DOCUMENT RESUME

ED 343 442 FL 020 293

AUTHOR Kenyon, Dorry Mann; Stansfield, Charles W.
TITLE Examining the Validity of a Scale Used in a Performance Assessment from Many Angles Using the Many-Faceted Rasch Model.
INSTITUTION Center for Applied Linguistics, Washington, D.C.
PUB DATE Apr 92
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (73rd, San Francisco, CA, April 20-24, 1992). For a related document, see FL 020 294.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Software; Elementary Secondary Education; Guidelines; *Language Proficiency; *Language Tests; Models; *Oral Language; *Rating Scales; Second Language Learning; Speech Skills; Surveys; *Test Validity; Verbal Ability
IDENTIFIERS *American Council on the Teaching of Foreign Langs; *Rasch Model; Texas Oral Proficiency Test

ABSTRACT

An attempt is made is this paper to examine the validity of the American Council on the Teaching of Foreign Languages (ACTFL) scale through a comparison of the scaling of speaking tasks and speech performances by the scale and by a Rasch analysis of judgments made by "naive" persons. The results of the multi-faceted Rasch analysis seem to support the use of the scale in assessing developing second language proficiency. The unifying element was the underlying ACTFL scale. The results indicate a tendency towards convergence of the judgments made by "naive" judges across three different groups, made during separate phases of the test development project, made on different aspects of the project, and made using different methods of indicating decisions with the ACTFL scale. It is concluded that the use of the ACTFL Proficiency Guidelines is justified for developing performance-based assessments of speaking ability. Documentation is presented in 10 tables, and appendices provide: (1) the Structure of the Texas Oral Proficiency Test (TOPT)--Spanish; and (2) the TOPT Bilingual Education Teachers Job-Relatedness Survey. Contains 16 references. (LB)

Examining the Validity of a Scale used in a Performance Assessment From Many Angles Using the Many-Faceted Rasch Model

Dorry Mann Kenyon

and

Charles W. Stansfield

Center for Applied Linguistics

1118 22nd Street, NW

Washington, DC 20037

(202) 429-9494

CAL@GUVAX

Paper presented at the annual meeting

of the American Educational Research Association

San Francisco, CA

April, 1992

Purpose

This study has two interrelated purposes. The first is to examine the validity of a widely-used scale of foreign language speaking ability through comparisons of scaling based on judgments made by "naive" judges with *a priori* scaling determined by "experts" on the basis of the established scale. The second purpose is to illustrate the application of the many-faceted Rasch model as a method of scaling.

Background to the Language Scale

The Proficiency Guidelines of the American Council on the Teaching of Foreign Languages (ACTFL)

> represent a hierarchy of global characterizations of integrated performance in speaking, listening, reading and writing. Each description is a representative, not an exhaustive, sample of a particular range of ability, and each level subsumes all previous levels, moving from simple to complex in an 'all-before-and-more' fashion. (ACTFL, 1986)

The ACTFL Guidelines have been widely used in the field of foreign language education in the United States since their original publication in 1982. A bibliography published in 1988 included over 400 articles in the literature focusing on the Guidelines and their application in measurement and teaching (Stansfield & Thompson, 1988; cf. Galloway et al, 1987). The Guidelines, by providing an *a priori* description of developing foreign language competence, have served as the basis for the widely-used, face-to-face tailored assessment of foreign language speaking ability known as the Oral Proficiency Interview (OPI). The Guidelines also form the basis for a series of tape-mediated speaking tests known as Simulated Oral Proficiency Interviews (SOPIs) developed by the Center for Applied Linguistics (Stansfield, 1989). In this performance-based assessment of speaking ability, the Guidelines guide both the development of the speaking tasks (i.e., the items) that appear on the test and the scoring of examinee performance.

The Guidelines describe foreign language proficiency at four main levels: Novice, Intermediate, Advanced and Superior. They also describe sublevels within the first three main levels. Table 1 presents the entire range of 9 level descriptions from lowest to highest.

Insert Table 1 About Here

Inherent in each description are the types of speaking tasks speakers at each level of ability can accomplish. Thus, an Intermediate Low level speaker can "perform such tasks as introducing self, ordering a meal, asking directions and making purchases" (ACTFL, 1986). Superior level speakers can, for example, "discuss special fields of competence and interest with ease" (ACTFL, 1986).

The Validity of the Guidelines

The development of the Guidelines dates back to the 1950s within the School of Language Studies at the Foreign Service Institute. The predecessor to the OPI, with accompanying scale level descriptors, was developed in response to the practical need of assessing the language performance of members of the United States diplomatic service corps. The Guidelines have been further refined within government agencies since that time by the coordinated efforts of an interagency committee acting under the auspices of the Federal Interagency Language Roundtable (FILR). Beginning in the early 70s, the Educational Testing Service (ETS) adapted the government work for use in the Peace Corps, and in the early 80s, ACTFL adapted the government work for use in academic settings. ACTFL disseminated the revised scale under the name of the ACTFL Guidelines (Lowe, 1988).

Despite the wide dissemination and application of the Guidelines and their demonstrated practical utility, their validity as a description of developing competence in a second language has been widely contested. Many have directly challenged their validity (e.g., Bachman & Savignon, 1986; Lantolf & Frawley, 1985), while others have cited the lack of research to validate the scale levels (Clark & Lett, 1988).

Context of the Current Studies: The Development of a Performance-based Assessment

This paper examines the validity of the ACTFL scale as a description of developing foreign language proficiency through two studies that compare judgements made by "naive" judges and by "expert" judges. The first compares the scaling of speaking tasks on the basis of the judgements of "naive" judges with *a priori* scaling of those tasks according to the ACTFL Guidelines. The second compares the scaling of speech performances by "naive" judges with the scaling of those performances by "experts" informed by the ACTFL Guidelines.

Data for the studies were collected during the development of the Texas Oral Proficiency

Test (Stansfield and Kenyon, 1991). Versions of the TOPT were developed in Spanish and French by the Center for Applied Linguistics (CAL) under contract with the Texas Education Agency (TEA). The TEA began using the TOPT for teacher certification purposes in November, 1991. The test is a SOPI, consisting of fifteen speaking tasks. The development of these tasks (items) was guided by the descriptions contained in the ACTFL Guidelines. In addition, the scoring of the test is also based entirely on the ACTFL scale. The data for the first study reported here was collected during a job-relevancy study conducted before the actual writing of test items began. The data for the second study reported here comes from a standard setting study conducted after the TOPT had been field tested and revised. The link between the TOPT development project and both these studies is the underlying ACTFL scale.

The Use of the Many-Faceted Rasch Model as a Method of Scaling

Although there are various approaches to and methods of scaling (e.g., Torgerson, 1958), the method used here is a many-faceted Rasch approach. Rasch methodology has provided practitioners with useful tools in the analysis of scales (e.g. Wright & Masters, 1982). The two studies reported in this paper provide an illustration of the information that may be gained from applying one of the newest Rasch computer programs, FACETS (Linacre & Wright, 1990) in a scalar analysis. FACETS was the only computer program that could adequately analyze the three facets involved in the study, handling both scalar and dichotomous data. Since the ACTFL scale assumes an underlying unidimensional trait of developing second language proficiency, it appears appropriate to consider using a Rasch model.

