Validity of Global Self-Ratings of ESL Speaking Proficiency Based on an FSI/ILR-Referenced Scale

by

Kenneth M. Wilson

in collaboration with
Ronald Lindsey
TOEIC Language Consulting, SA
Switzerland
ABSTRACT

This study explored the validity and potential utility for self-assessment of speaking proficiency in English as a second language (ESL), of a self-rating scale intended to nominally parallel—in terms of number of points or levels and corresponding descriptions of language-use behavior—a scale used by members of the Interagency Language Roundtable (ILR) to rate performance in Language Proficiency Interviews (LPI). The study sample was composed primarily of educated, adult ESL users/learners, most of whom were native-speakers of either French or German, who provided self-assessments just prior to taking the Test of English for International Communication (TOEIC), in employment-related ESL training contexts in Switzerland. Correlations between self-ratings and TOEIC scores centered around .7, and the mean ILR-referenced self-rating was very close to that for "predicted ILR-scaled LPI rating", where prediction was based on a previously developed equation for predicting LPI rating from TOEIC score. Lines of inquiry warranting further research are suggested.

Key words: TOEIC, Self-rating, ILR-referenced, speaking proficiency
ACKNOWLEDGEMENTS

This study is based on data generated in operational ESL proficiency assessments in Switzerland, involving operational administrations of the TOEIC and a global self-rating schedule, initiated and conducted by Ronald Lindsey, TOEIC Consulting SA, director of TOEIC testing operations in that country. At Mr. Lindsey's initiative, the English-language version of the self-assessment schedule, originally developed by staff members of the Educational Testing Service (ETS), was translated into both French and German, thus permitting use of French- and German-language versions as well as the English-language version in the study. Mr. Lindsey was directly responsible for all data collection and provided essential detail regarding the data and the study context. In addition, he reviewed and commented helpfully on various drafts of the study report. Without Mr. Lindsey's initiative and his very full collaboration, this study would not have been possible. Essential indirect support was provided by Educational Testing Service, especially the ETS Research Division. ETS colleagues, Anne Harvey and Ken Sheppard provided helpful reviews of the original draft. These direct and indirect contributions are acknowledged with appreciation.
INTRODUCTION

"... (M)ost of us could give a pretty accurate appraisal of our own abilities in other languages that we know" (Upshur, 1975: p. 329).

"... (M)ost learners (of another language) have a certain capacity for determining their own language ability--provided they have at their disposal a measuring standard by which they may express their intuitions" (Oskarsson [Oscarson], 1980: p. 13).

"One of the potentially most useful areas for research is in self-rating and self-assessment ... As yet, however, there have been few attempts made to develop self-rating instruments or to examine systematically the problems involved" (Ingram, 1985: p. 268).

"A number of important findings, both methodological and substantive, arose from this study (involving self-ratings of oral language proficiency). Perhaps most significant is the finding that it is possible to design a questionnaire that (language) teachers can fill out in twenty minutes and in which they can rate their own oral language proficiency with high validity and reliability" (Hilton, Grandy, Kline, and Liskin-Gasparro, 1985: p. 37).

"Although self-assessment has been prevalent for a number of years in such fields as psychology, sociology, business, and so on, its use in second language teaching/learning has remained rare. (However, based on our findings) . . ., at least under the conditions described, self-assessment must be considered a very valuable tool as a placement instrument" (LeBlanc and Painchaud, 1985: p. 673).

The foregoing comments by several professionals concerned with second-language teaching/learning/assessment, based on their clinical experience, suggest that interest in developing and evaluating (e.g., studying the validity and/or utility of) formal self-assessment (SA) "instruments" has not been widespread in the field of second-language proficiency assessment, but that given appropriate stimulus materials (e.g., SA instruments) users/learners of a target language are capable of providing potentially useful self-assessments of the level of proficiency at which they are functioning in the language.

The analyses reported herein were undertaken to extend inquiry into the validity of self-ratings of speaking proficiency in English as a second language (ESL), based on global, self-rating schedules. The schedules involved were designed to nominally parallel (with respect to number of levels and types of linguistic behavior at each level) the well-known oral language proficiency scale developed during the 1950s by the Foreign Service Institute (FSI) of the U.S. Department of State for use in rating behavior elicited in Language Proficiency Interviews (LPI). The speaking scale and conceptually comparable FSI-developed scales for rating second-language reading, listening and writing skills) were subsequently adopted for use by a number of U.S. governmental agencies.
known collectively as the Interagency Language Roundtable (ILR).\footnote{For historical perspective and detail regarding procedures used for the direct assessment of oral language proficiency, including the corresponding global rating scale, see Lowe and Stansfield (1988: passim); ETS (1982); Jones (1979); Clark (1978: passim). The scale does not appear to have been adapted for self-assessment purposes by the FSI or other U. S. government agencies (Ehrman, 1995, personal communication).}

The formal, "quasi-absolute" (after Carroll, 1967), ILR scale for speaking proficiency (as well as the scales for other macroskills) provides detailed descriptions of linguistic behavior at six basic levels (Level 0, Level 1, . . ., Level 5) ranging from "no proficiency" (Level 0) through "equivalent to educated native speaker" (Level 5) and at each of the corresponding intermediate (in-between) levels (Level 0+, Level 1+, . . ., Level 4+). Thus, the formal ILR scale involves 11 behaviorally defined levels of proficiency (see Appendix A for formal level descriptions for speaking proficiency).

Level-descriptions in the scale employed in the present study\footnote{The scale was originally developed as a collaborative, ETS internal enterprise, circa 1980-81, by John L. D. Clark, Steven A. Stupak, and Protase Woodford, then ETS staff members in the area of test development.} were adapted from the formal FSI/ILR (or ILR) scale descriptions with a view to maintaining "functional comparability" between (a) modified (that is, self-referenced, less formal and metalinguistic, paraphrased, briefer) behavioral descriptions to be used for self-assessment, and (b) the corresponding formal descriptions for the basic and intermediate levels referred to by professionals when rating performance in Language Proficiency Interviews.

The SA scale--as written originally in English--was introduced as an exploratory measure in a study (Wilson, 1989) that was concerned primarily with assessing concurrent relationships between formal LPI ratings and scores on the Test of English for International Communication \cite{TOEIC} (e.g., ETS, 1986),\footnote{Numbers refer to notes at end of text. See Endnote 1 for detail regarding the TOEIC, a norm-referenced ESL proficiency test developed by the Educational Testing Service and used primarily in corporate or other ESL assessment settings outside the United States, reporting a total score and separate section scores for listening comprehension and reading, respectively. Study data were collected by the representative TOEIC agency in Switzerland (TOEIC Consulting SA), Ronald Lindsey, Director.} in a sample of native-speakers of French--employees in ESL-essential positions with an international accounting firm, tested on-site in Paris. In 1994, the SA scale was translated into French and used with a sample of students enrolled in ESL training in the Paris area, who also took the TOEIC and the Test of English as a Foreign Language or TOEFL (e.g., ETS, 1991). Findings of unpublished analyses involving the second sample were consistent with those reported for the first sample studied which were interpreted as indicating that ILR-referenced self-ratings appeared to have sufficient validity to be used for research purposes: for example, for the purpose of estimating levels of ESL speaking proficiency in contexts in which Language Proficiency Interviews could not readily be conducted.
Implicit in the use of a self-assessment schedule written in a target-language is an assumption that the subjects involved have attained levels of reading skills in that language that are at least "minimally sufficient" for them to read and comprehend the content of the scale--and the latter appears to have been true for the initial French sample noted above. In any event, logical considerations suggest that some of the ambiguities inherent in self-assessments of target language proficiency can be obviated by using scales written in subjects' native language(s).

For the present study, French and German translations of the original English language version, as well as the English version, were used. All three versions excluded descriptions for levels beyond Level 3+, because few ESL users/learners in the TOEIC testing context are expected to exhibit ESL speaking proficiency at the higher levels represented in the ILR scale (see Appendix B for versions used in the present study; cf. formal descriptions in Appendix A). Global self-ratings of speaking proficiency based on the these rating scales were found to be relatively strongly related to scores on the TOEIC, in samples of educated, adult ESL users/learners involved in operational TOEIC assessments in Switzerland--primarily native-speakers of either French or German.

Before describing the present study in detail, attention is directed, first, to a review of previous studies--all that could be located in a search of related literature--involving conceptually similar, but abbreviated ILR-referenced self-assessment schedules which included descriptions corresponding only to the ILR basic or base levels; the two previous applications of the scale employed in the present study (the expanded ILR-referenced scale described generally above) are then reviewed. The present study is then described in detail.

Generally speaking, findings involving the particular ILR-referenced scale under consideration in the present study, as well as those of other studies cited in this report that have used conceptually similar, albeit less comprehensive, ILR-anchored global self-assessment schedules, appear to warrant the following general conclusion, offered here as a strong working proposition regarding the validity of ILR-referenced self-assessments (and other self-assessment models which involve graded, language-use descriptions [e.g., Clark, 1981; LeBlanc and Painchaud, 1985]):

Given descriptions of behavior that validly reflect different functional levels of proficiency in a second language as foci for organizing and expressing their intuitions, educated adult users/learners of that language are able to rate their own level of proficiency with a significant degree of validity.\(^d\)

Again, this general proposition appears to have been accepted as plausible a priori or intuitively by linguistic experts, buttressed by empirical findings, as suggested by the illustrative

---

\(^d\) The scale employed in this study describes levels of speaking proficiency. Conceptually comparable rating schedules are also available for rating aspects of proficiency other than speaking (e.g., reading, writing, listening). These scales do not appear to have been widely adapted for self-assessment purposes. However, generally related findings cited herein suggest that this working proposition might well be sustained for self-assessments of macroskills other than speaking.
Review of Studies Involving Abbreviated ILR Scales

Several studies involving nominally ILR-parallel self-rating scales with behavioral descriptions corresponding only to ILR basic levels have been conducted under general ETS auspices (Clark and Swinton, 1979; Clark, 1981; Hilton, Grandy, Kline and Liskin-Gasparro with Stupak and Woodford, 1985).

Clark and Swinton (1979), in the preliminary phase of a study concerned with development of measures of speaking proficiency in the TOEFL (Test of English as a Foreign Language: e.g., ETS, 1985) context, collected self-ratings in a linguistically heterogeneous sample of students in an intensive ESL-training program at Columbia University. Subjects provided self-assessments "... after reading over the FSI verbal descriptions for levels 0-5 ..." The level-descriptions were relatively formal and metalinguistic. The resulting self-ratings correlated moderately ($r = .48$) with external ratings based on formal interviews by LPI interviewer/rater trainees.

For the main study, Clark and Swinton (1979) used a modified version of the scale described generally above. The main-study scale, shown in Table 1, included "in-between" options not provided in the preliminary study. However, correlations involving the scale in Table 1 with other study variables were not reported. The self-assessment (SA) scales used in both the pilot study and the main study were written in English, a factor that theoretically may tend to have attenuating effects on the validity of self-assessment by non-native speakers of English who may be below some "threshold level" of ESL reading proficiency. However, the mean self-rating (2.4) reported for the main study sample was quite close to the mean LPI rating (2.2), indicating relatively good agreement between the two means.

As part of a national survey of U. S. college students' knowledge and beliefs regarding other cultures (Barrows, Ager, Bennett, Braun, Clark, Harris and Klein, 1981), Clark (1981) conducted a substudy designed to assess the validity and utility of a slightly modified version of the scale shown in Table 1, as well as other types of scales for obtaining self-assessments of speaking, reading and/or listening proficiency in the foreign language in which students indicated they were "most proficient."

Study variables included the following:

(a) the global FSI/ILR-parallel self-assessment scale shown in Table 1, but with no "in-between" options,

(b) a conceptually comparable scale for assessing reading proficiency,

(c) "can do" (Likert-type) scales for reporting relative ease or difficulty of performing each of a number of specific language-use tasks involving speaking, reading, and listening comprehension skills, respectively,
Table 1

"LPI-Parallel" Global Self Assessment Schedule for Speaking Ability in English, Developed by Clark and Swinton (1979) for Use with ESL TOEFL Candidates

INSTRUCTIONS

As part of the study, we would like to ask you to judge your own level of speaking ability in English. Below are six paragraphs. Please read each paragraph carefully and decide which paragraph best describes your ability to speak English. Please be as honest and as accurate as possible. This information will NOT affect your test score. If you feel your speaking ability is in-between two of the descriptions (in other words, better than one description but not as good as the next description), mark the box between the two descriptions. Mark one of the boxes below.

Level 0 [ ]. My speech in English is limited to a few words and I have great difficulty understanding English, even when it is spoken very slowly. I cannot really communicate any information in the language.

LEVEL 0+ [ ]. (My ability is in between these two descriptions).

LEVEL 1 [ ]. I can ask and answer questions about very familiar subjects and can understand simple questions and statements if they are spoken slowly and sometimes repeated. My vocabulary is limited to basic needs (food, asking directions, greeting people, and so forth). I make many grammatical mistakes but my teachers usually understand me. I can order food in a restaurant, get a room at a hotel, ask directions on the street, and introduce myself to people.

LEVEL 1+ [ ]. (My ability is in between these two descriptions.)

LEVEL 2 [ ]. I can talk with native speakers of English about myself and my family, my job or studies, hobbies, and current events. I can understand most conversations in English except when the speech is very fast. My grammar is fairly good but I make mistakes with complicated constructions. If I do not know the word for a particular thought or object, I can usually describe it by using other, easier words.

LEVEL 2+ [ ]. (My ability is in between these two descriptions.)

LEVEL 3 [ ]. I can understand almost everything spoken by native English users. My vocabulary is good enough that I usually know most or all of the words for what I want to say.

My grammar is good and any mistakes I make are usually with the more complicated constructions. I can pronounce English clearly but do not have a perfect accent.

LEVEL 3+ [ ]. (My ability is in between these two descriptions.)

LEVEL 4 [ ]. I can understand native English speakers even when they are speaking quickly and using sophisticated or colloquial expressions. My vocabulary is very extensive, even for technical
LEVEL 4, continued

matters, and I can talk fluently and accurately about almost any subject with which I am familiar. I make only a very few grammatical errors and my pronunciation is very good but not completely native.

LEVEL 4+ [ ]. (My ability is in between these two descriptions.)

LEVEL 5 [ ]. My speech is exactly like that of an educated native speaker of English.

Note. The behavioral descriptions in Clark (1981) were identical to those above (Clark and Swinton, 1979) as were those used in Hilton et al. (1985), except for designated target language. The other scales are not repeated here. However, Clark (1981) did not include the "in-between" option shown above. The Clark and Swinton scale descriptions were not numbered nor were the FSI-parallel levels (e.g., LEVEL 5) indicated. The latter are noted in this table to indicate the FSI/ILR level targeted by each description. Note that the behavioral descriptions occasionally include references to listening comprehension, reflecting the inherent (developmentally interdependent) linkage between listening comprehension and speaking proficiency.

(d) scores for Listening Comprehension and Reading Comprehension from subtests included in the Modern Language Association-Cooperative Foreign Language Tests (ETS, 1965), also called "MLA-Cooperative ... Tests", and

(e) FSI-ratings for speaking and reading proficiency, respectively.

Table 2a shows the intercorrelations and descriptive statistics reported by Clark (1981) for these variables. Regarding these data, Clark commented in part as follows:

"Observed correlations range from .462 (FSI-type self-appraisal of reading ability vs. MLA-Cooperative listening score) to .791 ("can do" reading vs. "can do" speaking scales). . . . For reasons previously discussed, extremely high correlations were not anticipated between the self-rating scales and the (MLA-Cooperative) tests, for which the test content and testing format are considered to be at some the considerably more complex FSI-type descriptions" (Clark, 1981, p. 34).

