEF* score, expressed as the percentage of personally achieved progress, were considered to have significantly correlated results. Of the 187 students in question, approximately 42% achieved scores within this same general range. While such results suggest a certain moderate level of coherence between individual rates of linguistic and cultural progress, they fail to confirm the consistency suggested by the similarity in the group average scores. This tendency towards fluctuation invites a close examination of individual and group results; at the very least, it seems important not to accept without question the nearly identical group average *TEF* and *IDI* progress rates.

Full-Year Participants and the TEF

To this point we have administered the *TEF* to 32 students who have attended the program for a full academic year, on an average of 8 per year over four years. As a small sample group, the results yielded and their subsequent provisional interpretation should thus be considered no more than preliminary steps towards a statistical understanding of the full-year students' experience. While the fall semester test results of these students are included in the overall results for one semester participants discussed above, we do not include full-year student achievement in the statistics for the spring semesters. We have examined the results of our full-year students in two ways: by comparing their one-semester achievement with that of their fall semester peers and by comparing their second semester achievement with that of their own first semester results.

As stated above, on average, one-semester program participants increase their raw *TEF* score by 170 points and fulfill 37% of their achievable progress according to the test's 900-point scale. Within this group, full-year students increase their raw score by 192 points in their first semester and thus bridge a superior 41% of their "fluency gap." Such encouraging results are perhaps unsurprising and may indicate that, given their longer-term plans, full-year students are more committed to language learning and more likely to plan for its longer-term benefits than their one-semester peers.

However, results show a comparative decline in the level of achievement for the second semester of study. On the average the same students increase their second-semester *TEF* score by only 17% of their remaining achievable progress. Thus, for the two semesters they make an average total of 45% of their achievable progress in French language acquisition, compared with the average one-semester collective achievement of 37%. In other words, in twice the time they make much less than twice the progress.

On-site, on a purely subjective level, we perceive profound change in most full-year students. In comparison with one-semester participants, most show deeper personal development, interiorizing cultural rhythms and growing in terms of what we might call emotional intelligence. This growth is perhaps reflected in our early statistical outcomes in the area of cultural sensitivity as measured by the *IDI*. The fact remains,

however, that the objectively-measured language learning curve for full-year students we have tested slumps significantly in the second semester.

Certain explanations for this statistical disparity suggest themselves. In the first fifteen weeks of their stay abroad, for example, most students experience a round-the-clock exposure to the language that is completely new to them. Not surprisingly, they make extensive, rapid progress in the basic vocabulary to which they are most intensely exposed: French home, meals, classroom, public transportation, and various social settings. The mastery of grammar and linguistic structures previously prepared in the classroom setting would suddenly take on new immediacy and utility, especially for students required to use the language in all settings and stimulated by the home stay context and extra-curricular required cultural contact. We suggest that, for students abroad, linguistic progress abroad may, by its very nature, be “front-loaded,” rising significantly in the first weeks and months, and then leveling off dramatically.

Such reasoning only partially accounts for this state of affairs, however. We speculate that a deeper explanation for the tailing-off of linguistic progress abroad may be cultural in origin. Most students make significant progress in their first fifteen weeks in France to the point at which they become at least superficially comfortable with the language. It is apparent from observation that, unlike their abilities in the days immediately after their arrival, after fifteen weeks, program participants can communicate with reasonable efficiency in most social situations. For most students, roughly successful communication is enough. To progress beyond this point, and to arrive at truly precise, subtle foreign language expression, means for most language learners a new, significant effort of concentration and attention. If students are more or less comfortable in their language usage, and can make themselves understood, they may become complacent.

Above we discuss the consistently weak *TEF* results in test sections demanding the exact command of detail, as well as the goal of functional competence at the heart of American language teaching. Both may be indicators of an American cultural reluctance to insist upon linguistic precision, moving beyond simply “making yourself understood.” The pragmatic strain in American culture suggests continually to young Americans that language is something fundamentally utilitarian. Theorists of culture speak of the American communication style’s emphasis upon the practical, bottom-line delivery of a message at the expense of contextualization. When such cultural conditioning is linked with the achievement of a certain superficial comfort and practical ease with French, it is not surprising that our full-year participants abroad tend to achieve less progress in the language in the second half of their stay.

While we naturally welcome all forms of concrete outcomes assessment complementing this study, we would be particularly eager to see second-semester language acquisition results from students enrolled in similarly-structured immersion programs.

The anomaly our very limited results suggest may dissolve in a larger sea of contrasting figures, but this discrepancy may also find itself confirmed. In that case, one line of inquiry, as we suggest above, may be that of ingrained American cultural attitudes towards language and its forms of use.

Full-Year Participants and the IDI

Full-year participants take the IDI three times, once upon arrival, a second time at the end of the fall semester, and a third time at the end of the spring term. They do not, during the spring term, take the Cultural Patterns course. Despite the small number of full-year students tested (25 to this point), IDI results suggest that full-year program participants make significantly more progress than others in areas of cultural understanding and cross-cultural communication and that their rate of progress increases significantly in the second term. These results serve at least to bring a certain numerical credence to the personal and cultural development we observe in second semester students.

In their second semester abroad the average rate of full-year participants' personal improvement in the *TEF* declines from the 41% achieved in the first term to 17%. On the IDI however, these students achieved an average 28% of their achievable progress in the first semester. While this is slightly less than the 36% average AP achieved by single-semester participants, these full-year students went on to achieve an average of 40% of their remaining AP in their second semester on-site.

These results are a small step towards the objective confirmation of the progress in cultural adaptation and in complexity of worldview that we observe on-site. These figures prevent U.S. from asserting any correlation between the rate of second-semester student progress in cultural sensitivity (40% total AP) and language acquisition (17% AP). For the sake of the legitimacy of future comparisons, we must emphasize that this disparity in progress in language and culture learning concerns a student group that has already achieved functional competence in the target language.

Looking Ahead

Admittedly limited in scope, this effort to measure learning outcomes abroad objectively and engage in preliminary analysis of the results constitutes, at the very least, a beginning. While immediately useful in the understanding of student outcomes at the host institution, this program-specific study offers the groundwork for future cross-program comparison.

Assessment results will never carry much meaning if they are not put in context. In this study we have tried to portray in precise terms the nature and design of the program in question, so that the specifics of the input could be established before

examining the quality of the outcomes. As more students are tested, clearer, more predictable patterns will emerge that will help us determine the continuity between educational goals abroad, program design, and the results obtained. Subsequent efforts to adjust program design for improved outcomes may well bring to light the greater or lesser effectiveness of certain program components.

By assessing outcomes from programs of otherwise similar design, the true weight of program components in the acquisition of linguistic and/or cultural skills may at last become statistically evident. When such outcomes efforts become widespread and are clearly linked to precise elements of program design both in individual programs and in international education as a whole, with meaningful comparison and a clear sense of what is being compared, we will be better able to understand what contributes to students' experience, its success, its failure, or the thousand gradations separating the two, and to adjust program design and focus in order to help students realize the personal and academic goals that bring meaning and purpose to education abroad.

Note

¹ For a more complete articulation of the issue of classifying study abroad programs, see our in-depth discussion in "Study Abroad Levels: Toward a Classification of Program Types." *Frontiers: The Interdisciplinary Journal of Study Abroad*, Fall 2003: 1-20.

References

Hammer, M. R., & Bennett, M. J. (2001). *The Intercultural Development Inventory (IDI)*. Portland, OR: Intercultural Communication Institute.