It is important to clarify that the motivation of the Rasch model is measurement construction, not data description. Accordingly, although the original data was not produced in an effort to build a measure, the analyses and interpretations in this paper will be in terms of measurement construction. The interpretation of support for the validity of the ACTFL Guidelines will be presented in the context of 1) whether the analysis shows evidence for the existence of an underlying scale that conforms to the Guidelines, and 2) whether further measurement construction, as indicated by information provided by the Rasch analysis, indicates development in a direction moving closer to the ACTFL scale or not.

Study 1: The Scaling of Speaking Tasks

In the TOPT, examinees are asked to perform 15 speaking tasks ranging from "giving

directions" to "supporting an opinion." Each of these tasks is *a priori* designated at one of the main proficiency levels on the ACTFL scale. (Novice level tasks were not included on the TOPT since it was assumed that teacher certification candidates would all be above that level.) Each item on the TOPT is designed to elicit performance at the ACTFL level associated with the item's speaking task. As an example, Appendix A presents an outline of the 15 speaking tasks on the Spanish TOPT and their levels on the ACTFL scale.

In the first phase of the test development project, a job relevancy survey was conducted to determine the relevancy of 38 individual speaking tasks. The survey presented teachers with a brief description of each speaking task and asked them to rate each on a five-point scale in response to the following question: *Is the level of ability required to perform this task needed by bilingual education* (Spanish language/French language, changed as appropriate) *teachers in Texas public schools?* A booklet sent with the survey contained the label for each task, followed by a more complete description of it (Appendix B). Teachers indicated their response on a machine-scoreable answer sheet. A rating of 5 indicated "Definitely Yes," 4 meant "Probably Yes," 3 meant "Maybe," 2 meant "Probably No," and 1 meant "Definitely No."

700 teachers from throughout the state of Texas were chosen in a geographically stratified random sampling design to receive the survey: 400 bilingual education teachers, 200 Spanish language teachers and 100 French language teachers. Four hundred two (402) teachers returned the survey for a response rate of 57%. Table 2 presents a summary of the demographic information of those returning the survey. It reveals an adequate response rate (57%) which was consistent across all three groups of teachers. In terms of the experience of the teachers and their sex, little difference appears across the three groups. In terms of educational level taught, Table 2 reflects the fact that bilingual education is offered only in K through 5th grade in Texas. In terms of ethnicity of respondents, there is great, though expected, variation among the groups. TEA staff and members of the test advisory boards felt that, based on the demographic data, the survey results may be seen as an accurate reflection of each group.

Insert Table 2 About Here

For the purposes of the test development project, all speaking tasks that received a mean rating above 3.50 were considered acceptable to appear on the TOPT. For this paper, the complete data matrix of ratings containing three facets [teachers, group (bilingual education,

Spanish or French), and speaking task] was analyzed using FACETS.

Table 3 presents the *a priori* classification of the 38 speaking tasks into the three highest main levels on the ACTFL scale. Within each level, tasks are listed in alphabetical order. These classifications were made by the test developers and were based on, as primary references, the ACTFL Guidelines (ACTFL, 1986) and the FILR Skill Level Descriptions (Liskin-Gasparro, 1987). As a secondary reference, Omaggio's influential text, Teaching language in context (Omaggio, 1986), was also used to classify the tasks.

Insert Table 3 About Here

Table 4 presents the results of the scaling of the 38 speaking tasks by the FACETS program. The reliability of the tasks measure was .99, with the scale extending about 3.50 logits. In the context of the survey, an "easier" task would receive a higher average rating, indicating more teachers felt that a Texas classroom teacher should have the ability to perform this task. Thus, tasks with a higher logit value may be considered as requiring less proficiency to perform, while tasks with a lower logit value may be considered as requiring greater proficiency.

Insert Table 4 About Here

Table 5 presents a comparison of the ranking of the 38 speaking tasks based on the FACETS analysis with their ranking based on their *a priori* designations. If the two rankings had completely matched (within measurement error), then the 12 tasks identified *a priori* as Intermediate would have been the first 12 tasks, the 14 tasks identified *a priori* as Advanced would have been the middle 14 tasks, and the 12 tasks identified *a priori* as Superior would have been the last 12 tasks on the FACETS scale. The bottom line of Table 4 would have shown 100% agreement in each category.

Insert Table 5 About Here

Table 5 presents a "best case" scenario. That is, the scaling takes measurement error into consideration so that "Order a Meal," designated an Intermediate task but located among the middle 14 tasks has been exchanged with "Compare and Contrast Two Objects or

Places," designated an Advanced task but located among the first 12 tasks. Similarly, "Propose and Defend a Course of Action" has been exchanged with "Lodge a Complaint."

Tables 4 and 5 indicate that while Superior tasks, in general, were scaled by the Texas teachers as expected, Intermediate and Advanced level tasks seem to be totally intertwined. This will be discussed further below.

One of the advantages of the Rasch model is the ability to examine fit and to incorporate information to continually assess and improve the quality of a measure. In an analysis of the fit of the tasks, a liberal criterion was adopted, as the purpose of the original project was not to construct a measure. Infit and outfit mean squares with a value greater than 1.4 or lower than 0.6 have been marked with an asterisk (*) in Table 4. In general, the tasks have scaled very well. This lends evidence to the hypothesis that a unidimensional construct underlies these data. Four tasks, however, have clearly misfitting infit and outfit values, and are problematical to the scale. These are: "Introduce Yourself," "Talk about Family Members," "Order a Meal," and "Make Purchases."

One of the facets in this analysis was group membership. Table 6 shows the results for this facet. The calibration logit indicates that the least and the most severe groups differed by less than 0.06 logit. This difference is not much greater than the model error (0.03) for the most lenient group, the French teachers. Thus, group membership of the teachers did not contribute much to the overall severity of the scaling of the tasks. In terms of fit, the outfit statistic is bordering on extreme for the French and bilingual education groups. This may indicate that members of those groups viewed the underlying construct differently.

The last facet was the teacher. Applying widely-used criteria of fit to the teachers indicated much misfit. In terms of the outfit mean squared statistic, of the 380 teachers without perfect ratings, 50% were "misfitting" when the criterion was above 1.3 or below 0.7. However, for the standardized outfit statistic (which is sensitive to sample size), only 36 teachers (8.9%) had a statistic above 2.0 or below -2.0. Of these, 75% were bilingual education teachers and 25% were Spanish language teachers. None of the French teachers were "misfitting" according to this criterion.

An analysis of the individual misfitting ratings is also possible. Of the 15,206 valid individual ratings, FACETS identified 195 (1.4%) as misfitting. These ratings were tabulated across the three facets to see if any consistent inconsistencies were present. When misfitting individual ratings consistently involve certain tasks, certain teachers, or certain

groups, individual problems may be highlighted. The three subtables in Table 7 present the results of these tabulations.

Insert Table 7 About Here

Table 7A indicates that 28% of the teachers had one or more misfitting ratings. However, only 9 teachers (2%) had 4 or more of their 38 ratings identified as misfitting. In other words, among the individual teachers, there does not appear to have been a significant cluster whose ratings were out of step with the entire group.

Table 7B shows that, on the task facet, there were consistencies in misfitting ratings. Table 7B reflects what was previously gleaned from the infit and outfit mean square statistics. The four tasks with the most misfitting ratings are those whose fit statistics were inadequate by the criterion used above. Five tasks were not involved in a misfitting rating.