The scale items used by Clark are shown in Table 2b, along with an illustrative distribution of responses (that for the sample of college seniors surveyed). These same items and formal Language Proficiency Interviews (but no gobal self rat-ing scales) were subsequently used by Carlson, Burn, Useem and Yachimowicz (1990), to assess effects on language proficiency
Table 2a. Intercorrelations and Descriptive Statistics for Designated Variables: From Barrows et al. (1981)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Global SA Speak</th>
<th>Global SA Read</th>
<th>&quot;Can Do&quot; scales Speak</th>
<th>&quot;Can Do&quot; scales Listen</th>
<th>&quot;Can Do&quot; scales Read</th>
<th>FSI Interview</th>
<th>MLA-Coop Listening</th>
<th>MLA-Coop Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak (SA)</td>
<td>1.00</td>
<td>.63</td>
<td>.73</td>
<td>.75</td>
<td>.72</td>
<td>.59</td>
<td>.60</td>
<td>.56</td>
</tr>
<tr>
<td>Read (SA)</td>
<td>.63</td>
<td>1.00</td>
<td>.63</td>
<td>.61</td>
<td>.72</td>
<td>.54</td>
<td>.46</td>
<td>.51</td>
</tr>
<tr>
<td>&quot;Can do&quot; scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak (SA)</td>
<td>.73</td>
<td>.63</td>
<td>1.00</td>
<td>.76</td>
<td>.79</td>
<td>.63</td>
<td>.59</td>
<td>.56</td>
</tr>
<tr>
<td>Listen (SA)</td>
<td>.75</td>
<td>.61</td>
<td>.76</td>
<td>1.00</td>
<td>.78</td>
<td>.58</td>
<td>.59</td>
<td>.53</td>
</tr>
<tr>
<td>Read (SA)</td>
<td>.72</td>
<td>.72</td>
<td>.79</td>
<td>.78</td>
<td>1.00</td>
<td>.64</td>
<td>.55</td>
<td>.60</td>
</tr>
<tr>
<td>Interview rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(FSI/ILR scale)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSI interview</td>
<td>.59</td>
<td>.54</td>
<td>.63</td>
<td>.58</td>
<td>.64</td>
<td>1.00</td>
<td>.72</td>
<td>.71</td>
</tr>
<tr>
<td>MLA-Coop Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>.60</td>
<td>.46</td>
<td>.59</td>
<td>.59</td>
<td>.55</td>
<td>.72</td>
<td>1.00</td>
<td>.78</td>
</tr>
<tr>
<td>Reading</td>
<td>.56</td>
<td>.51</td>
<td>.56</td>
<td>.53</td>
<td>.60</td>
<td>.71</td>
<td>.78</td>
<td>1.00</td>
</tr>
<tr>
<td>Mean</td>
<td>2.94</td>
<td>4.06</td>
<td>34.40</td>
<td>25.50</td>
<td>17.10</td>
<td>2.09</td>
<td>171.9</td>
<td>171.60</td>
</tr>
<tr>
<td>SD</td>
<td>.89</td>
<td>.72</td>
<td>.64</td>
<td>.50</td>
<td>.31</td>
<td>.90</td>
<td>11.8</td>
<td>13.14</td>
</tr>
<tr>
<td>Adjust mean</td>
<td>1.94</td>
<td>3.06</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>(2.09)</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Note. N = 142 for FSI Interview; between 300 and 320 for other variables. Coefficients were reported to three decimal places in the original report.

1 In-between FSI interview ratings (e.g. 0+, 1+ . . . 4+) were coded by adding .7 to the lower of the two levels involved.

2 The six-point self-report scale had values ranging between 1 and 6, inclusive, corresponding to FSI scale Levels 0-5.

associated with study abroad, in a sample of American undergraduate students. Correlations between self-ratings and the external measure in that sample were generally comparable to those reported by Clark (1981) for samples of U.S. undergraduate students--see Table 2a, above.

Based in part on the evidence of validity for self-assessments reported by Clark (1981), Hilton et al. (1985) used essentially the same global rating scale and similar "can do" scales in the preliminary, "validation" phase of a national survey designed to assess the oral language proficiency
of teachers of French and Spanish, respectively. In the preliminary study, an assessment was made of interrelationships among formal Language Proficiency Interviews, global self-ratings according to the LPI-parallel schedule shown in Table 1, and self-ratings according to "can do" scales (modified for use with language teachers as opposed to "college seniors generally") for assessing listening comprehension and speaking ability, respectively, in samples of high school teachers of either French \([N = 108]\) or Spanish \([N = 113]\).

Table 3 shows intercorrelations for global self-rating, the listening comprehension scale and the speaking scale, respectively, for the Spanish teachers (above the diagonal) and the French teachers (below the diagonal). For the global self-rating, correlations with LPI rating were quite consistent (.66, .69, and .68) for the French, Spanish, and combined samples, respectively. And, the distribution of LPI-parallel self-ratings was very similar to the corresponding distribution of actual LPI ratings, as is clearly indicated in Figure 1.3

From the foregoing it seems evident that after reading and reflecting upon the descriptions in the schedule, the foreign-language teachers in these samples were able to place themselves at about the same level as did their professionally trained interviewers/raters.4

Commenting on the validity of the global self-rating, Hilton et al. (1985: p. 24) had the following to say:

"Surprisingly, the interview rating correlated most highly with the single global self-rating. The reason this is unexpected is that it is a single item and therefore would not be expected to have very high reliability. We would expect the scale scores (involving 23 and 11 items, respectively) to correlate most highly. The fact that the global self-rating correlated so highly is probably because it is formatted so similarly to the ILR scale, and the teachers were able to rate themselves quite objectively" (p. 24).

Based on these findings, Hilton et al. (1985) decided to use the global scale for the main study. They established equivalencies between self-ratings and LPI ratings, and used the ILR-referenced SA item as a research surrogate for difficult-to-obtain LPI ratings in their national survey.

The Study Scale: Previous Findings

The ILR/FSI-referenced self-rating schedules used in the studies reviewed above,5 provided behavioral descriptions only for the basic FSI levels (Level 0, Level 1, . . ., Level 5).

---

5 We have cited all studies involving similarly designed nominally FSI-parallel global self-rating schedules that could be located in reviewing relevant published sources (e.g., journals, books) and formal organizational research reports (various ETS program publications, and so on). Further evidence of validity for self-ratings based on such schedules is also provided in an internal document prepared at the Iowa State University (Abraham, Plakans, Koehler, and Carley, 1988), reporting results of a study involving ESL proficiency assessment and training of nonnative-English speaking teaching assistants at Iowa State University. See Endnote 5 for a brief summary.
Table 2b. "Can Do" Scales for Speaking, Reading, and Listening Comprehension in Most Proficient Foreign Language (MPFL): Performance of College Seniors Surveyed (from Clark, 1981)

"Can do"* Speaking

<table>
<thead>
<tr>
<th>&quot;Can do&quot; Speaking</th>
<th>%</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Say the days of the week</td>
<td>56.5</td>
<td>28.9</td>
<td>14.5</td>
</tr>
<tr>
<td>b) Count to 10 in the language</td>
<td>93.1</td>
<td>5.7</td>
<td>1.1</td>
</tr>
<tr>
<td>c) Give the current date (month, day, year)</td>
<td>49.3</td>
<td>33.9</td>
<td>16.8</td>
</tr>
<tr>
<td>d) Order a simple meal in a restaurant</td>
<td>33.4</td>
<td>36.3</td>
<td>30.3</td>
</tr>
<tr>
<td>e) Ask for directions on the street</td>
<td>31.1</td>
<td>33.3</td>
<td>35.6</td>
</tr>
<tr>
<td>f) Buy clothes in a department store</td>
<td>26.2</td>
<td>28.6</td>
<td>45.2</td>
</tr>
<tr>
<td>g) Introduce myself in social situations, and use appropriate greetings and leave-taking expressions</td>
<td>45.4</td>
<td>31.5</td>
<td>23.1</td>
</tr>
<tr>
<td>h) Give simple biographical information about myself (place of birth, composition of family, early schooling, etc.)</td>
<td>29.8</td>
<td>30.7</td>
<td>39.5</td>
</tr>
<tr>
<td>i) Talk about my favorite hobby at some length, using appropriate vocabulary.</td>
<td>11.8</td>
<td>22.2</td>
<td>66.1</td>
</tr>
<tr>
<td>j) Describe my present job, studies, or other major life activities accurately and in detail</td>
<td>12.1</td>
<td>22.5</td>
<td>65.4</td>
</tr>
<tr>
<td>k) Tell what I plan to be doing 5 years from now, using appropriate future tenses</td>
<td>10.6</td>
<td>19.0</td>
<td>70.4</td>
</tr>
<tr>
<td>l) Describe the United States educational system in some detail</td>
<td>7.2</td>
<td>16.3</td>
<td>73.5</td>
</tr>
<tr>
<td>m) State and support with examples and reasons a position on a controversial topic (for example, birth control, nuclear safety, environmental pollution)</td>
<td>5.0</td>
<td>10.2</td>
<td>84.7</td>
</tr>
<tr>
<td>n) Describe the role played by Congress in the United States government system</td>
<td>3.5</td>
<td>10.4</td>
<td>86.2</td>
</tr>
</tbody>
</table>

"Can do" reading item

<table>
<thead>
<tr>
<th>&quot;Can do&quot; reading item</th>
<th>%</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Read personal letters or notes written to me in which the writer has deliberately used simple words and constructions</td>
<td>55.4</td>
<td>29.8</td>
<td>14.8</td>
</tr>
<tr>
<td>b) Read, on store fronts, the type of store or the services provided (for example, &quot;dry cleaning,&quot; &quot;bookstore,&quot; &quot;butcher,&quot; etc.)</td>
<td>53.8</td>
<td>33.2</td>
<td>13.0</td>
</tr>
<tr>
<td>c) Understand newspaper headlines</td>
<td>41.1</td>
<td>40.6</td>
<td>18.3</td>
</tr>
<tr>
<td>d) Read personal letters and notes written as they would be to a native speaker</td>
<td>17.1</td>
<td>34.2</td>
<td>48.8</td>
</tr>
<tr>
<td>e) Read and understand magazine articles at a level similar to those found in Time or Newsweek, without using a dictionary</td>
<td>9.4</td>
<td>21.5</td>
<td>69.1</td>
</tr>
<tr>
<td>f) Read popular novels without using a dictionary</td>
<td>6.7</td>
<td>12.6</td>
<td>80.7</td>
</tr>
</tbody>
</table>
**Reading, continued**

<table>
<thead>
<tr>
<th>Description</th>
<th>&quot;Can do&quot;*</th>
</tr>
</thead>
<tbody>
<tr>
<td>g) Read newspaper &quot;want ads&quot; with comprehension— even when many abbreviations are used</td>
<td>5.8 15.8 78.4</td>
</tr>
<tr>
<td>h) Read highly technical material in a particular academic or professional field with no use or only very infrequent use of a dictionary</td>
<td>2.6 9.4 88.0</td>
</tr>
</tbody>
</table>

*INSTRUCTIONS: Listed below are a number of "can do" statements about speaking ability [listening][reading] in the MPFL. Please read each description carefully and circle the appropriate number to indicate whether, at the present time, you would be able to carry out each task "quite easily" (1), "with some difficulty" (2), or "with great difficulty or not at all" (3). Circle only one response for each task.*

<table>
<thead>
<tr>
<th>Description</th>
<th>&quot;Can do&quot;*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>&quot;Can do&quot; listening comprehension</strong></td>
<td></td>
</tr>
<tr>
<td>a) Understand very simple statements or questions in the language (&quot;Hello,&quot; &quot;How are you?&quot;, &quot;What is your name?&quot;, &quot;Where do you live?&quot;, etc.)</td>
<td>88.9 9.3 1.7</td>
</tr>
<tr>
<td>b) In face-to-face conversation, understand a native-speaker who is speaking slowly and carefully (i.e., deliberately adapting his or her speech to suit me)</td>
<td>48.7 36.6 14.7</td>
</tr>
<tr>
<td>c) On the telephone, understand a native speaker who is speaking to me slowly and carefully (i.e., deliberately adapting his or her speech to suit me)</td>
<td>37.5 31.9 30.6</td>
</tr>
<tr>
<td>d) In face-to-face conversation with a native speaker who is speaking slowly and carefully to me, tell whether the speaker is referring to past, present or future events</td>
<td>33.4 34.8 31.8</td>
</tr>
<tr>
<td>e) In face-to-face conversation, understand a native speaker who is speaking to me as quickly and as colloquially as he or she would to another native speaker</td>
<td>7.8 23.8 68.4</td>
</tr>
<tr>
<td>f) Understand movies without subtitles</td>
<td>9.2 27.5 63.3</td>
</tr>
<tr>
<td>g) Understand news broadcasts on the radio</td>
<td>7.3 23.2 69.4</td>
</tr>
<tr>
<td>h) On the radio, understand the words of a popular song I have not heard before</td>
<td>5.4 26.6 68.0</td>
</tr>
<tr>
<td>i) Understand play-by-play descriptions of sports events (for example, a soccer match) on the radio</td>
<td>5.9 15.7 78.4</td>
</tr>
<tr>
<td>j) Understand two native speakers when they are talking rapidly with one another</td>
<td>6.1 14.5 79.4</td>
</tr>
<tr>
<td>k) On the telephone, understand a native speaker who is talking as quickly and as colloquially as he or she would to another native speaker</td>
<td>5.4 11.3 83.3</td>
</tr>
</tbody>
</table>
Table 3. Intercorrelations of Interview Rating, Global Self-Rating, and Scores on the Listening Comprehension and Speaking Scales (from Hilton et al., 1985, p. 23)

<table>
<thead>
<tr>
<th>Variables</th>
<th>LPI rating</th>
<th>Global Self-Rate</th>
<th>Listening scale</th>
<th>Speaking scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish (above diagonal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPI rating</td>
<td>--</td>
<td>.69</td>
<td>.57</td>
<td>.53</td>
</tr>
<tr>
<td>Global Self-rate</td>
<td>.66</td>
<td>--</td>
<td>.73</td>
<td>.69</td>
</tr>
<tr>
<td>Listening scale(^a)</td>
<td>.55</td>
<td>.71</td>
<td>--</td>
<td>.70</td>
</tr>
<tr>
<td>Speaking scale(^b)</td>
<td>.49</td>
<td>.64</td>
<td>.73</td>
<td>--</td>
</tr>
<tr>
<td>French (below diagonal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Sum of responses to 23 "can do" statements such as those shown in Table 1b, above (from Clark, 1981).

\(^b\) Sum of responses to 11 listening-related statements such as those shown in Table 1b, above (from Clark, 1981).

Figure 1. Distributions of self-ratings of speaking proficiency in language taught by U.S. teachers of French and Spanish, respectively, are comparable to corresponding distributions of LPI ratings according to the ILR/LPI scale.
The scale used in this study involved adaptations of formal ILR descriptions corresponding to the five intermediate as well as the six, basic levels. Illustrative excerpts of modified scale descriptions corresponding to arbitrarily selected FSI/ILR Levels (Level 0+ through Level 2) are provided in Table 4—see Appendix B for detail.

The two previous exploratory uses of the scale, as noted at the outset, involved native-French speakers in samples selected for operational assessments conducted by representatives of the Test of English for International Communication (TOEIC). In the initial assessment, the English-language version was administered in a study (Wilson, 1989) concerned primarily with evaluating concurrent relationships between formal Language Proficiency Interview (LPI) rating and scores on the TOEIC in a sample of native-French speaking ESL users/learners (N = 56) employed by an international accounting firm in Paris. A French-language version of the scale was administered in 1994, also as an exploratory measure, in an operational TOEIC assessment involving a sample (N=114) of students 20 to 24 years of age, enrolled in an intensive ESL program in the Paris area; no language proficiency interviews were conducted for this sample.

Table 5 summarizes descriptive statistics and intercorrelations of variables for the respective samples; results for the combined sample are shown in Table 6. Several outcomes appear to be noteworthy, including the following:

(a) in both samples, self-ratings correlated moderately with TOEIC scores,

(b) in the 1989 sample (data shown above the diagonal), for which LPI ratings were available, correlations involving self-rating and TOEIC scores were similar to those for the formal LPI rating and TOEIC scores,

(c) mean self-rating was lower for the 1994 sample, with lower TOEIC means, than for the 1989 sample, with higher means on the TOEIC—corresponding self-rating means of 1.8 and 2.8, respectively,

(d) in the 1989 sample, the mean, LPI-scaled self-rating was somewhat higher than expected based on TOEIC scores, whereas the opposite was true for the second sample,

During the course of reanalysis of data for the original sample for the present study, two miscoded LPISELF values were discovered and corrected. The coefficients involving Self Rating in Table 3 reflect outcomes after correction for miscoding. The coefficients originally reported were .64, .50 and .63 for TOEIC listening comprehension (LC), reading (R) and Total, respectively (from Wilson, 1989: Table 10). Thus the present findings differ only in detail from the originally reported findings; interpretive inferences are substantively the same for both.
Table 4

General Descriptions in the Self Rating Schedule Designed to Correspond to Designated ILR Basic and Intermediate Levels: ESL Speaking Proficiency *

0+. My English is limited to memorized words and phrases.
I ask questions only when I think I know the answer or when the answer is YES or NO. I am not able to use grammar.
   I am able to satisfy immediate survival needs using memorized material. There are long pauses in my speech and I must rely on single words and phrases . . .. I can understand sounds when they are isolated, but when they are in words and groups of words, I have a very difficult time understanding them . . . .

1. I am able to use limited English grammar. I must translate everything I say, but I can make sentences.
I know enough English to survive in the language, if put in an English-speaking situation.
   I can . . . order a meal and get a room in a hotel. I can ask and answer simple questions, ask for directions, respond to simple statements, and maintain a very simple face-to-face conversation . . ..

1+. I am able to speak English without translating every sentence. I know some past tense and future tense verbs but I am uncertain in their use. I feel comfortable in informal social situations, if not asked to speak very much.
   In conversation, I speak in short sentences, thinking about almost every one before peaking.
   I have difficulty producing certain sounds, but generally my speech is understood . . ..

2. I can discuss my work, home life, current events, hobbies, likes and dislikes in English. I can utter sentences in series, without having to pause, and can narrate and describe in past, present and future. My control of basic grammar, including past tense verbs, is good.
   I can discuss matters that are of a concrete nature with some confidence . . .. I can give detailed directions on how to get from one place to another . . .. In simple conversations, I rarely have to mentally translate what I am going to say.

* Each general level description (each numbered paragraph, above) is shown without modification (the ILR level-designations are not included in the operational scale), but only illustrative excerpts from the detailed elaborative comments are included. See Wilson (1989: Appendix B) for the full schedule, and Appendix B herein for levels 0+ thru 3+ only.
Table 5. Descriptive Statistics and Intercorrelations of Variables in the Initial (Wilson, 1989) French Sample (N = 56, Above Diagonal) and the Second (1994) French Sample (N = 114, Below Diagonal), Respectively

<table>
<thead>
<tr>
<th>TOEIC score</th>
<th>1989 sample above diagonal (N=56)</th>
<th>1994 sample below diagonal (N=114)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>R</td>
</tr>
<tr>
<td>TOEIC LC</td>
<td>--</td>
<td>.73</td>
</tr>
<tr>
<td>TOEIC R</td>
<td>.72</td>
<td>--</td>
</tr>
<tr>
<td>TOEIC Tot</td>
<td>.96</td>
<td>.89</td>
</tr>
<tr>
<td>RealLPI</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Self-Rate</td>
<td>.52</td>
<td>.53</td>
</tr>
<tr>
<td>PredLPI</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

1994 sample below diagonal (N=114)
Mean 345.6 318.9 664.5 n.a. 1.8 2.0
SD 78.1 49.8 119.2 n.a. .43 .27

a TOEIC LC and R scores, respectively, are reported on a standard scales ranging between 5 and 495; scores on TOEIC Total, which is the simple sum of the two section scores, range between 10 and 990.

b PredLPI is a predicted LPI value, based on TOEIC Total, using a regression equation previously developed in a general sample (N = 393) that included native-speakers of Japanese, Arabic and Spanish, respectively, as well as the 56 native-speakers of French here under consideration. “*” indicates coefficients not shown because they are identical with those for TOEIC Total (q.v.).

Table 6. Descriptive Statistics and Intercorrelations of Variables in the Combined (1989 and 1994) French Sample (Total N = 170)

<table>
<thead>
<tr>
<th>TOEIC score*</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>R</td>
</tr>
<tr>
<td>TOEIC LC</td>
<td>--</td>
<td>.79</td>
</tr>
<tr>
<td>TOEIC R</td>
<td>--</td>
<td>.92</td>
</tr>
<tr>
<td>TOEIC Total</td>
<td>--</td>
<td>.68</td>
</tr>
<tr>
<td>Self-Rating</td>
<td>--</td>
<td>2.1</td>
</tr>
</tbody>
</table>

* See note to Table 5, above, for scale values.
THE PRESENT STUDY

The present study extends analyses designed to shed light on the validity and potential usefulness of self-ratings according to the expanded ILR-parallel self-rating schedule described above, by assessing concurrent relationships between self ratings and TOEIC scores in samples of ESL users/learners involved in operational, on-site testing sessions conducted independently in Switzerland, under the auspices of TOEIC Language Consulting, SA.

Context of Data Collection

In the TLC operational assessments, self-ratings are obtained in order to give the ESL trainees, who are being evaluated primarily in terms of their performance on an external measure (the TOEIC), an opportunity to compare a functional description which they themselves perceive to represent their level of proficiency, with the functional description that on experience, reflects average expectancy for individuals with TOEIC scores similar to their own.6

In order to ensure that differences in ESL reading proficiency would not complicate the rating process for the majority of examinees (native-speakers of either French or German), the original English-language version of the self-rating schedule (Wilson, 1989; also Table 4, above) was translated into French and German versions. Because ESL users/learners taking the TOEIC seldom exhibit speaking proficiency above Level 3+, the self-rating schedules excluded descriptions corresponding nominally to ILR levels above Level 3+. Copies of English, French and German versions of the self-assessment scale developed for operational use in the TOEIC-Swiss setting, are shown in Appendix B.

Generally speaking, trainees who name either French or German as the native language are assigned a self-assessment schedule written in either French or German--thus removing possible interpretive complications associated with use of a scale written in a target language (English in this instance). ESL trainees who indicate that they are more proficient in either French or German than in English, use either the French or German version. The English version is assigned to those who report that they are more proficient in English than in either French or German, as second languages. Self-ratings and TOEIC scores typically are obtained on the same testing occasion in each testing site, and candidates complete the self-assessment just prior to taking the TOEIC.

Study Sample and Data

For the present study, TOEIC scores, self-ratings, and information regarding native-language and scale-version (French, German, Other; French, German, or either French, German or
English) were provided by TOEIC Consulting SA, for a total of 937 adult, ESL users/learners enrolled in diverse corporate-sponsored or other ESL training programs, during 1993-1994 and 1994-1995. Based on candidate self-report, the great majority were native-speakers of either French (677) or German (155); some 23 other native-language groups were represented by at least one trainee (total Other = 105). Distributions by self-reported native language and year of data collection are provided in Table 7.

Analytical Rationale and Procedure

Language Proficiency Interviews were not conducted so it was not possible to address directly questions regarding degree of agreement between ILR-scaled, LPI ratings and self-ratings based on the nominally ILR-parallel rating schedule. Instead, attention was focussed on an indirect assessment which involved evaluating "parallelism" in patterns of findings involving TOEIC scores and self rating in the Swiss sample, on the one hand, and patterns involving TOEIC scores and LPI rating in previously studied samples (e.g., Wilson, 1989; Wilson and Stupak, 1998; Wilson and Chavanich, 1989; Wilson, Komarakul & Woodhead, 1997), on the other hand.

To permit evaluation of "parallelism" in findings, the analyses undertaken in the study were designed primarily to shed light on the following questions:

1. Does the pattern of concurrent correlations involving TOEIC scores and the ILR-scaled self rating tend to parallel patterns that have been found to obtain between TOEIC scores and ILR-scaled LPI rating?

2. What is the extent of agreement between levels of ESL speaking proficiency as indexed, on the one hand, by self-assigned proficiency levels on a scale designed specifically to nominally parallel the ILR-scale used to rate performance in Language Proficiency Interviews and, on the other hand, by ILR-scaled levels that typically are assigned by professional interviewers/raters as descriptive of LPI performance for individuals with similar scores on the TOEIC--according to empirically validated, regression-based guidelines for inferring LPI rating from TOEIC scores in linguistically diverse samples (e.g., Wilson, 1989; Wilson & Chavanich, 1989; Wilson, Komarakul & Woodhead, 1997; Wilson & Stupak, 1998).

3. Do answers to these questions tend to be similar for subgroups classified according to native language: scale language--French: French, German: German, Other: French/ German/ English?

Analytical Procedures and Related Findings

To address the first question, above, descriptive statistics and intercorrelations of the primary study variables (TOEIC LC, TOEIC R, TOEIC Total [LC+R], and self rating) were computed for the total study sample (N = 937). These are shown in Table 8a. Corresponding statistics for TOEIC scores and formal LPI rating (from Wilson and Stupak, 1998), are shown in Table 8b. Both tables also show standard partial regression weights reflecting results of
### Table 7. Distribution of Subjects by Self-Reported Native Language and Year of Data Collection

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Arabic</td>
<td>9</td>
<td>1.0</td>
<td>7</td>
</tr>
<tr>
<td>Bengali</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Chinese</td>
<td>4</td>
<td>.4</td>
<td>2</td>
</tr>
<tr>
<td>Croatian</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Dutch</td>
<td>3</td>
<td>.3</td>
<td>3</td>
</tr>
<tr>
<td>French</td>
<td>679</td>
<td>2.1</td>
<td>496</td>
</tr>
<tr>
<td>French (Vietnam)</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>German</td>
<td>157</td>
<td>16.7</td>
<td>142</td>
</tr>
<tr>
<td>Hungarian</td>
<td>2</td>
<td>.2</td>
<td>1</td>
</tr>
<tr>
<td>Indonesian</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Italian</td>
<td>21</td>
<td>2.2</td>
<td>16</td>
</tr>
<tr>
<td>Japanese</td>
<td>3</td>
<td>.3</td>
<td>2</td>
</tr>
<tr>
<td>Norwegian</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Polish</td>
<td>5</td>
<td>.5</td>
<td>4</td>
</tr>
<tr>
<td>Portuguese</td>
<td>10</td>
<td>1.1</td>
<td>6</td>
</tr>
<tr>
<td>Romanian</td>
<td>2</td>
<td>.2</td>
<td>1</td>
</tr>
<tr>
<td>Spanish</td>
<td>28</td>
<td>3.0</td>
<td>19</td>
</tr>
<tr>
<td>Swedish</td>
<td>2</td>
<td>.2</td>
<td>2</td>
</tr>
<tr>
<td>Thai</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Turkish</td>
<td>4</td>
<td>.4</td>
<td>3</td>
</tr>
<tr>
<td>Urdu</td>
<td>1</td>
<td>.1</td>
<td>1</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>3</td>
<td>.3</td>
<td>2</td>
</tr>
<tr>
<td>Hebrew</td>
<td>1</td>
<td>.1</td>
<td>-</td>
</tr>
<tr>
<td>Serbo-Croatian</td>
<td>1</td>
<td>.1</td>
<td>-</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>.1</td>
<td>-</td>
</tr>
</tbody>
</table>

---

Total 942 100.0

Total 713 100.0

Total 229 100.0
Table 8a. Intercorrelations and Descriptive Statistics for TOEIC Scores and ILR-Scaled Self Rating for ESL Proficiency, in the Study Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>TOEIC score*</th>
<th>Self Rating</th>
<th>Mean</th>
<th>Sd</th>
<th>Beta coeff.**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>R</td>
<td>Total</td>
<td>Self Rating</td>
<td></td>
</tr>
<tr>
<td>TOEIC LC</td>
<td>1.000</td>
<td>.884</td>
<td>.973</td>
<td>.748</td>
<td>309.7</td>
</tr>
<tr>
<td>TOEIC R</td>
<td>1.000</td>
<td>.968</td>
<td>.698</td>
<td>.68</td>
<td>275.4</td>
</tr>
<tr>
<td>TOEIC Total</td>
<td>1.000</td>
<td>.746</td>
<td></td>
<td></td>
<td>585.1</td>
</tr>
<tr>
<td>SELF RATING</td>
<td>1.000</td>
<td></td>
<td>1.76</td>
<td>.68</td>
<td></td>
</tr>
</tbody>
</table>

* The standard score scales for LC and R, respectively, range between 5 and 495; TOEIC Total is the simple sum of the two section scores and ranges between 10 and 990.

** Standard partial regression coefficient in regression of Self Rating on TOEIC LC and R.

---

Table 8b. Intercorrelations and Descriptive Statistics for TOEIC Scores and ILR-Scaled LPI Rating in Previously Studied Samples (from Wilson & Stupak, 1998, Table 2): TOEIC/LPI Sample (N=1045)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>TOEIC score</th>
<th>LPI Rating</th>
<th>Mean</th>
<th>Sd</th>
<th>Beta coeff.**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>List</td>
<td>Read</td>
<td>Total</td>
<td>LPI Rating</td>
<td></td>
</tr>
<tr>
<td>TOEIC LC</td>
<td>1.000</td>
<td>.787</td>
<td>.945</td>
<td>.708</td>
<td>295.6</td>
</tr>
<tr>
<td>TOEIC R</td>
<td>1.000</td>
<td>.946</td>
<td>.614</td>
<td>.614</td>
<td>266.8</td>
</tr>
<tr>
<td>TOEIC Total</td>
<td>1.000</td>
<td>.699</td>
<td></td>
<td></td>
<td>562.4</td>
</tr>
<tr>
<td>LPI Rating</td>
<td>1.000</td>
<td></td>
<td>1.70</td>
<td>.68</td>
<td></td>
</tr>
</tbody>
</table>

* The TOEIC/LPI sample (N=1045) includes native-speakers of French (N=56), Japanese (N=285), Spanish-Mexico (N=42), Arabic -Saudi Arabia (N=10), Thai (N=583) and Korean (N=149), respectively. For detail regarding TOEIC/LPI relationships in the respective samples, see Wilson, 1989 for detail regarding the French, Japanese, Spanish and Arabic speakers; Wilson & Chavanich (1989) & Wilson, Komarakul & Woodhead (1997) for detail regarding the Thai samples; and Wilson & Stupak (1998) for detail regarding the Korean sample (and findings suggesting that the regression-based guidelines used in the present analysis for estimating LPI Rating from TOEIC scores may need to be modified for native-speakers of Korean).

** Standard partial regression coefficient in regression of LPI Rating on TOEIC LC and R.
regressing the pertinent rating (either self rating or LPI rating) on TOEIC LC and R in the respective samples.

The patterns of concurrent correlation between TOEIC scores and the self-rating (in Table 8a) appear to be very similar to the patterns for TOEIC scores and LPI ratings (shown in Table 8b).

More specifically, the findings in these tables suggest that,

a) TOEIC vs. criterion (self rating or LPI rating) correlations are of generally similar magnitude,

b) coefficients for TOEIC LC are larger than those for TOEIC R with respect to both criterion variables,

c) the LC coefficient is just slightly larger than that for TOEIC Total in each table, and

d) when LC and R were treated as separate predictors in regression analyses, the standard partial regression weights for LC and R were very similar (indicating that LC contributed relatively more than did R to prediction regardless of the criterion under consideration). By inspection, the descriptive statistics for the ILR-scaled self rating (mean = 1.67, Sd = .68), and those for the ILR-scaled LPI rating (mean = 1.60, Sd = .68), are also very similar; and the TOEIC score levels in the two samples are also quite similar (e.g., TOEIC Total scores of 585 and 562 for the study sample and the previously studied TOEIC/LPI sample).

The comparability of these two sets of statistics suggested that the trainees' average self-placed level on the ILR-referenced ESL proficiency scale tended to be generally consistent with expectation for individuals with the observed average level of performance on the TOEIC—that is, the self-assessed ILR-scaled levels, on the average, tended to be consistent with regression-based guidelines for predicting ILR-scaled LPI rating from TOEIC score(s).

To provide a systematic assessment of the foregoing inference, general regression equations developed in the basic calibration study (for detail, see Wilson, 1989: Endnote 27) for predicting LPI rating were used to generate three predicted LPI rating values based, respectively, on TOEIC LC only, TOEIC LC and R treated as separate predictors, and TOEIC Total score. Residual values reflecting the difference between criterion values (ILR-scaled self rating or the similarly scaled LPI rating) and the predicted values were computed for individuals in the respective samples.

The three residual variables, identified by the predictor(s) involved in the corresponding regression equations, were as indicated below:

- \( L\text{Conly} \) = LPI or Self - LPI predicted from TOEIC LC only;
- \( LC/R \) = LPI or Self - LPI predicted from TOEIC LC and R;
- \( \text{Total} \) = LPI or Self - LPI predicted from TOEIC Total.

Using oneway analysis of variance, differences in mean residuals associated with the respective predictive equations were analyzed for the TOEIC/Self sample (N=937) versus the
TOEIC/LPI sample (N=1045). Salient results, shown in Table 8c, indicate that regardless of the TOEIC score(s) involved in estimating LPI values, mean residuals for the two samples were very small (near zero) and not statistically significant, and the standard deviation of residuals was very similar in both samples, reflecting comparable consistency of prediction.

The results of the foregoing analyses thus suggest generally positive answers to the first two questions, above. They also suggest substantial validity for the self-assessment procedures involved.

To address the question of consistency of findings across subgroups within the TOEIC/Swiss sample, analytical procedures were used to evaluate consistency of findings across subgroups classified by native language (French, German, Other), scale language (French, German, or either French, German or English versions of the rating scale) and period of data collection (1993-4 ['93 sample] and 1994-5 ['95 sample], respectively). Descriptive statistics and intercorrelations of study variables were computed for subgroups classified, as indicated above, by native-language, scale language and data-collection period. Descriptive statistics for the observed ILR-scaled self rating were compared with corresponding statistics computed for "predicted LPI rating", where prediction was based on TOEIC Total score using the previously validated regression equation described generally above.

Salient correlational outcomes are shown in Table 9a, and Table 9b provides the descriptive statistics that permit general inferences regarding consistency across subgroups in degree of agreement between average "self-placement" (according to the ILR-scale-referenced self-rating schedule) and average "predicted placement" (according to ILR-scaled LPI rating expected for individuals with given scores on the TOEIC--in this instance, TOEIC Total score).