Table 7C indicates that, in terms of group membership, the number of misfitting ratings ascribed to the bilingual education group is disproportional to their size in the total population. In other words, there seems to be a tendency for the bilingual education teachers to award a greater number of misfitting ratings than for the French or Spanish language teachers.

Discussion of Study 1

As pointed out earlier, this discussion is in the context of test construction rather than data analysis. However, we do want to examine whether the ordering of the tasks by the randomly sampled Texas classroom teachers across three disciplines reflects the ordering based on the ACTFL scale. The FACETS program has provided much information that is useful to understanding what happened in this survey.

First, let's discuss the four misfitting speaking tasks. These are presented in Table 8. All were *a priori* designated as at the Intermediate level. Upon closer examination, two of them ("Make Purchases" and "Order a Meal") seem very different from the other 36 tasks (see Appendix B). These two tasks seem very concrete and less linguistically dependent. One may be able to fulfill these tasks in another country through signs and gestures without knowing any foreign language at all. On the other hand, each could potentially involve complications requiring much linguistic skill. Perhaps teachers had trouble picturing just

how much ability would be involved in performing these tasks. This could account for the misfit.

The two other tasks ("Introduce Yourself" and "Talk About Family Members") seem to involve a much more personal dimension that the other 36 tasks. It may also be noted that "Introduce Yourself" was located as item number 1, and very few teachers awarded it less than a "5." In summary, these four speaking tasks, as presented in the survey, appear to be of a slightly different nature than the majority.

Insert Table 8 About Here

However, even with discounting the four misfitting tasks, there is still an intermingling of the Intermediate and Advanced level tasks in the scaling. If the ACTFL scaling were valid, why might this have happened? First, in the set of speaking tasks as presented to the teachers, "text form," which is one of the characteristics that distinguishes the different main levels of the scale, appears to have been inadequately incorporated into the task descriptions, if at all. Briefly, Intermediate level speakers use "sentence-level" discourse. Their tasks can be accomplished at a sentence-level. Advanced level speakers use "paragraph-level" discourse. To carry out tasks at the Advanced level, more elaborate and more organized speech is required. Tasks at the Superior level require an extended level of discourse, in which thoughts are elaborated into "paragraphs" and these are solidly well-connected and organized to get meaning across.

This aspect of the response was not taken into account in the description of some of the speaking tasks that were designated *a priori* as Intermediate but scaled by the teachers as requiring much ability to perform. For example, "Describe Health Problems" was designated *a priori* as Intermediate. As an Intermediate level task, however, the expectation is that one can say, at the sentence level, "I have a pain in my stomach," but not necessarily go into great detail. In completing this survey, teachers may well have pictured much more complicated discourse. Similarly, the survey did not make clear that the expectation for fulfilling other high-ranking Intermediate level tasks, such as "Talk About Your Future Plans," "Make Arrangements for Future Activities," and "Give a Brief Personal History," was simple sentence-level discourse. Conversely, the Advanced level task "Express Personal Apologies" was scaled by the teachers as rather easy. For many teachers that task can be accomplished with a short "I'm sorry" at the sentence level. The survey did not clearly indicate that any elaboration (required of an Advanced level designation) was involved.

On the other hand, description of the Superior level tasks on the survey tended to convey the idea of complexity using words and phrases such as "abstract," "complex," "controversial," "explain in detail," and "discuss at length."

There is evidence that a second trend was also operating among this set of teachers that worked to place certain Advanced speaking tasks lower on the scale than expected. Three of the four easiest ranked Advanced tasks ("Give Instructions, "Describe Typical Routines," and "Explain a Familiar Simple Process") are tasks that may actually occur in the classroom on a frequent basis. Thus, when the teachers were asked whether a teacher in Texas needed the ability to perform this task, they would have ranked these as "5", "Definitely Yes." This may be particularly true for the bilingual education teachers. It is interesting that "Give Instructions" was ranked second in the scaling. In terms of linguistic ability, it cannot really be completed at the sentence level, since in most cases elaboration would definitely be required.

In summary, the FACETS analysis has revealed a wealth of information for helpful in understanding what may have been going on in this survey. The "naive" teachers did perceive a single trait as underlying the tasks. Where the task description matched the intent of the Guidelines, results were as expected. In our opinion, this study does provide evidence to support the validity of the Guidelines as a scale of speaking ability. Were such a survey to be undertaken again and a greater effort made to better match the task descriptions to the levels of the Guidelines, we believe that naive language teachers would even more closely scale the speaking tasks in accordance to the Guidelines.

Study 2: The Scaling of Speakers

As part of the test development project, CAL conducted three separate standard setting studies following the model described in Livingston (1978) and adapted by Powers and Stansfield (1982) in order to provide additional data to assist the TEA and the Texas State Board of Education in setting passing scores for the test. These studies required a sample of examinee performances at known levels and a panel of judges to rate the performances as acceptable or unacceptable.

Examinee performances were selected according to the following procedure. First, two Texas ACTFL-certified testers for Spanish and French independently assigned a rating on each of the 15 TOPT speaking tasks to approximately 40 examinees. The examinee tapes had been recorded during the field testing of the TOPT. After these ratings were examined,

three tasks were selected from 25-31 examinees to be indicative of various level of speech performance between Intermediate Mid and Superior on the ACTFL scale. These were edited onto a preliminary tape, which contained the words "This is Speaker X," followed by that speaker's performance on the three speaking tasks. The preliminary tapes for French and for Spanish were each sent to five ACTFL-certified testers for independent confirmatory ratings. Only those speakers for which at least three of the five raters agreed with the original level description were retained. The final tape for the French TOPT contained 17 speakers and for the Spanish TOPT, 22 speakers. For each speaker, the original level established by the two Texas judges and independently confirmed by five additional judges from the confirmation study designated the *a priori* ACTFL level for that speaker's performance.

These master tapes were played to representative groups of judges selected by the TEA from throughout the state of Texas. One group was for French language teaching, one for Spanish language teaching, and one for bilingual education. As the judges heard each speaker, they were asked to indicate whether or not the speaker demonstrated enough second language ability to perform successfully in a Texas public school classroom. The response options were "Yes" or "No." The mean number of positive responses across the examinees at each different level assigned *a priori* was presented to the TEA to assist them in setting a passing score for the TOPT.

For this paper, the data was analyzed using a multi-faceted Rasch analysis to scale the speakers from the Master tape. These are then compared with the *a priori* scalings according to the ACTFL guidelines. The ratings of the French speakers and the ratings of the Spanish speakers are considered separately.

Thirty judges made dichotomous decisions for the Spanish speakers (17 for the Spanish study and 13 for the bilingual education study). Sixteen judges rated the French speakers. Table 9 presents a summary of the demographic information on these judges.

Insert Table 9 About Here

Table 9 reveals that the greatest difference among the groups was that the bilingual education and Spanish judges (who listened to the Spanish examinees) were much more likely to be Hispanic than the French teachers.

Table 10 presents the results of the scaling of the speakers on each tape by the FACETS program. Table 10A presents the results for the French speakers. The reliability of the French speakers' measure is .87. The scale extends almost 9 logits. The speakers perceived to have greater ability have a higher logit value. Nine of the 19 speakers received perfect scores, indicating that all judges agreed that they demonstrated enough ability to perform in a Texas public school classroom. Although these cannot be ranked in comparison to each other, they have been presented in the table according to the *a priori* ordering assigned by the ACTFL-certified testers.