Based on inspection, several aspects of the findings in Table 9a and Table 9b appear to be noteworthy:

Correlations between TOEIC scores and the self-rating are generally relatively strong. It can be determined that 19 of 36 coefficients shown in Table 9a are .70 or higher, 34 of 36 are .60 or higher, and only two are below .60 (.57 and .59). Correlations by year of data collection are generally similar in both level and pattern to those for the total sample (.75, .70, and .75 for TOEIC LC, TOEIC R and TOEIC Total, respectively, as reported earlier).

For each of the subgroups in Table 9b, the mean ILR-scale-referenced self rating is relatively close to the average ILR-scaled LPI rating that would be expected for individuals with average TOEIC Total scores corresponding to those reported in the table for the respective subgroups (cf. self rating mean and predicted LPI rating mean). For example, the differences between the two means (not shown in the table) for all possible comparisons ranged between -.14 and .25, and several are essentially zero; for the total sample, the mean difference was -.01, as previously reported in Table 8c.
Table 8c
Consistency of Self-Ratings with Expectancy: Results of Residual Analysis for the Total Swiss, TOEIC/Self Sample vs. the Previously Studied, TOEIC/LPI Sample

<table>
<thead>
<tr>
<th>Prediction based on</th>
<th>LC only</th>
<th>LC (residual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>TOEIC/LPI</td>
<td>1045</td>
<td>-.06</td>
</tr>
<tr>
<td>TOEIC/Self</td>
<td>937</td>
<td>-.08</td>
</tr>
<tr>
<td>Total</td>
<td>1982</td>
<td>-.07</td>
</tr>
<tr>
<td>(df 1,1980) F = .470, p = .493</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prediction based on LC and R

| LC&R (residual) |
|------------------|---------|---------|
| TOEIC/LPI        | 1045    | -.03    | .51 |
| TOEIC/Self       | 937     | -.04    | .49 |
| Total            | 1982    | -.04    | .50 |
| (df 1,1980) F = .144, p = .705 |

Prediction based on TOEIC Total

| Total (residual) |
|------------------|---------|---------|
| TOEIC/LPI        | 1045    | -.01    | .51 |
| TOEIC/Self       | 937     | -.02    | .49 |
| Total            | 1982    | -.01    | .50 |
| (df 1,1980) F = .171, p = .678 |

Note: The predictive equations were developed in a sub-sample (N=393) of the TOEIC/LPI sample (N=1045); see text for detail. Residuals for the TOEIC/LPI sample reflect differences between observed and predicted ILR-scaled LPI rating; those for the TOEIC/Self sample reflect corresponding differences when the observed ILR-scaled self-rating is the criterion (or dependent variable).
### Table 9a

Correlations for TOEIC Listening, Reading and Total Scores, Respectively, with Self Rating for ESL Trainees Grouped by Native Language and Year

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Correlation with self rating</th>
<th>TOEIC score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>LC</td>
</tr>
<tr>
<td>Native language:</td>
<td>Scale language/Year</td>
<td>r</td>
</tr>
<tr>
<td>Total '93</td>
<td>708</td>
<td>.75</td>
</tr>
<tr>
<td>Total '95</td>
<td>229</td>
<td>.74</td>
</tr>
<tr>
<td>Total</td>
<td>937</td>
<td>.75</td>
</tr>
<tr>
<td>Fr:Fr '93</td>
<td>494</td>
<td>.77</td>
</tr>
<tr>
<td>Fr:Fr '95</td>
<td>183</td>
<td>.75</td>
</tr>
<tr>
<td>Fr:Fr Tot</td>
<td>677</td>
<td>.76</td>
</tr>
<tr>
<td>Gr:Gr '93</td>
<td>140</td>
<td>.64</td>
</tr>
<tr>
<td>Gr:Gr '95</td>
<td>15</td>
<td>.74</td>
</tr>
<tr>
<td>Gr:Gr Tot</td>
<td>155</td>
<td>.66</td>
</tr>
<tr>
<td>Oth:Fr/Gr/EN '93</td>
<td>94</td>
<td>.57</td>
</tr>
<tr>
<td>Oth:Fr/Gr/EN '95</td>
<td>31</td>
<td>.70</td>
</tr>
<tr>
<td>Oth:Fr/Gr/EN Tot</td>
<td>105</td>
<td>.59</td>
</tr>
</tbody>
</table>

Note. Fr = French; Gr = German; Oth = Other; En = English.
Table 9b

Means and Standard Deviations of TOEIC Scores, Self-Ratings (ILR/LPI Scale) and Estimated LPI Rating Based on the Corresponding TOEIC Total Score, for the Subgroups Designated

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>N</th>
<th>LC Mean</th>
<th>SD</th>
<th>R Mean</th>
<th>SD</th>
<th>Total Mean</th>
<th>SD</th>
<th>Self-Predicted Rating Mean</th>
<th>SD</th>
<th>LPI Rating** Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total 93</td>
<td>708</td>
<td>312</td>
<td>116</td>
<td>275</td>
<td>106</td>
<td>586</td>
<td>216</td>
<td>1.8</td>
<td>.7</td>
<td>1.8</td>
<td>.6</td>
</tr>
<tr>
<td>Total 95</td>
<td>229</td>
<td>303</td>
<td>114</td>
<td>276</td>
<td>99</td>
<td>579</td>
<td>207</td>
<td>1.6</td>
<td>.7</td>
<td>1.8</td>
<td>.6</td>
</tr>
<tr>
<td>Total</td>
<td>937</td>
<td>310</td>
<td>115</td>
<td>275</td>
<td>105</td>
<td>586</td>
<td>214</td>
<td>1.8</td>
<td>.7</td>
<td>1.8</td>
<td>.6</td>
</tr>
<tr>
<td>French 93</td>
<td>494</td>
<td>289</td>
<td>110</td>
<td>261</td>
<td>104</td>
<td>552</td>
<td>208</td>
<td>1.7</td>
<td>.6</td>
<td>1.7</td>
<td>.6</td>
</tr>
<tr>
<td>French 95</td>
<td>183</td>
<td>300</td>
<td>114</td>
<td>278</td>
<td>97</td>
<td>579</td>
<td>206</td>
<td>1.6</td>
<td>.7</td>
<td>1.8</td>
<td>.6</td>
</tr>
<tr>
<td>French Tot</td>
<td>677</td>
<td>292</td>
<td>111</td>
<td>266</td>
<td>102</td>
<td>559</td>
<td>208</td>
<td>1.6</td>
<td>.6</td>
<td>1.7</td>
<td>.6</td>
</tr>
<tr>
<td>German 93</td>
<td>140</td>
<td>394</td>
<td>86</td>
<td>332</td>
<td>82</td>
<td>728</td>
<td>161</td>
<td>2.3</td>
<td>.6</td>
<td>2.2</td>
<td>.5</td>
</tr>
<tr>
<td>German 95</td>
<td>15</td>
<td>343</td>
<td>117</td>
<td>278</td>
<td>115</td>
<td>622</td>
<td>230</td>
<td>1.8</td>
<td>.6</td>
<td>1.9</td>
<td>.7</td>
</tr>
<tr>
<td>German Tot</td>
<td>155</td>
<td>389</td>
<td>91</td>
<td>327</td>
<td>87</td>
<td>718</td>
<td>171</td>
<td>2.2</td>
<td>.6</td>
<td>2.2</td>
<td>.5</td>
</tr>
<tr>
<td>Other 93</td>
<td>74</td>
<td>307</td>
<td>131</td>
<td>259</td>
<td>126</td>
<td>565</td>
<td>250</td>
<td>2.0</td>
<td>.7</td>
<td>1.7</td>
<td>.7</td>
</tr>
<tr>
<td>Other 95</td>
<td>31</td>
<td>298</td>
<td>111</td>
<td>264</td>
<td>106</td>
<td>561</td>
<td>206</td>
<td>1.6</td>
<td>.8</td>
<td>1.7</td>
<td>.6</td>
</tr>
<tr>
<td>Other Tot</td>
<td>105</td>
<td>304</td>
<td>125</td>
<td>260</td>
<td>120</td>
<td>564</td>
<td>237</td>
<td>1.9</td>
<td>.7</td>
<td>1.7</td>
<td>.7</td>
</tr>
</tbody>
</table>

* TOEIC LC and R scores, respectively, are reported on standard scales ranging between 5 and 495; TOEIC Total, the simple sum of the two section scores, ranges between 10 and 990.

** This is estimated LPI mean for TOEIC examinees with the corresponding TOEIC Total mean, based on the regression of LPI rating on TOEIC total in a general sample (see Wilson, 1989, for detail; see also note to Table 8b, above, and related discussion).

Analysis of Residuals

Although all the observed TOEIC/criterion coefficients are relatively large, the relative strength of relationship between TOEIC scores and the self rating appears to vary across subgroups. That is, for example, coefficients are larger in the French sample(s) than in the German sample(s), both involving use of native-language versions of the rating schedule; and coefficients in the German sample(s), in turn, tend to be larger than corresponding coefficients in the Other sample(s), in which self rating was based on a second-language version of the rating schedule. Moreover, although the mean self-rating is very close to the corresponding mean for the predicted LPI rating in the Other subgroups, the two means appear to differ somewhat more for these subgroups than for subgroups using a native-language version of the rating schedule.
To provide a more systematic assessment of differences among subgroups with respect to overall degree of fit between distributions of ILR-scaled self ratings and distributions of predicted ILR-scale LPI ratings, where prediction is based on TOEIC Total, several one-way analyses of variance were conducted involving the corresponding residual variable (self rating minus predicted LPI rating, as described earlier [see note to Table 8b, above, and related discussion]).

Analysis 1 involved the six principal subgroups (by native language and year of data collection), Analysis 2 involved subgroups by native language, and Analysis 3 examined differences for subgroups classified according to "scale language"—that is, the native-French and -German speaking trainees who used a native-language version of the self rating schedule versus trainees representing some 21 different native languages, who used a second-language version.

In each of the analyses, results indicated that the observed differences in mean residuals were statistically significant but, consistent with findings reported above in Table 9b, the mean residuals were relatively small in absolute value, indicating relatively close agreement in all instances, between the average self-assessed level of ESL speaking proficiency and that predicted for individuals with TOEIC Total scores. Results of the analyses are shown in Table 10.

There is no a priori basis for evaluating the direction of the mean differences in residuals for the respective subgroups in terms of, say, possible differences among language groups with respect to tendencies toward over- or underestimation. Mean differences for the two linguistically homogeneous subgroups are slight. And the largest deviation from average expectancy, indicating self-ratings averaging .15 points higher than predicted LPI ratings, is registered by the linguistically heterogeneous sample, which includes representatives of some 21 different native-language groups. Although this sample used a second-language version of the self rating schedule, with potential for validity-attenuating effects associated with limitations in second-language reading ability, there appears to be no basis for hypothesizing "systematic overestimation of proficiency" as a possible effect.

Diminished correlation, of course, is a plausible effect, and as noted above, correlations between TOEIC scores and the self-rating tended to be lower in the Other sample than in the samples using a native-language scale version. In this connection, it is perhaps noteworthy that in each of the three analyses, standard deviations of residuals are larger for the subgroup(s) composed of trainees who used a second-language version of the rating schedule than for their counterparts who used a native-language version. The larger dispersions are consistent with the lower levels of correlation (reported in Table 9a, above) for this subgroup. And, a test of differences in variance estimates involving residuals for subgroups in Analysis 3 (with standard deviations of .61 and .41, respectively), suggested statistical significance at beyond the .001 level of confidence—hence less consistent prediction for the sample using a second-language version of the rating schedule. While it is possible that this outcome reflects effects that plausibly may tend to be associated with level of second-language reading proficiency—as indexed by TOEIC Reading scores, for example—the present study was not designed to assess the effects of scale-language (for example, through randomized distribution of native-language vs. target-language rating schedules), or to establish levels of performance on the TOEIC that are "sufficient" to warrant the use of self-rating instruments written in English (or other target languages).
### Table 10

Results of Analyses of Residuals (ILR-Scaled Self Rating minus Predicted ILR-Scaled LPI Rating) for Designated Subgroups: Prediction Based on TOEIC Total Score

**Analysis 1 (all Subgroups)**

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>French 93*</td>
<td>494</td>
<td>-.03</td>
<td>.43</td>
</tr>
<tr>
<td>French 95*</td>
<td>183</td>
<td>-.14</td>
<td>.44</td>
</tr>
<tr>
<td>German 93*</td>
<td>140</td>
<td>.07</td>
<td>.47</td>
</tr>
<tr>
<td>German 95*</td>
<td>15</td>
<td>-.09</td>
<td>.48</td>
</tr>
<tr>
<td>Other 93</td>
<td>74</td>
<td>.25</td>
<td>.63</td>
</tr>
<tr>
<td>Other 95</td>
<td>31</td>
<td>-.07</td>
<td>.52</td>
</tr>
<tr>
<td>Total (93,95)</td>
<td>937</td>
<td>-.01</td>
<td>.47</td>
</tr>
</tbody>
</table>

F(df=5,931) 9.00(p=.001)

**Analysis 2 (by Language: French, German, Other)**

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>677</td>
<td>-.06</td>
<td>.43</td>
</tr>
<tr>
<td>German</td>
<td>155</td>
<td>.06</td>
<td>.47</td>
</tr>
<tr>
<td>Other</td>
<td>105</td>
<td>.15</td>
<td>.61</td>
</tr>
<tr>
<td>Total</td>
<td>937</td>
<td>-.01</td>
<td>.47</td>
</tr>
</tbody>
</table>

F(df 2,934) 16.10(p=.001)

**Analysis 3 (by Scale Language: Native vs. Second Language)**

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native language</td>
<td>832</td>
<td>-.04</td>
<td>.44</td>
</tr>
<tr>
<td>Second language</td>
<td>105</td>
<td>.15</td>
<td>.61</td>
</tr>
<tr>
<td>Total</td>
<td>937</td>
<td>-.01</td>
<td>.47</td>
</tr>
</tbody>
</table>

F (df=1,935) 16.43(p=.001)

---

The predicted LPI-rating values reflect the regression of ILR-scaled LPI rating on TOEIC Total (simple sum of LC and R), in a sample (N=393) composed of native speakers of Japanese, French, Spanish and Arabic, respectively. Interviews and TOEIC scores were collected under operational assessment conditions in Japan, France, Mexico, and Saudi Arabia, respectively, by professionals trained in the LPI procedure (see Wilson, 1989 for detail).
However, it can be inferred from the fact that correlations between TOEIC scores and self rating in the Other sample were relatively strong—despite being somewhat lower than those in the French and German samples—that a significant number of the trainees using the second-language version were at least "minimally" proficient in either French, German or English to permit them to read and understand relatively lengthy level descriptions written in one of these languages; also that this may tend to be true for other samples of ESL users/learners with average TOEIC scores like those of the study sample.

DISCUSSION

Notwithstanding differences in detail for subgroups, the findings summarized above in Tables 8a through 10, indicate generally strong relationships between scores on the TOEIC—a norm-referenced test of ESL listening comprehension and reading skills—and global self-assessments based on an ILR-referenced scale.

Findings that have been reviewed indicate clearly that levels and patterns of correlation between TOEIC scores and the self rating tend to strongly parallel corresponding levels and patterns of correlation between TOEIC scores and formal LPI rating. Moreover, there was close fit at the mean between (a) the observed distributions of ILR-referenced self-ratings according to an informal version of the formal scale used to rate performance in Language Proficiency Interviews, and (b) distributions of predicted Language Proficiency Interview ratings for individuals with scores on the TOEIC like those observed in the study sample.

More generally, it is believed that the findings reviewed herein tend to reinforce the validity of the working proposition stated at the outset, namely, that given descriptions of behavior that validly reflect different functional levels of proficiency in a second language as foci for organizing and expressing their intuitions, educated adult users/learners of that language are able to rate their own level of proficiency with a significant degree of validity.

The findings clearly appear to suggest the potential value of the present scale (and other similar scales) for self-assessment purposes. Because the self-assessment schedule is directly "linked" to the ILR-scale—that is, scale points were anchored by descriptions adapted from the corresponding ILR-scale descriptions—the findings also appear to extend, albeit indirectly, evidence regarding the validity of the FSI/ILR scale itself. Regarding the latter, it appears that the language-use descriptions that comprise the ILR scale for rating level of speaking proficiency, can be thought of as constituting a valid standard for use not only by professional interviewers/raters to organize their clinical impressions of proficiency levels exhibited by examinees in Language Proficiency Interviews, but also—in modified form—by educated, albeit "linguistically naive" second-language users/learners to organize intuitive impressions regarding their language-use abilities. It is possible, of course, that educated users/learners of a target language might not "need" to use "modified" versions of the formal scale-descriptions—that is, validity of self-placement at least comparable to that reported herein for modified descriptions, possibly might be attained simply by asking subjects to to read the several level-descriptions in the formal ILR scale for rating.
ESL speaking proficiency (see Appendix A) and select the one description that best characterizes their own level of functioning in the language.

It is important to keep in mind that empirical findings involving the self-assessment schedule used in the present study are currently available only for samples composed primarily of educated, ESL users/learners who are native-speakers of French or German. Questions regarding generalizability of conclusions about this particular scale need to be and can be resolved by conducting studies involving appropriate samples from other language groups, less highly educated subjects, EFL learners with limited experience in using English, and so on.

It is possible that there may be individual and/or culturally related differences with respect to general response tendencies in ESL (or other second language) self-assessment situations--e.g., to overestimate or underestimate when asked to provide self-ratings of "proficiency" (e.g., Heilenman, 1990; Oskarsson [Oscarson], 1981; Wangsotorn, 1981). The question of tendencies toward over- or underestimation as an individual or "population" characteristic appears to be an important one for further research.\(^1\)

The findings of the present study and others that have been reviewed herein reflect self-assessment outcomes under conditions that are theoretically conducive to "objectivity" in self-assessment. For example, the respondents were aware that the information they provided would not be used to certify proficiency or in making decisions about admission/hiring/citizenship/loss of standing and so on. Such decisions logically call for external measures and observations. In any event, the findings that have been reviewed do not provide evidence regarding the validity and utility of ESL self-assessments that may be collected for purposes other than those related to research and/or in similarly "unthreatening" contexts. At the same time, the findings do give added emphasis to--and suggest the potential generalizability to other research contexts of--conclusions reached by Hilton et al. (1985) regarding the potential research uses of the self-assessment instrument employed in the context of research involving teachers of French and Spanish, respectively, in the United States. Based on their findings (see Table 3, above, and related discussion) they concluded, in part, as follows:

"The fact that a valid self-administering instrument to measure oral language proficiency now exists raises many possibilities for its application, particularly in a research setting where the subjects are guaranteed anonymity. The questionnaire as it currently exists, can be easily adapted for teachers of any language... Large national samples of language teachers, for example, (if) surveyed each year could provide useful trend data. The same questionnaire could be used for collecting pretest and posttest data as one indicator of the effectiveness of a foreign language training program" (Hilton et al., 1985: p. 38).

\(^1\) For added perspective on some of the considerations noted here, see Oscarson (forthcoming). In a comprehensive review, Oscarson identifies problems, issues, and research findings regarding questions as to how, under what conditions, and with what effects learners and other users of a foreign or second language may judge their own ability in the language.
Additional Research Questions and Issues

Warranting further attention are questions regarding scale-language/native-language effects, the validity of ILR-referenced self-ratings of proficiency in reading, writing and listening, respectively, as well as that of multiple-item, "can do" self rating scales for the four macroskills.

Scale Language: Native Language versus Second Language?

Neither the present study nor any of the other studies reviewed herein was designed to assess possible effects associated with choice of scale language. In the present study, observed test/self-rating correlations were somewhat lower in subgroups composed of individuals who needed to rely on their second-language reading skills, than in subgroups using a native-language version of the self-rating scale, to complete the self-rating process. At the same time, the observed levels of concurrent test/self-rating correlations in these same samples were relatively strong--thus, by inference, the average level of second-language reading proficiency in the samples involved was at least "minimally sufficient" to enable subjects to read, comprehend, and validly evaluate the functional implications of the relatively complexly worded behavioral descriptions in the rating schedule. In any event, logical considerations suggest, as a general rule, that native-language versions of rating schedules should be used for the purpose of obtaining self-assessments from nonnative-speakers of a target language.

At the same time, such considerations also suggest that questions regarding minimum or threshold levels of target-language reading proficiency required for valid use of self-assessment schedules written in a target language need to be explored--e.g., by randomized distribution of native- and target-language versions in various samples; or along lines suggested by Alderman's (1981) exploration of the hypothesis that level of English proficiency (as measured by the TOEFL) would "moderate" correlations between scores on English-language and Spanish-language versions of a scholastic aptitude test in a sample of Puerto Rican secondary-school students--more specifically that the observed correlation between the native- and nonnative-language aptitude-test versions would tend to vary directly across samples classified according to TOEFL score level.

Related to the foregoing are questions regarding effects on validity possibly associated with reduction in the amount of supplementary detail in behavioral descriptions and/or inclusion/exclusion of detailed descriptions for the several "in-between" levels--modifications which would involve loss of detail that may or may not be deemed critical for research and other appropriate uses of self-ratings. Briefer scales would tend to reduce the overall English-language reading load for nonnative-speakers (cf. descriptions in Table 1 and Table 4, above).

Extend Research on "Can Do" Scales

Several major studies (for example, Clark, 1981; Hilton et al., 1985; Carlson, Burn, Useem, and Yachimowicz, 1990; LeBlanc and Painchaud, 1985) have found relatively high levels of concurrent correlation between "can do" scales and direct measures. "Can do" scales typically ask subjects to indicate the relative ease (difficulty) with which they can perform communicative acts represented by each of a number of descriptive statements "... intended to represent increasingly sophisticated language-use situations" (Clark, 1981: p. 34). The studies just cited included "can do"
scales for listening, reading, writing and/or speaking proficiency.

The findings reported by LeBlanc and Painchaud (1985) are particularly noteworthy in that they constitute evidence of quite useful levels of pragmatic validity for academically oriented "can do" scales similar in design to those used in Barrows et al. (1981) and Hilton et al. (1985). LeBlanc and Painchaud reported that, at the University of Ottawa, self-assessments based on "can do" scales written in English and in French, respectively, when used for for placement purposes proved to be as valid as was the set of external measures administered concurrently--when "validity" was defined as incidence of change due to "misplacement" involving the two methods. Illustrative copies of the English and French versions of the University of Ottawa self-assessment inventory, supplied by Professor LeBlanc, are shown in Appendix C.9

Studies designed to assess the validity and potential utility of "can do" scales clearly appear to be warranted. In corporate settings, for example, it would seem to be useful to explore the validity and utility of self-ratings based on "can do" statements describing characteristic workplace scenarios.

For example, studies designed to classify a variety of specific ESL proficiency-related workplace-tasks, formulated in "can do" format, according to their relative self-perceived ease or difficulty for employees, generally and by ESL proficiency levels (e.g., particular ILR-scale levels [self-as-sessed and/ or directly assessed] and/or corresponding TOEIC-score levels), would appear to have substantial potential value. In the same general vein, it would appear to be useful to have expert LPI interviewers/raters independently classify the same "can do" statements.

Generally speaking, classification by experts according to "judged LPI level" of a large number of "can do" statements representing a wide array of "increasingly sophisticated language-use situations . . ." (after Clark, 1981, p. 34), appears to represent a useful goal for further research.

*Explore ILR-Parallel Self-Assessment for the Four Basic Macroskills*

The ILR-referenced scale under consideration in this study was designed for self-assessment of speaking proficiency in English; self-assessments of other basic macroskills have not been widely studied, but results of several such studies (e.g., Clark, 1981; Oskarsson [Oscarson], 1981; Wangsotorn, 1981; Hilton et al., 1985; LeBlanc and Painchaud, 1985) suggest that attention should be given to the development and evaluation of self-assessment scales for the four basic macro-skills. Generally illustrative global scales for self-assessment of second-language listening, reading, writing, and/or speaking proficiency, respectively, are provided in Appendix D--from Oskarsson and Wangsotorn, respectively, cited above. These scales, though not specifically ILR-referenced, encompass a similar "quasi-absolute" range of proficiency--that is, levels ranging from no proficiency through native-like proficiency.
REFERENCES


APPENDIX

A. ILR Speaking Proficiency Scale

B. Self-Rating Schedule for Oral English Proficiency:
   English-, French-, and German-Language Versions Used
   by TOEIC Language Consulting, SA

C. University of Ottowa Can Do Scale for Self-Assessment
   of Proficiency in French as a Second Language

D. Illustrative Global Scales for Self- or External
   Assessment of English and/or Other Second Language
   (Oskarsson/Oscarson, 1981; Wangsotorn, 1981)
APPENDIX A. Levels of Oral English Proficiency in the FSI/ILR Scale (see, e.g., Duran, Canale, Penfield, & Stansfield, 1985)

Preface

The following proficiency level descriptions characterize spoken language use. Each of the six "base levels" (coded 00, 10, 20, 30, 40, and 50) implies control of any previous "base level's" functions and accuracy. The "plus level" designation (coded 06, 16, 26, etc.) will be assigned when proficiency substantially exceeds one base skill level and does not fully meet the criteria for the next "base level." The "plus level" descriptions are therefore supplementary to the "base level" descriptions.

A skill level is assigned to a person through an authorized language examination. Examiners assign a level on a variety of performance criteria exemplified in the descriptive statements. Therefore, the examples given here illustrate, but do not exhaustively describe, either the skills a person may possess or situations in which he/she may function effectively. Statements describing accuracy refer to typical stages in the development of competence in the most commonly taught languages in formal training programs. In other languages, emerging competence parallels these characterizations, but often with different details.

Unless otherwise specified, the term "native speaker" refers to native speakers of a standard dialect.

"Well-educated," in the context of these proficiency descriptions, does not necessarily imply formal higher education. However, in cultures where formal higher education is common, the language-use abilities of persons who have had such education is considered the standard. That is, such a person meets contemporary expectations for the formal, careful style of the language, as well as a range of less formal varieties of the language.

Speaking 0 (No Proficiency)

Unable to function in the spoken language. Oral production is limited to occasional isolated words. Has essentially no communicative ability. (Has been coded S-0 in some nonautomated applications.)

Speaking 0+ (Memorized Proficiency)

Able to satisfy immediate needs using rehearsed utterances. Shows little real autonomy of expression, flexibility, or spontaneity. Can ask questions or make statements with reasonable accuracy only with memorized utterances or formulae. Attempts at creating speech are usually unsuccessful.

Examples: The individual's vocabulary is usually limited to areas of immediate survival needs. Most utterances are telegraphic; that is, functors (linking words, markers, and the like) are omitted, confused, or distorted. An individual can usually differentiate most significant sounds when produced in isolation, but, when combined in words or groups of words, errors may be frequent. Even with repetition, communication is severely limited even with people used to dealing with foreigners. Stress, intonation, tone, etc. are usually quite faulty. (Has been coded S-0+ in some nonautomated applications.)
Speaking 1 (Elementary Proficiency)

Able to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics. A native speaker must often use slowed speech, repetition, paraphrase, or a combination of these to be understood by this individual. Similarly, the native speaker must strain and employ real-world knowledge to understand even simple statements/questions from this individual. This speaker has a functional, but limited proficiency. Misunderstandings are frequent, but the individual is able to ask for help and to verify comprehension of native speech in face-to-face interaction. The individual is unable to produce continuous discourse except with rehearsed material.

Examples: Structural accuracy is likely to be random or severely limited. Time concepts are vague. Vocabulary is inaccurate, and its range is very narrow. The individual often speaks with great difficulty. By repeating, such speakers can make themselves understood to native speakers who are in regular contact with foreigners, but there is little precision in the information conveyed. Needs, experience, or training may vary greatly from individual to individual; for example, speakers at this level may have encountered quite different vocabulary areas. However, the individual can typically satisfy predictable, simple, personal, and accommodation needs; can generally meet courtesy, introduction, and identification requirements; exchange greetings; elicit and provide, for example, predictable and skeletal biographical information. He/she might give information about business hours, explain routine procedures in a limited way, and state in a simple manner what actions will be taken. He/she is able to formulate some questions even in languages with complicated question constructions. Almost every utterance may be characterized by structural errors and errors in basic grammatical relations. Vocabulary is extremely limited and characteristically does not include modifiers. Pronunciation, stress, and intonation are generally poor, often heavily influenced by another language. Use of structure and vocabulary is highly imprecise. (Has been coded S-1 in some nonautomated applications.)

Speaking 1+ (Elementary Proficiency, Plus)

Can initiate and maintain predictable face-to-face conversations and satisfy limited social demands. He/she may, however, have little understanding of the social conventions of conversation. The interlocutor is generally required to strain and employ real-world knowledge to understand even some simple speech. The speaker at this level may hesitate and may have to change subjects due to lack of language resources. Range and control of the language are limited. Speech largely consists of a series of short, discrete utterances.

Examples: The individual is able to satisfy most travel and accommodation needs and a limited range of social demands beyond exchange of skeletal biographic information. Speaking ability may extend beyond immediate survival needs. Accuracy in basic grammatical relations is evident, although not consistent. May exhibit the more common forms of verb tenses, for example, but may make frequent errors in formation and selection. While some structures are established, errors occur in more complex patterns. The individual typically cannot sustain coherent structures in longer utterances or unfamiliar situations. Ability to describe and give precise information is limited. Person, space, and time references are often used incorrectly. Pronunciation is understandable to natives used to dealing with foreigners. Can combine most significant sounds with reasonable comprehensibility, but has difficulty in producing certain sounds in certain positions or in certain combinations. Speech will usually be labored. Frequently has to repeat utterances to be understood by the general public. (Has been coded S-1+ in some nonautomated applications.)

Speaking 2 (Limited Working Proficiency)
Able to satisfy routine social demands and limited work requirements. Can handle routine work-related interactions that are limited in scope. In more complex and sophisticated work-related tasks, language usage generally disturbs the native speaker. Can handle with confidence, but not with facility, most normal, high-frequency social conversational situations including extensive, but casual conversations about current events, as well as work, family, and autobiographical information. The individual can get the gist of most everyday conversations but has some difficulty understanding native speakers in situations that require specialized or sophisticated knowledge. The individual's utterances are minimally cohesive. Linguistic structure is usually not very elaborate and not thoroughly controlled; errors are frequent. Vocabulary use is appropriate for high-frequency utterances, but unusual or imprecise elsewhere.

Examples: While these interactions will vary widely from individual to individual, the individual can typically ask and answer predictable questions in the workplace and give straightforward instructions to subordinates. Additionally, the individual can participate in personal and accommodation-type interactions with elaboration and facility; that is, can give and understand complicated, detailed, and extensive directions and make non-routine changes in travel and accommodation arrangements. Simple structures and basic grammatical relations are typically controlled; however, there are areas of weakness. In the commonly taught languages, these may be simple markings such as plurals, articles, linking words, and negatives or more complex structures such as tense/aspect usage, case morphology, passive constructions, word order, and embedding. (Has been coded S-2 in some nonautomated applications.)

Speaking 2+ (Limited Working Proficiency, Plus)

Able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective. The individual shows considerable ability to communicate effectively on topics relating to particular interests and special fields of competence. Often shows a high degree of fluency and ease of speech, yet when under tension or pressure, the ability to use the language effectively may deteriorate. Comprehension of normal native speech is typically nearly complete. The individual may miss cultural and local references and may require a native speaker to adjust to his/her limitations in some ways. Native speakers often perceive the individual's speech to contain awkward or inaccurate phrasing of ideas, mistaken time, space, and person references, or to be in some way inappropriate, if not strictly incorrect.

Examples: Typically the individual can participate in most social, formal, and informal interactions; but limitations either in range of contexts, types of tasks, or level of accuracy hinder effectiveness. The individual may be ill at ease with the use of the language either in social interaction or in speaking at length in professional contexts. He/she is generally strong in either structural precision or vocabulary, but not in both. Weakness or unevenness in one of the foregoing, or in pronunciation, occasionally results in miscommunication. Normally controls, but cannot always easily produce, general vocabulary. Discourse is often incohesive. (Has been coded S-2+ in some nonautomated applications.)

Speaking 3 (General Professional Proficiency)

Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Nevertheless, the individual's limitations generally restrict the professional contexts of language use to matters of shared knowledge and/or international convention. Discourse is cohesive. The individual uses the language acceptably, but with some noticeable imperfections; yet, errors virtually never interfere with understanding and rarely disturb the native speaker. The individual can effectively combine structure and vocabulary to convey his/her meaning accurately. The individual speaks readily and fills pauses suitably. In face-to-face conversation with natives speaking the standard dialect at a normal rate of speech,
comprehension is quite complete. Although cultural references, proverbs, and the implications of nuances and idiom may not be fully understood, the individual can easily repair the conversation. Pronunciation may be obviously foreign. Individual sounds are accurate; but stress, intonation, and pitch control may be faulty.

Examples: Can typically discuss particular interests and special fields of competence with reasonable ease. Can use the language as part of normal professional duties such as answering objections, clarifying points, justifying decisions, understanding the essence of challenges, stating and defending policy, conducting meetings, delivering briefings, or other extended and elaborate informative monologues. Can reliably elicit information and informed opinion from native speakers. Structural inaccuracy is rarely the major cause of misunderstanding. Use of structural devices is flexible and elaborate. Without searching for words or phrases, the individual uses the language clearly and relatively naturally to elaborate concepts freely and make ideas easily understandable to native speakers. Errors occur in low-frequency and highly complex structures. (Has been coded S-3 in some nonautomated applications.)

Speaking 3+ (General Professional Proficiency, Plus)

Is often able to use the language to satisfy professional needs in a wide range of sophisticated and demanding tasks.