Insert Table 10 About Here

Table 10B presents the results for the Spanish speakers. Four of the 22 speakers received perfect scores. The reliability of the measures is .93. The length of the logit scale is about 8.50 logits.

The two subtables in Table 11 are similar to Table 4. They present comparisons of the ranking of the speakers tasks based on the FACETS analysis with their ranking based on their *a priori* designations.

Insert Table 11 About Here

As with Table 4, Table 11 presents a "best case" scenario. Speakers receiving the maximum score are ordered according to the *a priori* designations. Measurement error has been taken into consideration such that, when possible, speakers have been re-ranked to be in the *a priori* ordering. This has occurred with one Intermediate Mid level French speaker and one Superior level Spanish speaker.

Table 11A indicates that, under the "best case" scenario, the French judges ranked the speakers in the same order as the experts. There was more disagreement for the Spanish speakers. However, it may be noted that no Intermediate level speaker was ranked in the Advanced level, or vice versa, and only one Superior level speaker ranked in the Advanced level. It may be noted that ACTFL considers a misrating within a main level to be of lesser importance than one between main levels.

In examining the fit, there were only two individual misfitting ratings for the French

TOPT, both involving different judges (4 and 12) and different speakers (15 and 18). In comparing this information with Table 10, it can be seen how sensitive the mean square fit statistics were in this situation. Speakers 15 and 18 have the highest combined infit (both 1.6) and outfit (1.3 and 2.3, respectively) statistics of the group. This extreme sensitivity may be due to the fact that the raw data, upon closer examination, is very close to forming a deterministic Guttman scale. When this happens, the fit statistics of the probabilistic Rasch model show extreme sensitivity to outliers (Linacre, personal communication). This situation is the same for Judges 4 and 12. The standardized statistic, however, indicates no problems with fit. None of the individual judges and none of the individual speakers appear problematic. This lends support to the argument that these judges were ranking these speakers on a unidimensional construct of ability to speak French.

For the Spanish speakers, nine individual ratings were misfitting. Five of these were in the bilingual group, and 4 in the Spanish group. Three of them involved Judge 6 from the bilingual education group. The rest involved different judges. This is reflected in the fit statistics for the judges. Judge 6 has an infit mean square fit statistic of 3.3 and outfit mean square of 3.9. This judge also had the only standardized fit statistic above 2.

None of the speakers were involved in more than one misfitting rating. Although Table 10 shows some rather large mean square statistics, it again appears that these are due more to the fact that the raw data was very close to approaching a Guttman scale. None of the standardized fit statistics reveal any problem with fit.

The group facet also showed no misfit. In terms of the severity of judgement, the bilingual education group was slightly more severe, with a logit value of 0.25 (error of .21), while the Spanish group's value was -0.25 (error of .18). Given the seven- logit scale and the size of the error, there was very little actual difference in their severity. This analysis of fit lends support to the argument that the Spanish and bilingual judges were also ranking the Spanish speakers on a unidimensional construct of ability to speak Spanish.

Discussion of Study 2

The results for the French speakers appear to provide support for the validity of the Guidelines, although due to perfect scores the scaling actually effected only three sublevels rather than five. Study 2 also presents support for the main level distinctions of the Guidelines, though there were unexpected rankings within the sublevels. These, however, may have been due to the presence of many Hispanics, who may consider Spanish as their

native language, among the both the judges and the Spanish speakers on the tape. In making a judgment about such speakers, other, non-linguistic, standards may have been used by Hispanic judges. This possibility would need to be further investigated.

Through the process by which the master tape was created in Study 2, the *a priori* levels assigned to the speakers was very closely aligned with the intent of the ACTFL Guidelines, and the comparison of the results was also closer to what was expected. This supports the contention that if the tasks in Study 1 had more appropriately matched the descriptions of the Guidelines, the results would have shown a closer agreement between the rankings based on the ratings of the "naive" teachers and the *a priori* ACTFL scale designations.

Conclusions

An attempt was made in this paper to examine the validity of the ACTFL scale through a comparison of the scaling of speaking tasks and speech performances by the scale and by a Rasch analysis of judgments made by "naive" persons. In our opinion, the results of the multi-faceted Rasch analyses support the use of the scale in assessing developing second language proficiency. The unifying element across these two studies and the entire test development project was the underlying ACTFL scale. The results indicate a tendency towards convergence of the judgements made by "naive" judges across three different groups, made during separate phases of the test development project, made on different aspects of the project, and made using different methods of indicating decisions, with the ACTFL scale. We believe the results also support the use of the ACTFL Proficiency Guidelines to guide the development of performance-based assessments of speaking ability.

REFERENCES

American Council on the Teaching of Foreign Languages. (1986). ACTFL Proficiency Guidelines. Hastings-on-Hudson, NY: ACTFL.

Bachman, L.F., & Savignon, S.J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. The Modern Language Journal, 70, 380-90.

Clark, J.L.D., & Lett, J. (1988). A research agenda. In P. Lowe, Jr., and C.W. Stansfield. (Eds.), Second language proficiency assessment: Current issues. Englewood Cliffs, NJ: Prentice Hall Regents.

Galloway, V., Stansfield, C.W., & Thompson, L. (1987). Topical bibliography of proficiency-related issues. In C.W. Stansfield & C. Harmon (Eds.), ACTFL Proficiency Guidelines for the less commonly taught languages. Washington, DC: Center for Applied Linguistics and ACTFL. (ERIC Document Reproduction Service, ED 289 345).

Lantolf, J.P. & Frawley, W. (1985). Oral proficiency testing: a critical analysis. The Modern Language Journal, 69, 337-45.

Linacre, J.M., & Wright, B.D. (1990). FACETS: Rasch-Model Computer Program Version 2.4. Chicago: MESA Press.

Liskin-Gasparro, J. (1987). Testing and teaching for oral proficiency. Boston, MA: Heinle and Heinle Publishers.

Livingston, S.A. (1978). Setting standards of speaking proficiency. In J.L.D. Clark (Ed.), Direct testing of speaking proficiency: Theory and application. Princeton, NJ: Education Testing Service.

Lowe, P., Jr. (1988). The ACTFL proficiency guidelines: The unassimilated history. In P. Lowe, Jr., and C.W. Stansfield. (Eds.), Second language proficiency assessment: Current issues. Englewood Cliffs, NJ: Prentice Hall Regents.

Omaggio, A.C. (1986). Teaching language in context. Boston: Heinle & Heinle.

Powers, D.E. & Stansfield, C.W. (1982). The Test of Spoken English as a measure of communicative proficiency in the health-related professions (TOEFL Research Report 13). Princeton, NJ: Educational Testing Service.

Stansfield, C.W. (1989). Simulated Oral Proficiency Interviews. ERIC Digest. Washington, DC: Center for Applied Linguistics.

Stansfield, C.W., & Kenyon, D.M. (1991). Development of the Texas Oral Proficiency Test (TOPT): Final report. Washington, DC: Center for Applied Linguistics. (ERIC Document Reproduction Service, ED 332 522)

Stansfield, C.W., & Thompson, L. (1989). Topical bibliography of proficiency-related publications: 1987-1988. In K. Buck (Ed.), The ACTFL Oral Proficiency Interviewer Tester Training Manual. Yonkers, NY: American Council for the Teaching of Foreign Languages.

Torgerson, W.S. (1958). Theory and methods of scaling. New York: John Wiley & Sons.

Wright, B.D., & Masters, G.N. (1982). Rating scale analysis: Rasch measurement. Chicago: MESA Press.

Table 1
Level Descriptors of the ACTFL Scale

Main Level	Sublevels
NOVICE	Novice Low
	Novice Mid
	Novice High
INTERMEDIATE	Intermediate Low
	Intermediate Mid
	Intermediate High
ADVANCED	Advanced
	Advanced High
SUPERIOR	Superior

Table 2
TOPT Job-relevancy Survey Sample: Summary of Demographic Information

TOTAL NUMBER OF SURVEYS SENT: 700

Bilingual Education (BE) Teachers	400
Spanish Language (SP) Teachers	200
French Language (FR) Teachers	100

TOTAL NUMBER OF VALID RETURNED SURVEYS: 402 (57%)

	n	%Ret'd	% of Responses Total Group
Bilingual Education	229	57%	57%
Spanish	113	57%	28%
French	60	60%	15%

LEVEL TAUGHT:

	BE	SP	FR
Elementary	96%	14%	0%
Jun High/Middle	1%	21%	14%
High School	1%	65%	86%
Other	2%	0%	0%

EXPERIENCE:

	BE	SP	FR
1-5 years	41%	37%	34%
6-10 years	28%	24%	25%
11-15 years	20%	17%	20%
16+ years	11%	22%	21%

SEX:

	BE	SP	FR
Male	10%	18%	12%
Female	90%	82%	88%

ETHNICITY:

	BE	SP	FR
Hispanic	87%	43%	9%
White	11%	52%	89%
Black	1%	3%	2%
Other	1%	2%	0%

Table 3
A Priori Scaling of the 38 Speaking Tasks
Used in the TOPT Job-relevancy Survey

INTERMEDIATE TASKS

Describe a Place
Describe Health Problems
Describe Your Daily Routine
Give a Brief Personal History
Give Directions
Introduce Yourself
Make Arrangements for Future Activities
Make Purchases
Order a Meal
Talk About Family Members
Talk About Personal Activities
Talk About Your Future Plans

ADVANCED TASKS

Compare and Contrast Two Objects or Places
Correct an Unexpected Situation
Describe a Sequence of Events in the Past
Describe Expected Future Events
Describe Habitual Actions in the Past
Describe Typical Routines
Explain a Familiar Simple Process
Express Personal Apologies
Give a Brief Organized Factual Summary
Give Advice
Give Instructions
Hypothesize About a Personal Situation
Lodge a Complaint
State Advantages and Disadvantages

SUPERIOR TASKS

Change Someone's Behavior through Persuasion
Describe a Complex Object in Detail
Discuss a Professional Topic
Evaluate Issues Surrounding a Conflict
Explain a Complex Process in Detail
Explain a Complex Process of a Personal Nature
Give a Professional Talk
Hypothesize About an Impersonal Topic
Hypothesize About Probable Outcomes
Propose & Defend a Course of Action with Persuasion
State Personal Point of View (Controversial Subject)
Support Opinions

Table
Scaling of the 38 Speaking Tasks by the FACETS Program

Level	Tasks	Measure Logit	Model Error	Infit MnSq	Outfit MnSq
(I)	Introduce Yourself	2.98	0.15	1.8*	4.6*
(A)	Give Instructions	2.63	0.13	1.1	0.8
(A)	Describe Typical Routines	2.04	0.10	1.2	1.0
(I)	Give Directions	2.02	0.10	1.2	0.9
(A)	Describe a Sequence of Events in the Past	1.89	0.09	0.9	0.7
(A)	Explain a Familiar Simple Process	1.80	0.09	1.3	1.4
(I)	Describe Your Daily Routine	1.68	0.09	1.3	1.3
(I)	Describe a Place	1.59	0.08	1.1	0.9
(A)	Express Personal Apologies	1.44	0.08	1.1	1.5*
(I)	Talk About Family Members	1.38	0.08	1.8*	1.6*
(A)	Describe Expected Future Events	1.33	0.08	0.8	0.7
(A)	Compare and Contrast Two Objects or Places	1.05	0.07	1.0	0.9
(I)	Order a Meal	1.01	0.07	1.9*	2.2*
(I)	Talk About Personal Activities	1.00	0.07	1.2	1.1
(I)	Give a Brief Personal History	0.97	0.07	1.0	1.0
(I)	Make Purchases	0.91	0.07	1.6*	2.8*
(A)	Give a Brief Organized Factual Summary	0.87	0.07	0.9	0.9
(I)	Make Arrangements for Future Activities	0.79	0.07	0.9	0.8
(A)	Give Advice	0.77	0.07	0.8	0.8
(A)	State Advantages and Disadvantages	0.63	0.06	0.7	0.6
(A)	Describe Habitual Actions in the Past	0.62	0.06	1.1	1.0
(I)	Describe Health Problems	0.55	0.06	1.0	0.9
(S)	Change Someone's Behavior through Persuasion	0.44	0.06	1.1	1.0
(I)	Talk About Your Future Plans	0.42	0.06	1.0	0.9
(S)	Support Opinions	0.39	0.06	0.7	0.6
(S)	Propose & Defend a Course of Action with Persuasion	0.35	0.06	0.9	0.8
(A)	Lodge a Complaint	0.28	0.06	0.6	0.6
(S)	State Personal Point of View (Controversial Subject)	0.24	0.06	0.8	0.8
(A)	Hypothesize About a Personal Situation	0.21	0.06	0.6	0.6
(A)	Correct an Unexpected Situation	0.13	0.06	0.8	0.7
(S)	Hypothesize About an Impersonal Topic	0.11	0.06	0.8	0.8
(S)	Hypothesize About Probable Outcomes	-0.02	0.06	0.8	0.8
(S)	Evaluate Issues Surrounding a Conflict	-0.18	0.06	0.8	0.7
(S)	Discuss a Professional Topic	-0.19	0.06	0.9	0.9
(S)	Explain a Complex Process of a Personal Nature	-0.31	0.06	0.8	0.8
(S)	Explain a Complex Process in Detail	-0.33	0.06	1.2	1.1
(S)	Give a Professional Talk	-0.41	0.06	1.3	1.3
(S)	Describe a Complex Object in Detail	-0.48	0.06	1.1	1.1

* Inadequate fit

Table 5
Comparing the A Priori Classifications With the Actual Scaling "Best Case Scenario"

A Priori Expected Ordering	Actual Scaling		
	I	A	S
I (12)	6	6	0
A (14)	6	6	2
S (12)	0	2	10
Correct Order	6 (50%)	6 (43%)	10 (83%)

Table 6
Results of Analysis of the "Groups" Facet

Groups	Calib. Logit	Model Error	Infit MnSq	Outfit MnSq
French	-0.68	0.03	0.8	0.7
Spanish	-0.73	0.02	0.9	1.0
Bilingual	-0.74	0.01	1.1	1.3

N Groups		Calib. Logit	Model Error	Infit MnSq	Outfit MnSq
Count:	Mean:	-0.72	0.02	0.9	1.0
3	S.D.:	0.03	0.00	0.1	0.2

Table 7
Tabulation of Misfitting Ratings

A. Number of Misfitting Ratings Across Teachers (with the percent of all misfitting ratings)

Teacher #	N	%
123	8	4.10
28	7	3.59
301	6	3.08
178	5	2.56
64	4	2.05
81	4	2.05
148	4	2.05
170	4	2.05
175	4	2.05