Examples: Despite obvious strengths, may exhibit some hesitancy, uncertainty, effort, or errors which limit range of language-use tasks that can be reliably performed. Typically there is particular strength in fluency and one or more, but not all, of the following: breadth of lexicon, including low- and medium-frequency items, especially socio-linguistic/cultural references and nuances of close synonyms; structural precision, with sophisticated features that are readily, accurately, and appropriately controlled (such as complex modification and embedding in Indo-European languages); discourse competence in a wide range of contexts and tasks, often matching a native speaker's strategic and organizational abilities and expectations. Occasional patterned errors occur in low-frequency and highly complex structures. (Has been coded S-3+ in some nonautomated applications.)

Speaking 4 (Advanced Professional Proficiency)

Able to use the language fluently and accurately on all levels normally pertinent to professional needs. The individual's language usage and ability to function are fully successful. Organizes discourse well, using appropriate rhetorical speech devices, native cultural references, and understanding. Language ability only rarely hinders him/her in performing any task requiring language; yet, the individual would seldom be perceived as a native. Speaks effortlessly and smoothly and is able to use the language with a high degree of effectiveness, reliability, and precision for all representational purposes within the range of personal and professional experience and scope of responsibilities. Can serve as an informal interpreter in a range of unpredictable circumstances. Can perform extensive, sophisticated language tasks, encompassing most matters of interest to well-educated native speakers, including tasks which do not bear directly on a professional specialty.

Examples: Can discuss in detail concepts which are fundamentally different from those of the target culture and make those concepts clear and accessible to the native speaker. Similarly, the individual can understand the details and ramifications of concepts that are culturally or conceptually different from his/her own. Can set the tone of interpersonal official, semi-official, and non-professional verbal exchanges with a representative range of native speakers (in a range of varied audiences, purposes, tasks, and settings). Can play an effective role among native speakers in such contexts as conferences, lectures, and debates on matters of disagreement. Can advocate a position at length, both formally and in chance
encounters, using sophisticated verbal strategies. Understands and reliably produces shifts of both subject matter and tone. Can understand native speakers of the standard and other major dialects in essentially any face-to-face interaction. (Has been coded S-4 in some nonautomated applications.)

Speaking 4+ (Advanced Professional Proficiency, Plus)

Speaking proficiency is regularly superior in all respects, usually equivalent to that of a well-educated, highly articulate native speaker. Language ability does not impede the performance of any language-use task. However, the individual would not necessarily be perceived as culturally native.

Examples: The individual organizes discourse well, employing functional rhetorical speech devices, native cultural references, and understanding. Effectively applies a native speaker's social and circumstantial knowledge. However, cannot sustain that performance under all circumstances. While the individual has a wide range and control of structure, an occasional non-native slip may occur. The individual has a sophisticated control of vocabulary and phrasing that is rarely imprecise, yet there are occasional weaknesses in idioms, colloquialisms, pronunciation, cultural reference, or there may be an occasional failure to interact in a totally native manner. (Has been coded S-4+ in some nonautomated applications.)

Speaking 5 (Functionally Native Proficiency)

Speaking proficiency is functionally equivalent to that of a highly articulate well-educated native speaker and reflects the cultural standards of the country where the language is natively spoken. The individual uses the language with complete flexibility and intuition, so that speech on all levels is fully accepted by well-educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references. Pronunciation is typically consistent with that of well-educated native speakers of a non-stigmatized dialect. (Has been coded S-5 in some nonautomated applications.)
APPENDIX B

SELF-RATING SCHEDULE FOR ORAL ENGLISH PROFICIENCY: ENGLISH-, FRENCH-, AND GERMAN-LANGUAGE VERSIONS USED BY TOEIC LANGUAGE CONSULTING SA

The ILR-referenced rating schedules included in this appendix were used to generate the self-ratings that constituted the basis for the present study. Level descriptions are provided for the first seven levels only, intended to correspond nominally to FSI/ILR Levels 0+ through 3+ for use with populations of ESL users/learners for which this is a sufficient range--relatively few are expected to function at higher levels on the FSI/ILR speaking proficiency scale. In the initial version (in English only), Levels 4, 4+ and 5 were included (see Wilson, 1989, for the full scale).
Read the following paragraphs which describe levels of English from emerging, to advanced professional proficiency. Then choose the description that corresponds most accurately to your current level of ability. Mark the number of the description chosen in space following the last paragraph. Be as objective as possible. Your choice has no influence on your test result.

1. **MY ENGLISH IS LIMITED TO MEMORIZED WORDS AND PHRASES. I ASK QUESTIONS ONLY WHEN I THINK I KNOW THE ANSWER OR WHEN THE ANSWER IS YES OR NO. I AM NOT ABLE TO USE GRAMMAR.**

I am able to satisfy immediate survival needs using memorized material. There are long pauses in my speech and I must rely on single words and phrases for all communication, except for occasional uses of one or two verbs. I can ask questions and make statements with reasonable accuracy only by using short memorized material. My vocabulary is limited to survival needs. I can understand sounds when they are isolated, but when they are in words or groups of words, I have a very difficult time understanding them. When I talk with people in English, even if they are used to speaking with learners, I have a very difficult time making myself understood.

2. **I AM ABLE TO USE LIMITED ENGLISH GRAMMAR. I MUST TRANSLATE EVERYTHING I SAY BUT I CAN MAKE SENTENCES. I KNOW ENOUGH ENGLISH TO SURVIVE IN THE LANGUAGE, IF PUT IN ENGLISH-SPEAKING SITUATION.**

I am able to survive in English, to order a meal and get a room in a hotel. I can ask and answer simple questions, ask for directions, respond to simple statements, and maintain very simple face-to-face conversation. I am able to ask questions using limited grammar, but usually with much inaccuracy. Almost all of my statements contain fractured syntax and other grammatical errors. If I repeat what I say, I can make myself understood to people who have regular contact with learners of English. If I think about it, I am able to create sentences, using simple verbs, nouns, pronouns, etc.
3. I AM ABLE TO SPEAK ENGLISH WITHOUT TRANSLATING EVERY SENTENCE. I KNOW SOME PAST TENSE AND FUTURE TENSE VERBS BUT I AM UNCERTAIN IN THEIR USE. I FEEL COMFORTABLE IN INFORMAL SOCIAL SITUATIONS, IF NOT ASKED TO SPEAK VERY MUCH.

I am able to speak English as described above, as well as satisfy limited social demands. I am able to produce language more spontaneously (than in paragraph 2), but it is still difficult for me to speak. I am able to use some verbs in the past and future tenses. Because my vocabulary is limited, I often have to think about words. I am unable to control grammar in longer sentences and/or unfamiliar situations. I have only limited ability to describe and cannot give precise information. In conversation, I speak in short sentences, thinking about almost every one before speaking. I have difficulty producing certain sounds, but generally my speech is understood. While I am not comfortable speaking English for any length of time, I think that with some practice I could become a confident speaker.

4. I CAN DISCUSS MY WORK, HOME LIFE, CURRENT EVENTS, HOBBIES, LIKES AND DISLIKES IN ENGLISH. I CAN UTTER SENTENCES IN SERIES, WITHOUT HAVING TO PAUSE, AND CAN NARRATE AND DESCRIBE IN PAST, PRESENT AND FUTURE. MY CONTROL OF BASIC GRAMMAR, INCLUDING PAST TENSE VERBS, IS GOOD.

I am able to speak English as described above, and I am also able to discuss my work in English. I can discuss matters that are of a concrete nature with some confidence. I am comfortable in social situations, casual conversations, discussing current events, work, family, and my likes and dislikes. I can understand most conversations on non-technical subjects, if they require no specialized information. I can give detailed directions on how to get from one place to another. I am able to narrate and describe in the past, present, future, and I can string sentences together in conversation with no difficulty. I do not have a thorough or confident control of all grammar, but I am able to use simple language quite accurately. While my vocabulary is not very precise, it is broad enough to discuss a wide variety of topics, most of which are familiar to me. In simple conversations, I rarely have to mentally translate what I am going to say.

5. I "KNOW" ALMOST ALL THE ENGLISH GRAMMAR, BUT I AM NOT YET ABLE TO USE IT. I USE SIMPLE GRAMMAR WITH NO PROBLEM BUT EXPERIENCE DIFFICULTY WITH MORE ADVANCED GRAMMAR. MY COMPREHENSION IS ALMOST 100%. WHEN EXPRESSING AN OPINION, I HAVE TO STOP AND THINK ABOUT HOW I AM GOING TO EXPRESS IT.

I am able to speak English as described above, and my vocabulary is adequate to discuss a wide range of topics, many of which are unfamiliar to me. I have quite good control of verbs in the past and future, and am able to use other less frequent forms, including conditionals. I still make mistakes with regard to grammar and vocabulary that I consider simple, but those errors are more the result of carelessness than anything else. I have been exposed to all of the important grammar points, but I cannot always recall them as I speak. I find it difficult to discuss concepts, thoughts, opinions, or hypothetical situations in English, as I am required to both think and speak at the same time. When I offer an opinion, etc., I have to stop and think about how I will express
myself. My accent is quite good and I am very fluent, except when I must stop to think about what I am going to say.

6. I AM ABLE TO SAY NEARLY EVERYTHING I WANT TO SAY IN ENGLISH, ALTHOUGH MY VOCABULARY IS NOT ALWAYS PRECISE. I CAN EXPRESS OPINIONS, DISCUSS CONCEPTS, AND HYPOTHESIZE. I AM ABLE TO USE ALL GRAMMATICAL CONSTRUCTIONS, BUT MAKE SOME MISTAKES. I am able to participate effectively in most formal and informal conversations on practical, social, and professional topics. I can discuss particular interests and special fields of competence with reasonable ease. I understand nearly everything I hear in normal, everyday speech. My vocabulary is broad enough that if I do not know a word, it does not prevent me from expressing myself. One way or another, I am able to express myself on every topic on which I wish to express myself and I am never driven to silence because of lack of control of grammar or vocabulary limitations. Grammatical errors in my speech are still fairly common, but they never prevent me from being understood. I am able to discuss any topic that is presented, on which I would be able to comment in my native language.

7. I CAN USE ALL VERB FORMS CONSISTENTLY AND ACCURATELY, BUT SOMETIMES MAKE MISTAKES. I HAVE MINOR PROBLEMS WITH CERTAIN GRAMMAR POINTS, SUCH AS ARTICLES AND PREPOSITIONS. I CAN FUNCTION EFFECTIVELY IN UNFAMILIAR SITUATIONS. I am able to speak as described above, but the error rate in my speech is quite low. My control of grammar is such that I can use all verb forms consistently and accurately, although there are still certain patterns of error in my speech. I am able to speak English with a structural accuracy and a breadth of vocabulary sufficient to extensively discuss my professional needs. I understand and use a great many idiomatic expressions, but in that regard my speech is still foreign. I sometimes have difficulty with certain verb forms, articles, and prepositions. I would not be taken for a native speaker, but I am able to respond appropriately in unfamiliar or surprise situations.

WRITE THE NUMBER OF THE PARAGRAPH THAT, IN YOUR OPINION, BEST CORRESPONDS TO HOW YOU SPEAK ENGLISH:
German language version

NAME:_________________________________ VORNAME :_____________________________________
ORT :  _________________________________ DATUM :  ______________________________________

Die folgenden Abschnitte beschreiben Sprachfähigkeiten in Englisch, die eine Einstufung vom Anfänger bis zum Fortgeschrittenen erlauben. Lesen Sie die einzelnen Abschnitte durch, bis Sie die Beschreibung finden, die am besten Ihren Kenntnissen entspricht. Die **FETTGEDRUCKTEN SÄTZE** enthalten jeweils eine Zusammenfassung.

1. **MEIN ENGLISCH BESCHRÄNKT SICH AUF AUSWENDIG GELERNTE WÖRTER UND AUSDRÜCKE. FRAGEN STELLE ICH NUR, WENN ICH DENKE, DASS ICH DIE ANTWORT KENNE, ODER WENN SIE JA ODER NEIN IST. ICH BIN NICHT FÄHIG, GRAMMATIK ANZUWENDEN.**


2. **ICH BIN IN DER LAGE, DIE EINFACHEN GRUNDLAGEN DER ENGLISCHEN GRAMMATIK ZU BENUTZEN. ICH MUSS ALLES, WAS ICH SAGEN WILL, ÜBERSETZEN, KANN ABER SÄTZE BILDEN. ICH BEHERRSCHE GENUG ENGLISCH, UM MICH IN EINER SITUATION, IN DER ENGLISCH GESPROCHEN WIRD, IN EINFACHER SPRACHE ZU VERSTÄNDIGEN.**


5. Ich "kenne" fast die ganze Englische Grammatik, aber ich bin noch nicht in der Lage, sie immer anzuwenden. Ich benutze einfache Grammatik ohne Probleme, aber es tauchen Schwierigkeiten mit komplexen Grammatikalischen Strukturen auf. Ich verstehe beinahe 100%. Wenn ich über theoretisches sprechen oder meine Meinung ausdrücken will, muss ich zuerst innehalten und überlegen, wie ich mich ausdrücken soll.

6.
ICH BIN IN DER LAGE, FAST ALLES IN ENGLISCH ZU SAGEN, WAS ICH SAGEN WILL, OBWOHL MEIN WORTSCHATZ NICHT IMMER PRÄZIS IST. ICH KANN MEINUNGEN AUSDRÜCKEN, IDEEN DISKUTIEREN UND HYPOTHESEN AUFSTELLEN. ICH BIN IN DER LAGE, ALLE GRAMMATIKALISCHEN KONSTRUKTIONEN ANZUWENDEN, MACHE ABER EINIGE FEHLER.

7.

NOTIEREN SIE BITTE HIER DIE NUMMER DES ABSCHNITTS, DER NACH IHRER MEINUNG AM BESTEN IHREN ENGLISCHFÄHIGKEITEN ENTSPRicht:__________
Lisez les paragraphes suivants et choisissez celui qui, selon vous, décrit le mieux votre propre capacité d'expression. La partie en MAJUSCULES résume le texte qui suit pour que vous puissiez vous situer rapidement. Ensuite, écrivez le numéro du paragraphe choisi dans l’espace prévu au verso.

1. **MON ANGLAIS SE LIMITE A DES MOTS ET DES PHRASES APPRIS PAR COEUR. JE POSE DES QUESTIONS UNIQUEMENT LORSQUE JE PENSE CONNAITRE LA REPONSE, OU BIEN QUAND LA REPONSE EST OUI OU NON. JE NE PEUX PAS UTILISER LA GRAMMAIRE.**

   Je fonctionne assez pour satisfaire à mes besoins élémentaires en utilisant des éléments mémorisés. Il y a de longues pauses quand je parle et je dois recourir aux mots et phrases assez isolés les uns des autres pour toute communication à part deux ou trois verbes. Je peux formuler des questions et des phrases avec précision uniquement avec des éléments que j'ai mémorisés. Mon vocabulaire se limite à mes besoins élémentaires. Je comprends des sons isolés, mais j'ai beaucoup de peine à les distinguer s'ils font partie intégrante de mots et de phrases. Lors d'un échange en anglais, j'ai beaucoup de peine à me faire comprendre, même si mon interlocuteur a l'habitude de parler avec des non-anglophones.

2. **JE PEUX UTILISER UN PEU DE GRAMMAIRE ANGLAISE. IL FAUT QUE JE TRADUISE TOUT CE QUE JE VEUX DIRE, MAIS JE PEUX FORMULER DES PHRASES. JE SAIS ASSEZ D'ANGLAIS POUR ME DEBROUILLER DANS UN CONTEXTE ANGLOPHONE.**

   Je peux me débrouiller en anglais, par exemple pour commander un repas et prendre une chambre d'hôtel. Je peux poser des questions simples et y répondre, demander le chemin et m'entretenir dans des dialogues très simples. Je peux poser des questions en utilisant la grammaire de base, mais souvent avec beaucoup d'imprécisions. Presque toutes mes phrases contiennent des erreurs de syntaxe et d'autres fautes de grammaire. Si je me répète, je peux me faire comprendre par des personnes qui sont en contact régulier avec des non-anglophones. Si j'y réfléchis, je peux formuler des phrases en utilisant des verbes simples, des substantifs, des pronoms, etc.

3. **JE PEUX PARLER L'ANGLAIS SANS DEVOIR TRADUIRE CHAQUE Phrase. JE CONNAIS UN CERTAIN NOMBRE DE VERBES AU PASSE ET AU FUTUR, AIS JE NE LES EMPLOIE PAS TOUJOURS CORRECTEMENT. JE ME SENS A L'AISE DANS DES SITUATIONS COURANTES, SI L'ON NE ME DEMANDE PAS TROP SOUVENT**
DE PRENDRE LA PAROLE.
Je peux m'exprimer comme décrit ci-dessus, de même que satisfaire à un nombre limité d'exigences courantes. Je peux parler plus spontanément (par rapport à la description 2) mais il m'est encore difficile de parler en anglais. Je peux employer certains verbes au passé et au futur. Parce que mon vocabulaire est limité, je dois souvent chercher mes mots. Je ne maitrise pas la grammaire dans des phrases longues ou compliquées ni dans des situations qui ne me sont pas familières. Ma capacité de donner des descriptions et de transmettre des informations précises est limitée. En conversation, j'utilise des phrases courtes et, normalement, les prépare avant de les exprimer. J'éprouve des difficultés à produire certains sons, mais, en général, on me comprend. Bien que je ne sois pas à l'aise si je dois prendre la parole pendant un long moment, je pense que si je m'exerçais je pourrais m'exprimer avec assurance.

4.
JE SAIS DISCUTER EN ANGLAIS DE MON TRAVAIL, MA VIE DE FAMILLE, DES ACTUALITES, MES HOBBIES ET MES PREFERENCES. JE PEUX ENCHAINER DES PHRASES SANS HESITATION POUR TRADUIRE (DE MA LANGUE MATERNELLE), DE MEME QUE RAPPORTER ET DONNER DES DESCRIPTIONS AU PASSE, PRESENT ET FUTUR. J'AI UNE BONNE MAITRISE DE LA GRAMMAIRE DE BASE, Y COMPRIS DES VERBES AU PASSE.

Je peux m'exprimer comme décrit ci-dessus et je peux aussi discuter de mon travail en anglais. Je peux traiter de sujets concrets avec une certaine assurance. Je me sens à l'aise dans des situations sociales, des conversations de tous les jours, ainsi que dans des discussions relatives aux actualités, travail, famille, et mes préférences. Je comprends la plupart des conversations sur des sujets non spécialisés, si une information technique n'est pas requise. Je peux indiquer le chemin avec beaucoup de détails. Je peux rapporter et donner des descriptions au passé, présent et futur, ainsi qu'enchaîner, sans difficultés, des phrases lors de conversations. Je ne maitrise pas la totalité de la grammaire, mais mon anglais de base est très précis. Tandis que mon vocabulaire n'est pas très précis, il est assez étendu pour discuter d'une grande variété de sujets, dont la plupart me sont familiers. Lors de conversations simples, il est rare que je traduise mes pensées avant de m'exprimer.

5.
JE CONNAIS PRESQUE LA TOTALITE DE LA GRAMMAIRE ANGLAISE, MAIS JE NE L'EXPLOITE PAS ENCORE EN PRATIQUE. LA GRAMMAIRE DE BASE NE ME POSE AUCUN PROBLEME, MAIS J'EPROUVE UNE CERTAINE DIFFICULTE AVEC LA GRAMMAIRE AVANCEE. JE COMPRENDS PRESQUE TOUT. LORSQUE JE VEUX EXPRIMER MON OPINION, JE DOIS D'ABORD REFLECHIR A LA MANIERE DE LA FORMULER EN ANGLAIS.

Je m'exprime en anglais comme indiqué ci-dessus et mon vocabulaire est suffisant pour discuter d'une grande variété de sujets dont beaucoup ne me sont pas familiers. J'ai une bonne maîtrise des verbes au passé et au futur et je sais employer d'autres temps moins habituels, tels que les conditionnels. Je commets encore, plutôt par manque d'attention, des erreurs mineures de grammaire et de vocabulaire. J'ai étudié tous les points principaux de la grammaire anglaise, mais je ne les applique pas systématiquement en m'exprimant. J'éprouve des difficultés à discuter en anglais de concepts, d'idées, d'opinions ou de situations hypothétiques, puisqu'il faut à la fois réfléchir et m'exprimer. Quand je donne une opinion, je dois d'abord réfléchir à sa formulation. Mon accent est très bon et je parle très couramment, sauf lorsque j'hésite pour réfléchir à l'idée que je veux exprimer.
6. JE PEUX EXPRIMER EN ANGLAIS PRATIQUEMENT TOUT CE QUE JE VEUX, BIEN QUE MON VOCABULAIRE NE SOIT PAS TOUJOURS PRECIS. JE PEUX EXPRIMER DES OPINIONS, DISCUTER DE CONCEPTS ET EMETTRE DES HYPOTHESES. JE SUIS MEME D'UTILISER TOUTE LA GRAMMAIRE, MAIS JE COMMETS QUELQUES ERREURS. 

Je suis capable de participer avec efficacité à des conversations courantes et formelles sur des thèmes sociaux, pratiques et professionnels. Je peux m'exprimer avec une relative facilité sur des problèmes particuliers et des domaines spécialisés. Je comprends quasiment tout ce que je comprends lors de conversations courantes. Mon vocabulaire est suffisamment étendu pour me permettre de m'exprimer même si un mot précis me manque. Je suis capable de m'exprimer sur tous les sujets sur lesquels je souhaite m'exprimer, sans devoir hésiter à cause d'insuffisances grammaticales ou manque de vocabulaire. Mon anglais contient encore un certain nombre d'erreurs, mais qui n'empêchent pourtant pas la compréhension. Je peux discuter de n'importe quels thèmes sur lesquels je pourrais aussi m'exprimer en français.

7. JE PEUX UTILISER TOUS LES TEMPS AVEC REGULARITE ET PRECISION, MAIS JE COMMETS PARFOIS UNE ERREUR. J'AI DES PROBLEMES MINEURS, TELS QUE LES ARTICLES ET LES PREPOSITIONS. JE PEUX M'EXPRIMER AVEC EFFICACITE DANS DES SITUATIONS IMPREVUES.

Je peux parler comme indiqué ci-dessus, et le nombre d'erreurs dans mon anglais est très limité. Ma maîtrise de la grammaire me permet d'utiliser tous les temps avec régularité et précision, bien qu'il existe quelques erreurs qui se répètent. Mon anglais est suffisamment précis et mon vocabulaire suffisamment étendu pour discuter à fond de mes exigences professionnelles. Je comprends et j'emploie un grand nombre d'expressions idiomatiques, mais à cet égard mon language reste perceptiblement étranger. Parfois, j'éprouve de la difficulté avec certaines structures, tels que les articles et les prépositions. On ne m'prend pas pour un(e) anglophone, mais je m'exprime de façon appropriée dans toutes situations.

**INDIQUEZ ICI LE NUMERO DU PARAGRAPHE QUI, SELON VOUS, REFLETE LE MIEUX VOTRE CAPACITE DE VOUS EXPRIMER EN ANGLAIS: __________**

---

**TOEIC Switzerland: Rte de Moncor 14, 1752 Villars-sur-Glâne**

Tel: (026) 401 2626     Fax: (026) 401 2627

✈ e-mail: test@toeic.ch  ✈ [http://www.toeic.ch](http://www.toeic.ch)
APPENDIX C: UNIVERSITY OF OTTAWA SELF-ASSESSMENT INVENTORY
FOR FRENCH AS A SECOND LANGUAGE

This appendix includes basic instructions and corresponding "can do" statements included in the self-assessment scales developed by the University of Ottowa for use in placing students according to level of listening and reading skills, respectively, in French. A French-language version of this scale is used for placement in English.

SELF-ASSESSMENT QUESTIONNAIRE (French as a Second Language)

Second Language Institute
UNIVERSITY OF OTTAWA

If you are a new student, you must complete the following questionnaire before registering either in person or by mail. It is the first step in the university's evaluation of your second language proficiency.

COMPLETING THE QUESTIONNAIRE

Read each statement carefully, think about it for a few seconds and indicate on your ANSWER SHEET how well you think you can do what is described in each statement. If you are not familiar with the situation, try to imagine how well you would perform and give yourself the appropriate score. The meaning of the scores for each item is as follows:

a: I cannot do this at all
b : I can seldom do this
c: I can do this about half the time
d: I can often do this
e: I can do this all the time

Remember to read each statement carefully, as the level of difficulty varies from one to another. Keep in mind also that if you need (or want) to take a course in French as a second language, this score will be used to place you provisionally in a group at your level. It is obviously important for you to be as careful and as honest as you can in answering each question.

N.B. The Second Language Institute reserves the right to change your placement subsequent to further evaluation in class by your teacher.

PLEASE RECORD YOUR ANSWERS ON YOUR ANSWER SHEET

SCORE CODE: a- never; b- seldom; c- half; d- often; e- always

PART 1: LISTENING

1. When people speak foreign languages around me, I can tell those speaking French
2. When I hear two persons speaking French, I can catch a few words here and there
3. In French, I can distinguish between a question and an order
4. I understand when someone gives me the time in French
5. I understand when someone tells me in French the room number where classes will be held
6. I understand questions in French about my name, address, phone number, etc. a b c d e
7. If someone gives me an appointment in French, I understand details such as the place and the time of the appointment
8. When someone is introduced to me in French, I understand his name and his field of work
9. I understand when someone explains to me in French how to get to a university building
10. Over the phone, when the caller only speaks French, I can understand what the person wants
11. If a course announcement is made in French over the campus radio, I understand the title, the place and the time of the lecture

1 A French translation of this self-assessment questionnaire is used for self-assessment of proficiency in English as a Second Language
SELF-ASSESSMENT QUESTIONNAIRE  (French as a Second Language)

Listening, continued
12. In the usual exchange of greetings between two people, I understand everything spoken to me in French
13. In a set of directions given in French at normal speed, I can catch enough details to be able to find my way (including numbers, etc.)
14. If someone comments in French on the food quality at the cafeteria, I understand the general meaning of the remarks and some details
15. If I phone a friend and someone else answers me in French, I understand what I am being told about my friend's absence
16. I can understand the directives given in French in class without having to hear them again
17. If someone addresses me in French, I can understand the gist of what I am being told
18. In radio advertising in French, I understand the main details (who is selling what, where, when, and at what price)
19. I can follow a short conversation in French between two native speakers
20. If someone describes the campus to me in French, I will understand where the main buildings and services are located
21. I can follow a short conversation in French between two native speakers
22. If I am asked in French for my opinion in an opinion poll, I understand enough to be able to answer all the questions
23. If I attend a lecture given in French in my field of specialization, I recognize most of the technical words
24. From a detailed explanation given in French in class, I can get the main idea and one or two other supporting ones
25. From the content of a discussion in French recorded on tape, I can determine the psychological state of the participants (anger, happiness, sadness, etc.)
26. When I watch a movie in French with a lot of dialogue, I can easily recognize the role of each character
27. If I hear a conversation in French where two professors are being compared, I understand enough to be able to choose the one that corresponds to my tastes
28. During discussions in student meeting, I understand enough to form an opinion about some of the ideas expressed
29. If I need help with an assignment, and the only help available is in French, I understand enough for it to be useful
30. If a professor who is replacing my usual professor states that he will give his course in French, I remain in class since I will understand almost as well as if it were in English
31. When a course is given in French in my field, I understand enough to be able to do the work related to that course
32. I understand a speaker giving a lecture in French on a familiar topic even when his sentences are very long
33. If I happen upon a round-table in French concerning fee increases for foreign students, I understand each contribution despite the diversity of accents
34. I am sufficiently fluent in French to understand most of the jokes made in French
35. At a party, I understand most of the jokes made in French
36. I understand discussions in French just as well as those in my mother tongue.

PART 2: READING

37. I understand what is written on a poster in French, when it is also illustrated by some form of picture
38. When I see something being offered for sale in French, I can give a description of the product, when and where it is being sold and how much it costs
39. I can understand unilingual traffic signs in French
40. In a short text describing some action in French, I can determine who is doing what and to whom
41. If a menu is written only in French, I can order knowing what I will be eating
42. When I get a postcard in French, I understand the usual formulae people write in postcards
43. If a student offers me a flyer in French describing a club to which he belongs, I understand well enough to decide if I would like to join
44. I can understand a one-page text in my field of specialization when there are titles and subtitles to guide me
45. When I read a short report on an accident written in French, I can tell how many persons were injured and what their condition is
46. If I have to fill out a detailed bursary form in French, I understand most of the information requested
47. When I read a detailed course description, I can understand how the course will be organized
SELF-ASSESSMENT QUESTIONNAIRE (French as a Second Language)

Part 2: Reading, continued

48. In a 2-3 page text, I can find the details that I need to answer specific questions
49. When I read an article in French on budget cuts at the university, I understand what they are and what their effect will be
50. From an interview written in French in the student newspaper, I can understand the thrust of the answers and draw some personal conclusions
51. When I look at a well detailed table of contents of a book written in French, I can tell whether the book will be useful to me or not
52. When I get unilingual advertising mail in French, I can tell what it is at first sight
53. I can understand a telegram written in French even though many words are omitted in telegraphic style
54. If a unilingual administrative memo in French on exam regulations is distributed in my class, I will understand most of the regulations
55. If I borrow a friend's course notes written in French, I am able to use them efficiently to answer that day's home-work assignment
56. I understand written French well enough so that if exam questions for a course were written in that language only, I would have no problem
57. I can read an adventure story in French (about 5 pages long) and rarely have to use a dictionary
58. I can read an editorial in French and determine the areas of agreement and disagreement between the author's views and mine
59. Faced with a new proposal for changing academic regulations written in French and distributed on the spot during a meeting I understand enough at first reading to know what it is all about
60. I know written French well enough to be able to spot mistakes and misprints in a text.

UNIVERSITY OF OTTAWA
SECOND LANGUAGE INSTITUTE
600 KING EDWARD, OTTAWA, ONTARIO
CANADA K1N 6N5 (613) 564-2973
APPENDIX D


The scales in this appendix are from what appear to be rare, published reports involving application of the global rating concept to self-ratings of the four basic language macroskills. Neither of the studies involved raised or addressed questions regarding the "functional, developmental equivalence" of similarly positioned scale levels for rating listening, speaking, reading and writing, respectively.

The scales reported by Oskarsson [Oscarson] (1981) have 11 levels that appear to nominally parallel FSI/ILR levels 0, 0+, 1, 1+ . . . , 5, but FSI scales were not specifically cited (cf. speaking scale in this appendix with Table 1 in text). Positive relationships between teachers' ratings and self-ratings were reported in analyses involving all four ratings considered collectively rather than in separate analyses. The native languages of the subjects were German and (by inference) Swedish, and corresponding native-language versions of the English-versions of the scales shown in this appendix appear to have been used. However, questions regarding the implications of agreement versus disagreement between scale-language and native-language were not specifically raised.

The scales used by Wangsotorn (1981) reportedly were adapted from scales developed by Wilkins (not otherwise identified). They were used with undergraduate- and graduate-level students in several Thai universities, including Chulalongkorn University and Kasetsart University. Students were enrolled in EFL courses. It is possible that native-language (Thai) versions were used in the study, but this is not indicated. In any event, only the English-language versions of the several scales, copies of which are shown in this appendix, were reported.
Self-Assessment Scales Developed by Oskarsson (Oscarson), 1981

Reading

_____ 5. I read and understand the language as well as a well-educated native

_____ 4.5

_____ 4. I understand everything or nearly everything written in the language within non-specialized fields. There may be words I do not understand in difficult texts.

_____ 3.5

_____ 3. I understand most of what I read in simple texts dealing with familiar subjects such as leisure interests, current affairs and living conditions. I understand most of a normal private letter dealing with everyday things such as family and their doings, etc. I understand the main contents of a normal newspaper article about a plane crash or the opening of a new underground line, for example, but not all of the details.

_____ 2.5

_____ 2. I understand the meaning of simple written instructions about the way, time, place and similar things, and also understand the essential things in simple texts dealing with familiar subjects such as common leisure interests, current affairs, and living conditions.

_____ 1.5

_____ 1. I understand the main points of a simple text and simple written directions for familiar things.

_____ 0.5

_____ 0. I cannot read the language at all.

Writing

_____ 5. I write the language as well as a well-educated native

_____ 4.5

_____ 4. I write the language rather easily and for the most part correctly. I only make occasional grammatical mistakes and spelling mistakes. When writing about subjects in which I have no opportunity to practise the language I need to use a dictionary.
Self–Assessment Scales Developed by Oskarsson (Oscarson), 1981

Writing, continued

____3.5

____3. I can formulate written messages or give a coherent account of things connected with my own life, my interests, needs and wishes, but I make mistakes in both grammar and spelling. I often cannot find the words for what I want to express. I can write down from dictation a normal prose text about a familiar subject. I may make spelling mistakes and mistakes due to a lack of words.

____2.5

____2. I can formulate simple written messages connected with my own life and my needs, but there are often grammatical and spelling mistakes and a wrong choice of words. I can write down from dictation a simple text about everyday subjects. There are often spelling mistakes and mistakes due to a lack of words.

____1.5

____1. I can formulate very simple messages connected with my own life. I only have a command of very simple words and phrases.

____0.5

____0. I cannot write the language at all.

Listening

____5. I understand the language as well as a well-educated native.

____4.5

____4. I understand most of what is said in the language, even when said by native speakers, but have difficulty in understanding extreme dialect and slang. It is also difficult for me to understand speech in unfavourable conditions, for example, through bad loudspeakers outdoors.