9 Teachers had 3 misfitting ratings
26 Teachers had 2 misfitting ratings
70 Teachers had 1 misfitting rating

114 Teachers (28%, involved in misfitting ratings

B. Number Across Tasks with the percent of all misfitting ratings

Speaking Task	N	%
(I) Introduce Yourself	19	9.74
(I) Make Purchases	17	8.72
(I) Order a Meal	16	8.21
(I) Talk About Family Members	16	8.21
(I) Describe Your Daily Routine	13	6.67
(A) Explain a Familiar Simple Process	13	6.67
(I) Describe Typical Routines	11	5.64
(I) Give Directions	10	5.13
(A) Express Personal Apologies	9	4.62
(A) Give Instructions	8	4.10
(S) Describe a Complex Object in Detail	6	3.08
(S) Give a Professional Talk	6	3.08

3 Tasks involved in 5 misfitting ratings
0 Tasks involved in 4 misfitting ratings
3 Tasks involved in 3 misfitting ratings
8 Tasks involved in 2 misfitting ratings
6 Tasks involved in 1 misfitting rating

33 Tasks (87%) involved in misfitting ratings

C. Number Across Group Membership with the percent of all misfitting ratings compared with the % of total membership of the sample

Group	N	%	TOTAL MEMBERSHIP
Bilingual	159	81.5	57%
Spanish	31	15.9	29%
French	5	2.6	15%

Table 8
Speaking Tasks
That Were "Misfitting"
As presented to the teachers

Task No.	Task
1.	Introduce Yourself Be able to give your name and basic personal information such as would be given at a first meeting.
6.	Make Purchases Be able to request items, discuss prices, and handle currency in a situation involving a purchase.
9.	Talk About Family Members Be able to give the names of the members of your family and simple descriptive information, such as their occupations and physical characteristics.
14.	Order a Meal Be able to ask questions about menu items, order food, and ask for and settle a bill.

Table 9
Summary of Demographic Information
on the TOPT Standard Setting Studies

TOTAL NUMBER OF JUDGES

Spanish TOPT	30	(13 BE, 17 SP)
French TOPT	16	

POSITION:

	BE	SP	FR
Classroom Teacher	77%	18%	44%
Department Chair	0%	47%	19%
District Supervisor	8%	18%	6%
Teacher Trainer	15%	18%	31%

SEX:

	BE	SP	FR
Male	15%	24%	31%
Female	85%	76%	69%

ETHNICITY:

	BE	SP	FR
Hispanic	77%	53%	6%
White	23%	47%	81%
Black	0%	0%	13%

Table 10
Scaling of the Speaking Performances by the Standard Setting Judges

A. French Speakers

Speaker (Level)	Measure Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std
Spkr8 (Sup)	Maximum					
Spkr10 (Sup)	Maximum					
Spkr19 (Sup)	Maximum					
Spkr6 (Adv High)	Maximum					
Spkr13 (Adv High)	Maximum					
Spkr14 (Adv High)	Maximum					
Spkr1 (Adv)	Maximum					
Spkr3 (Adv)	Maximum					
Spkr7 (Adv)	Maximum					
Spkr18 (Adv)	3.18	0.84	1.6	1	2.3	1
Spkr11 (Int High)	1.07	0.70	0.6	-1	0.4	0
Spkr15 (Int Mid)	0.02	0.77	1.6	1	1.3	0
Spkr16 (Int High)	0.02	0.77	0.8	0	0.5	0
Spkr5 (Int High)	-0.64	0.85	0.5	0	0.3	0
Spkr2 (Int Mid)	-3.93	1.27	2.3	1	0.7	0
Spkr4 (Int Mid)	-3.93	1.27	0.2	-1	0.1	0
Spkr9 (Int Mid)	-5.58	1.32	0.9	0	0.2	0
Spkr12 (Int Mid)	-5.58	1.32	0.9	0	0.2	0
Spkr17 (Int Mid)	-5.58	1.32	0.9	0	0.2	0

Nu Speakers		Measure Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std
Count:	Mean:	-2.10	1.04	1.0	-0.0	0.6	0.2
19	S.D.:	3.04	0.26	0.6	1.0	0.7	0.5

B. Spanish Speakers

Speakers	Measure Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std
Spkr18 (Sup)	Maximum					
Spkr3 (Sup)	Maximum					
Spkr6 (Sup)	Maximum					
Spkr5 (Adv High)	Maximum					
Spkr21 (Adv)	4.22	1.05	1.0	0	0.4	0
Spkr9 (Sup)	3.43	0.77	0.8	0	0.3	0
Spkr13 (Sup)	3.43	0.77	1.3	0	2.9	1
Spkr15 (Sup)	3.43	0.77	1.2	0	0.8	0
Spkr20 (Sup)	2.92	0.66	0.9	0	0.5	0
Spkr19 (Adv High)	2.21	0.55	1.1	0	1.2	0
Spkr17 (Adv)	2.21	0.55	1.0	0	1.9	1
Spkr14 (Adv)	1.92	0.52	1.2	0	0.9	0
Spkr2 (Adv)	1.20	0.47	0.9	0	0.8	0
Spkr7 (Adv)	1.20	0.47	0.9	0	0.8	0
Spkr8 (Int High)	0.18	0.44	1.2	0	1.0	0
Spkr16 (Int High)	-0.60	0.45	1.0	0	0.9	0
Spkr10 (Int Mid)	-1.46	0.49	1.0	0	0.8	0
Spkr12 (Int Mid)	-1.70	0.51	1.0	0	0.7	0
Spkr11 (Int High)	-2.27	0.56	0.8	0	1.0	0
Spkr4 (Int Mid)	-3.02	0.68	1.0	0	0.7	0
Spkr22 (Int High)	-3.48	0.81	0.6	0	0.2	0
Spkr1 (Int High)	-4.41	1.08	1.3	0	1.1	0

Speakers		Measure Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std
Count:	Mean:	0.52	0.64	1.0	0.1	0.9	0.2
22	S.D.:	2.63	0.19	0.2	0.4	0.6	0.5

Table 11
Comparing the A Priori Classifications of the Speakers With the Actual Scaling "Best Case Scenario"

A. French Speakers

A Priori Expected Ordering	Actual Scaling				
	Int Mid	Int High	Adv	Adv High	Sup
Int Mid (6)	6	0	0	0	0
Int High (3)	0	3	0	0	0
Adv (4)	0	0	4	0	0
Adv High (3)	0	0	0	3	0
Sup (3)	0	0	0	0	3
Correct Order	6 (100%)	3 (100%)	4 (100%)	3 (100%)	3 (100%)

B. Spanish Speakers

A Priori Expected Ordering	Actual Scaling				
	Int Mid	Int High	Adv	Adv High	Sup
Int Mid (3)	1	2	0	0	0
Int High (5)	2	3	0	0	0
Adv (5)	0	0	4	1	0
Adv High (2)	0	0	1	0	1
Sup (7)	0	0	0	1	6
Correct Order	1 (33%)	3 (60%)	4 (80%)	0 (0%)	6 (86%)

APPENDICES

STRUCTURE OF THE TOPT - Spanish

Task	Item	Level	Speaking Task
	Warm-up	I	Answer personal questions
1	Picture 1	I	Give Directions
2	Picture 2	I	Describe a place/activities
3	Picture 3	A	Narrate in present time
4	Picture 4	A	Narrate in past time
5	Picture 5	A	Narrate in future time
6	Topic 1	A	Give instructions
7	Topic 2	A	State advantages/disadvantages
8	Topic 3	A	Give a brief factual summary
9	Topic 4	S	Support an opinion
10	Topic 5	S	Hypothesize on an impersonal topic
11	Situation 1	A	Speak with tact
12	Situation 2	S	Speak to persuade someone
13	Situation 3	S	Propose and defend a course of action
14	Situation 4	S	Give a professional talk
15	Situation 5	A	Give advice
	Wind down	I	

Key

A = Advanced
I = Intermediate
S = Superior

Texas Oral Proficiency Test (TOPT)
Bilingual Education Teachers

JOB-RELATEDNESS SURVEY

RETURN BY MAY 4, 1990

INTRODUCTION

The Texas Education Agency is developing a test of oral proficiency in Spanish which will be required of individuals seeking a certificate or an endorsement for bilingual education. The Texas Oral Proficiency Test in Spanish (TOPT-Spanish) will be a tape-mediated test. From a master tape and via a test booklet, examinees will be presented with approximately twenty speaking tasks. These tasks will allow them to demonstrate their ability to speak Spanish. Successful performance of these tasks requires various levels of Spanish speaking ability; some are fairly easy to perform, while others are considerably more challenging. The examinees' responses will be recorded on examinee response tapes. After examinees complete the test, their performance, as recorded on the tapes, will be scored by trained raters.

This survey presents you with 38 speaking tasks, such as may appear on the TOPT-Spanish. For each task, you are to indicate whether, in your professional opinion, bilingual education teachers need to have the **ABILITY** to carry out this task in order to perform successfully in bilingual education classrooms in the state of Texas. Note that the question is not whether bilingual education teachers need to carry out the task in the classroom, but whether bilingual education teachers need the **level of ability** necessary to carry out the task.

You are one of a sample of Texas bilingual education teachers selected to receive this survey. The results will assist the TEA in determining the level of speaking skills in Spanish needed by bilingual education teachers in Texas. Your responses are important and your assistance to the TEA is appreciated.

DIRECTIONS

Your survey packet contains: this survey booklet, a blue and white machine-readable survey response sheet, and a stamped, pre-addressed return envelope. Note that data for this survey are being collected with machine-readable response sheets. Please do not fold the survey response sheets.

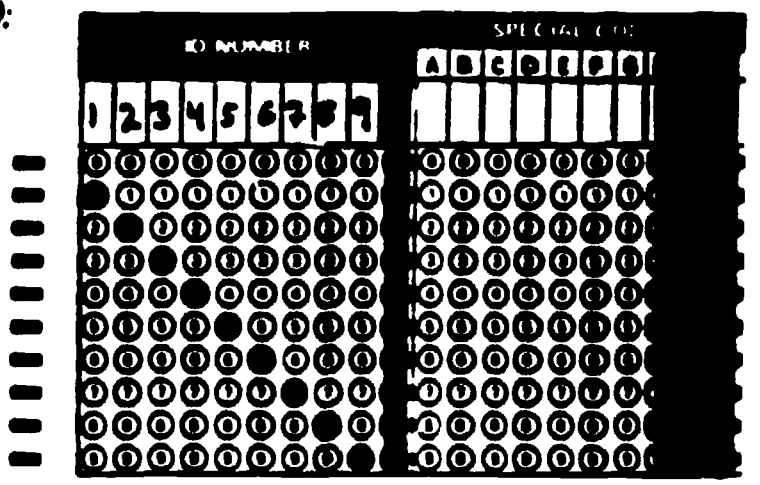
There are five steps to completing this survey. Follow all directions carefully and use a No. 2 pencil. It is estimated that this survey will require **15 to 20 minutes** to complete.

STEP 1 ID NUMBER

Please write your social security number in the boxes in the area entitled ID NUMBER on the top left-hand corner of the machine-readable survey response sheet. Then fill in the circle corresponding to the number in each box. NOTE: Your social security number will only be used for data processing purposes and will not be used to identify any individual respondent to this survey.

EXAMPLE

This is what your response sheet would look like if your social security number were 123-45-6789:

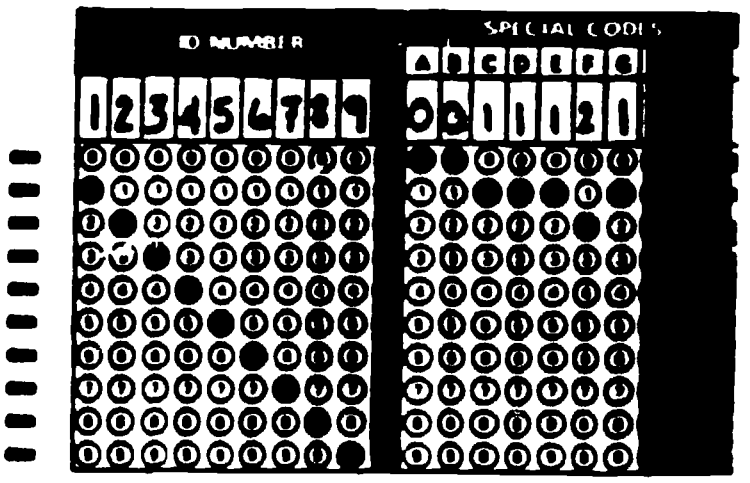


STEP 2 DEMOGRAPHIC INFORMATION

For demographic purposes, please answer each lettered question presented on the next page in the box labeled DEMOGRAPHIC INFORMATION. Write your answer in the area entitled SPECIAL CODES on the top left-hand corner of the response sheet. For each lettered question (A through G), write the number of your answer in the block on the answer sheet. Then fill in the circle corresponding to the number of your answer.

EXAMPLE

This is what your response sheet would look like if you were an elementary school teacher (Question A) with a certificate in bilingual education (Question B) and between 3 and 5 years of experience (Question C), etc.:



DEMOGRAPHIC INFORMATION

A. What is your current level of assignment?

(0) Elementary
(1) Junior High or Middle School
(2) High School
(3) Other

B. Do you hold a certificate or endorsement in bilingual education?

(0) Yes
(1) No

C. How many years of bilingual education teaching experience do you have?

(0) 1-2 years
(1) 3-5 years
(2) 6-10 years
(3) 11-15 years
(4) 16-19 years
(5) 20 or more years

D. What levels of bilingual classes have you taught during the past three years? (select only one)

(0) Early Childhood
(1) Grades 1-3
(2) Grades 4-6

E. What is the highest degree that you hold?

(0) No degree
(1) Bachelor's
(2) Master's
(3) Doctorate

F. What is your ethnic group?

(0) Hispanic
(1) Black
(2) White
(3) Other

G. What is your sex?

(0) Male
(1) Female

STEP 3 RESPONSES TO SPEAKING TASKS

Listed on the survey response sheet is a series of speaking tasks requiring various degrees of language ability to perform. For each task, indicate whether, in your professional opinion, bilingual education teachers need to have the language ability necessary to carry out the task in order to perform successfully in a bilingual classroom. In other words, for each task, ask yourself:

> Is the level of ability
> required to perform this task
> needed by bilingual education teachers
> in Texas public schools?