____3.5

____3. I can follow and understand the essential points concerning everyday and general things when spoken slowly and clearly, but do not understand native speakers if they speak very quickly or use some slang or dialect.
Self-Assessment Scales Developed by Oskarsson (Oscarson), 1981

Listening, continued

____2.5

____2. I can follow and understand the essential points concerning everyday and general things when spoken slowly and clearly, but in the course of conversation I often have to ask for things to be repeated or made clearer. I only understand occasional words of statements made in unfavourable conditions, for example, through loudspeakers outdoors.

____1.5

____1. I understand the meaning of simple requests, statements and questions if they are spoken slowly and clearly and if I have a chance of asking for them to be repeated. I only understand common words and phrases.

____0.5

____0. I do not understand the language at all.

Speaking

____5. I speak the language as well as a well-educated native.

____4.5

____4. I speak the language fluently and for the most part correctly. I have a large vocabulary so I seldom have to hesitate or search for words. On the other hand I am not completely fluent in situation in which I have had no practice with the language.

____3.5

____3. I can make myself understood in simple everyday situations, but my language is not without mistakes and sometimes I cannot find the words for what I want to say. It is difficult for me to express myself in situations in which I have had no opportunity to practise the language. I can give a short summary of general information that I have received in my native language.

____2.5

____2. I can make myself understood in simple everyday situations, for example asking and given simple directions, asking and telling the time, asking and talking about simpler aspects of work and interests. My vocabulary is rather limited, so it is only by a great deal of effort that I can use the language in new and unexpected situations.
Self-Assessment Scales Developed by Oskarsson (Oscarson), 1981 (concluded)

Speaking, continued

_____ 1.5

_____ 1. I can just about express very simple things concerning my own situation and my nearest surroundings, for example asking and answering very simple questions about the time, food, housing and directions. I only have a command of very simple words and phrases.

_____ 0.5

_____ 0. I do not speak the language at all.
Scales Developed by Wangsotorn (1981)

Reading

___ Level 1 Able to read English alphabet and understand simple public announcements

___ Level 2 Know the topic of short statements and roughly understand some complicated public announcements

___ Level 3 Able to grasp main ideas of passages read without a dictionary. Details will be comprehended with the help of a dictionary. Read very slowly.

___ Level 4 Have adequate comprehension of passages of various types. A lost of re-reading is done for full comprehension. Still unable to tell about changes in the style of writing. Recognize only changes in types of vocabulary used. Still need a dictionary. Read rather slowly.

___ Level 5 Able to comprehend all passages written in various styles. Very little re-reading. Need a dictionary when reading passages not in (familiar) areas of study. Able to read with ease passages of general interest. Still reading is slower than that of the native speaker.

___ Level 6 Able to fully comprehend passages written in various styles and having different topics. Differ from the native speaker only in reading speed and ease in reading. This level of proficiency belongs to those using English very often.

___ Level 7 Have native-like proficiency. Fully comprehend all messages and styles. This level of proficiency belongs to those using all English skills, listening, speaking, reading, and writing integratively.

Writing

___ Level 1 Able to copy the English alphabet

___ Level 2 Able to write from dictations given at a very slow speed, but with errors. Can transform certain words in a sentence, e.g., “She wrote with a pen” to “She wrote with a pencil”.

___ Level 3 Able to write single sentences from memory, but with grammatical and mechanical errors. Can transform certain parts of the content. Still unable to change the style of writing. Need a dictionary.

___ Level 4 Able to write connected sentences under familiar topics, for instance, to report events and give information. Use wrong structures. Very limited skill in varying the style of writing. Need a dictionary.
Writing, continued

___Level 5 Able to write messages of various types, such as explanatory and descriptive prose. Transfers from Thai are still frequent. Write at a slower speed than the native speaker. Need a dictionary.

___Level 6 Able to write all types of English with ease. Need no dictionary. Differ from the native-speaker only in choice of vocabulary to absolutely fit the context. Write at a slightly lower speed than the native speaker.

___Level 7 Able to write at the same level as an educated native speaker. Limited only by background knowledge and intelligence. This level belongs to those who have been in situations in which English is the only medium of communication for a number of years.

Listening

___Level 1 Merely recognize the heard utterances as English languages. May know a single word repeated several times.

___Level 2 Barely understand slow and repeated utterances in a face to face communication. Know certain memorized formulaic expressions, e.g., “hello”.

___Level 3 Know the topics of slow and repeated utterances in a face-to-face communication. Have faulty understanding of messages heard through radio, tape, or telephone.

___Level 4 When listening attentively, will be able to understand clear and simple utterances spoken in a face-to-face situation. Need some repetitions and still unable to grasp some details but able to get the overall meaning through context. Unclear, ambiguous, and slangly expressions are not understood. Lack the ability to appreciate stylish, idiomatic expressions, or humour.

___Level 5 Able to understand native speakers of English when the topics are in one’s area of study. Able to grasp main ideas of lectures. Still unable to comprehend complex arguments in discussions, or stylish and slangy expressions. Have a limited understanding of various language styles and registers.

___Level 6 When listening attentively, will be able to understand all language styles like a native speaker with similar background. Rarely become confused because of tonal variations.
Scales Developed by Wangsotorn (1981), continued

Listening, continued

Level 7 Able to understand every topic spoken in a variety of styles and frequencies like a native speaker with similar background. Comprehend nearly all details even though the messages are not in one’s field of experience. This skill level belongs to those who have lived for a long time in countries where English is the mother tongue and are used to ways of life and customs of the native speakers of English.

Speaking

Level 1 Use Thai intonation to utter certain formulaic expressions, such as “good morning”, and “goodbye”. To understand these utterances the hearer has to be used to the Thai intonation. Have no knowledge of English grammar.

Level 2 Able to read English orally and utter memorized utterances to express one’s needs. Unable to communicate freely. Commit grammatical and phonological errors. Need to repeat oneself several times. Very slow in answering any questions. Need much help from the other party in continuing the conversation.

Level 3 Able to use simple sentences to talk about various topics. Language is still faulty and uncertain. Long pauses are needed before answering any questions. The other party has a difficult time grasping the meaning of the message.

Level 4 Able to communicate about events. But still unable to speak freely enough to argue in discussions. Messages have fewer grammatical errors, but heavy Thai intonation. Pauses are needed because the speaker cannot think of the right vocabulary right off. Generally, the messages are understood by the hearer who may have to help during pauses.

Level 5 Able to communicate in various situations. Capable of expressing one’s views freely. Need preparations for complicated arguments or discussions. Rarely commit errors in grammar. The errors are stylistic most of the time. The other party, at times, may have to ask for clarification.

Level 6 Able to speak on all topics and in all situations. Errors are very few and at times undetectable.

Level 7 Have a native-like proficiency to communicate for all purposes. This level belongs to those who have been residents in countries where English is the mother tongue.
ENDNOTES

1. The Test of English for International Communication (TOEIC) is a test of proficiency in English as a second language, developed by Educational Testing Service (ETS), and introduced in Japan in 1979, at the request of the Ministry of International Trade and Industry, to meet the need for a test of ESL proficiency for use by corporations and other agencies concerned with selecting, placing, and/or training employees for ESL-essential or ESL-desirable positions. TOEIC and TOEIC-related assessment services are now (1995) being offered under the general aegis of The Chauncey Group International (CGI), by professional administrative agencies in more than 25 countries. Test scoring and score reporting are the responsibility of the local TOEIC representative agency involved. TOEIC scores for listening comprehension (LC) and reading (R) are reported on a standard scale (5-495); the Total score is the sum of the two scaled section scores (10-990). An independent review (Perkins, 1987) characterized TOEIC’s psychometric properties, in part, as follows:

"In sum, the TOEIC is a standardized, highly reliable and valid measure of English, specifically designed to assess real-life reading and listening skills of candidates who will use English in a work context. . . . (and) in addition to being an integrative test, the TOEIC also appears to tap communicative competence in that the items require the examinee to utilize his or her sociolinguistic and strategic competence" (p. 82).

For additional detail see, for example, ETS (1986); also Woodford, (1982) for a detailed description of TOEIC’s development and initial validation.

2. For a comprehensive review and evaluation of "self-assessment of second/foreign language proficiency," see Oscarson (1989), an early and persistent advocate (e.g., Oskarsson [Oscarson] 1978, 1980, 1981) who, in 1978, made the point that "... self assessment is a new field in language testing and consequently there is very little accumulated knowledge and experience to draw on" (Oskarsson, 1978: p. 13, emphasis added); but more recently (Oscarson, 1989) stated that there had been "... a surge of interest in methods of self-assessment of foreign language proficiency" (p. 1). The more recent (Oscarson, 1989), comprehensive review suggests continued, marked diversity in the purposes, methods, and lines of inquiry subsumed under the "self-assessment" rubric (a theme which is reinforced by Oscarson's [forthcoming] review, based on examination of pre-publication copy, received too late for full consideration in this paper). For perspective regarding conceptual and methodological diversity see, for example, Ingram (1985), Bachman and Palmer (1989), Blanche and Merino (1989), Heilenman (1990), Xu (1991). For present purposes, it seems fair to say that in the field of second language proficiency assessment, interest in developing and evaluating formal self-assessment "instruments" has not been widespread, and very few studies appear to have employed the same or clearly similar instruments/models. Accordingly questions regarding the types of circumstances (e.g., purposes for which self-assessments are to be used) and demographic/psychological/linguistic/cultural variables that may affect the validity and/or the reliability of particular "self-rating models" have been relatively neglected--a point made explicitly by LeBlanc and Painchaud (1985) with respect to the status of self-assessment generally in the field of second-language testing. In any event, relatively few studies involving formal self-assessment models such as those under consideration in the present study--global scales with levels clearly referenced to those in corresponding ILR rating scales, multi-item, Likert-type "can do" scales, or other similar scales--could be located in searching potentially relevant sources. For perspective on
use and validity of self-assessment in the field of psychology, for example, see a comprehensive
review (Shrauger and Osberg, 1981) of research on the relative accuracy of self-predictions and
judgments by others in psychological assessment.

3. LPI ratings by two interviewers/raters (and a master-rater for discrepant cases) were averaged,
leaving to some "in-between" values (1.25, 1.5, 1.75 and so on). In classifying cases into the five
basic levels, average ratings of .75 thru 1.5 were assigned to Level 1, ratings of 1.75 thru 2.5 were
assigned to Level 2, and so on.

4. Related findings suggested that few non-native-speakers of the target languages received high LPI
ratings (e.g., Level 3 or 1 higher) -- for example, by inference from correlational data, teachers with
higher LPI (and self) ratings tended to be those who reported, for example, that their parents' native
language was the language taught, that they themselves were native speakers of the language
involved, and that they were born in the country of the language taught; single items coded to reflect
the foregoing correlated in the .57 to .53 range with LPI rating -- coefficients higher than those
reported for these background variables with either of the "can do" scales (for listening and
speaking, respectively). Training-related variables, on the other hand, were essentially unrelated to
either the interview rating or the self rating -- for example, correlations with LPI rating ranged
between -.08 and .13 (combined French and Spanish samples) for items involving reported emphasis
on listening/speaking as opposed to reading/writing in high school and college, and high school and
college grades in language. Such findings attest, albeit indirectly, to the construct validity of the LPI
procedure.

5. Abraham, et al. (1988), among other things, collected self-ratings from more than 400
international teaching assistants at Iowa State, for whom scores on a semi-direct measure of
speaking proficiency -- versions of the Test of Spoken English made available for local administration
(SPEAK: ETS, 1982) -- were also available. Shown below are the level descriptions and
corresponding average SPEAK scores (estimated from their Figure 4).

1. I am able to satisfy routine travel needs and minimum courtesy requirements
   SPEAK (Mn = 170)
2. I am able to satisfy routine social demands and limited work requirements
   . . . SPEAK (Mn = 175)
3. I am able to speak the language correctly enough to participate effectively in most
   conversations SPEAK (Mn = 205)
4. I am able to use the language fluently and accurately on all levels for professional
   purposes. . . SPEAK (Mn = 235)
5. I am able to speak as fluently as an educated native speaker of English. . .
   SPEAK (Mn = 275)

The self-assessment schedule included only brief functional summaries for corresponding
basic, FSI speaking proficiency levels. Based on the self-rating/TSE correspondencies suggested
above, according to the report (p. 12), "Examinees seemed well aware of their own English-speaking
capabilities, sorting themselves out in much the same way as did our (SPEAK) raters."

6. This use of the self-assessment scale reflects an operational application predicated on empirical
findings--buttressed by clinical observation--of consistent, positive relationships between LPI ratings of ESL speaking proficiency and scores on the external measure involved (the TOEIC). The rationale for this "instructional" use of ILR-referenced self-ratings has been described by Lindsey (1993, personal communication), as follows: "Although TOEIC candidates (using the nominally LPI-parallel rating schedule) only get a thumbnail sketch of the (formally defined) LPI levels to go by, and they may feel it is in their interest to under/over-estimate their ability, they may reasonably conclude (when later informed regarding the expected correspondence between their TOEIC scores and the selected self-descriptions) that their TOEIC scores seem to be a good reflection of what they themselves feel about their English proficiency. This alone should be reassuring to those of them who might feel that a multiple choice test will not do them justice . . . . or (may tend to) overstate their ability.

7. Among the several studies reviewed herein in which correlations are reported between LPI ratings and self-assessments according to a nominally LPI-parallel rating schedule, Clark and Swinton (1979) reported the lowest coefficient (r = .48). The study involved LPI ratings by interviewer-rater trainees and self-assessments according to a scale written in English, in a nationally/linguistically heterogeneous sample of nonnative-English speaking, international students enrolled in an intensive ESL training program in the United States, and by inference probably functioning at a comparatively "low level" of English proficiency. TOEFL scores were not available for this particular sample, which was used in a pilot study. However, for the conceptually similar sample involved in the main study, Clark and Swinton (1979) reported a TOEFL mean of 434, indicating relatively low standing in basic TOEFL reference groups (e.g., below the 20th centile). Empirical evidence regarding the functional ESL proficiency level in samples at this TOEFL score level does not appear to be available. In any event, it is possible that use of a self-assessment scale written in English may have had attenuating effects on the LPI/self-assessment correlation in this sample; the fact that LPI ratings were provided by trainees in the LPI procedure (as opposed to experienced interviewers/raters), may also have been involved.

8. The amount of time needed to complete an English-language version of the schedule probably would tend to be greater than that needed to complete a native-language version--a pragmatically important consideration in operational contexts. However, in some testing contexts (e.g., intensive ESL training programs) self-assessments might be obtained on a "take-home" basis, thus not requiring any formal allocation of limited assessment time. In any event, according to Lindsey (1995, personal communication), the candidates involved in this study, most of whom used a native-language version of the relatively lengthy rating schedule, completed the relatively lengthy self-assessment item quickly, just prior to taking the TOEIC.

9. According to the study report, the self-assessment inventories were used for initial placement decisions which were subject to revision after observation of students' proficiency in the actual placement settings. Incidence of change in initial placement was generally comparable to that observed when locally developed proficiency tests were used. In a personal communication in May, 1994, Professor LeBlanc indicated that self-assessments were still being used as the primary basis for placement at the University. Students not natively-fluent in either French or English, indicate which of these languages is their better second-language and use the inventory written in that language, according to Professor LeBlanc (personal communication, 1996) who indicated that questions regarding possible differences in validity for self-assessments provided by these students...
as opposed to those using a native-language version of the self-assessment inventory, have not been addressed systematically. Generally speaking it would seem that when self-assessment of second-language proficiency is provided for placement purposes, under conditions providing for change in placement based on direct observation, there is little or no incentive to deliberately over- or underestimate one's level of proficiency.

10. No "in-house" adaptations of the four ILR/FSI scales for self-assessment purposes appear to be available, although Ingram (1985) cites unpublished FSI-related self-assessment findings dating back to the 1970s. In response to a query regarding the foregoing, Madeline Ehrman (1995, personal communication), Director of Research and Development, FSI Language School, indicated that she knew of no such adaptations involving agencies within the ILR. The following historical note (Bruhn, 1989: 246-247) appears to be pertinent in connection with the foregoing:

"The development of the so-called FSI guidelines and FSI test was promoted in the 1950s (in response to legislation directing) the Civil Service Commission to inventory and develop a register of persons in government with foreign language skills. . . . It began as a self-rating based on six language user descriptions (without differentiation of skill domains). . . . The self-rating proved to be unreliable and the use of a single rating unworkable" (emphasis added).

Thus, an unfavorable initial experience with self-assessment, based on guidelines "(lacking) differentiation of skills or components", may to some extent be responsible for what appears to have been lack of interest within the FSI itself in self-assessments based on adaptations of the well-differentiated quasi-absolute scales subsequently developed by FSI staff for assessment of the four basic second-language macroskills.