Important: The question is NOT "Do bilingual teachers need to carry out this task in the classroom?" Rather, the question is "Do bilingual education teachers need to have the Spanish language ability to carry out this task?"

Fill in the letter that represents your response to this question in the appropriate column on the response sheet. The columns are as follows:

A	=	Definitely Yes
B	=	Probably Yes
C	=	Maybe
D	=	Probably No
E	=	Definitely No

Following the examples below are detailed descriptions of the speaking tasks. Be sure to read them before making your response.

EXAMPLES

Here are two example tasks with responses completed for you.

Example A

Extend an Invitation

Be able to politely invite someone to your home for a party or other social function.

*If, in your opinion, bilingual education teachers should **definitely** have the **level of ability** required to perform this speaking task (independent of whether they would need to do the task in the classroom), then you would darken circle "A" in the first column of the response sheet.*

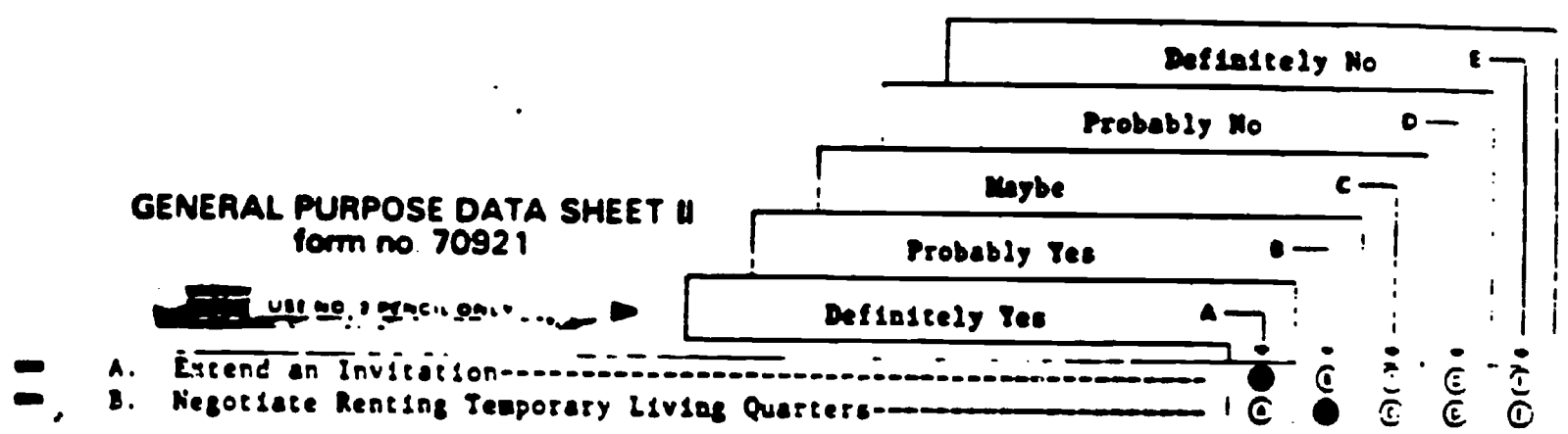
Example B

Negotiate Renting Temporary Living Quarters

Be able to negotiate a rental agreement with a landlord, ask questions about what is included in the rent, and ask for clarification of the rental agreement.

If, in your opinion, bilingual education teachers should probably have the level of ability required to perform this speaking task (independent of whether they would need to do the task in the classroom), then you would darken circle "B" in the second column of the response sheet.

If you made the above two responses to the example tasks, your survey response sheet would look like this:



Now please make your response for each of the 38 speaking tasks listed on the following pages on the appropriate line of the survey response sheet.
Remember to ask yourself, for each task:

> **Is the level of ability**
> **required to perform this task**
> **needed by bilingual education teachers**
> **in Texas public schools?**

SPEAKING TASKS

1. Introduce Yourself

   Be able to give your name and basic personal information such as would be given at a first meeting.

2. Explain a Familiar, Simple Process

   Be able to explain how to accomplish everyday processes such as writing a check, borrowing a book from the library, or taking attendance in the classroom.

3. Describe a Sequence of Events in the Past

   Be able to use and sequence language indicating past time in order to narrate an event or incident which occurred recently.

4. Propose and Defend a Course of Action with Persuasion

   In light of at least two possible choices of action, be able to propose and defend a course of action in such a way as to persuade others to accept your choice.

5. Describe Typical Routines

   Be able to use and sequence language indicating present or habitual time in order to narrate recurring events or routines, everyday activities, etc.

6. Make Purchases

   Be able to request items, discuss prices, and handle currency in a situation involving a purchase.

7. Talk About Personal Activities

   Be able to talk about your leisure activities, favorite pastimes, and preferred hobbies.

8. Hypothesize About an Impersonal Topic

   Be able to discuss various possibilities ("what if" situations) surrounding an abstract, impersonal topic.

[FOR SURVEY PURPOSES ONLY]

9. Talk About Family Members

Be able to give the names of the members of your family and simple descriptive information, such as their occupations and physical characteristics.

10. Give a Brief, Organized Factual Summary

Be able to summarize in an "oral report" fashion factual information about topics of a personal or professional nature.

11. State Your Personal Point of View on a Controversial Subject

Be able to state what you believe on a controversial subject and why you hold those beliefs.

12. Describe Expected Future Events

Be able to use and sequence language indicating future time in order to narrate expected occurrences of a personal nature, such as a planned trip or activity.

13. Explain a Complex Process in Detail

Be able to explain in detail a non-routine process of an impersonal nature, such as how to carry out a scientific investigation or how to write a term paper.

14. Order a Meal

Be able to ask questions about menu items, order food, and ask for and settle a bill.

15. Express Personal Apologies

Be able to apologize clearly and appropriately to an offended party.

16. Give Advice

Be able to give advice to someone faced with making a decision between two or more choices, giving supporting reasons for the advice given.

17. Hypothesize About a Personal Situation

Be able to say what you would do in a hypothetical situation.

[FOR SURVEY PURPOSES ONLY]

18. Describe Your Daily Routine

Be able to narrate your typical daily activities.

19. Give Instructions

Be able to give instructions and explain the steps involved in carrying out an activity.

20. Give a Brief Personal History

Be able to talk about your personal background.

21. State Advantages and Disadvantages

Be able to state the advantages and disadvantages of a situation (such as living in a big city), a decision (such as going to college), or an object that has affected society (such as the computer).

22. Support Opinions

Be able to state, support and defend a personally-held opinion or belief about an issue.

23. Describe Health Problems

Be able to describe health problems or conditions.

24. Discuss a Professional Topic

Be able to discuss at length and in detail a topic of professional interest.

25. Describe a Complex Object in Detail

Be able to describe a complex object such as a car or bicycle in detail and with precise vocabulary.

26 Lodge a Complaint

Be able to lodge a complaint, giving the reasons for and details behind the complaint.

[FOR SURVEY PURPOSES ONLY]

27. Talk About Your Future Plans

Be able to state and describe your personal or professional plans, goals and ambitions.

28. Give a Professional Talk

Be able to present a talk on a topic of professional interest.

29. Make Arrangements for Future Activities

Be able to inquire about and to make arrangements for future activities, and to set the date, time and place.

30. Evaluate Issues Surrounding a Conflict

Be able to present arguments on both sides of a familiar issue or topic and evaluate their relative merits.

31. Give Directions

Be able to give directions on how to get from one place to another.

32. Describe a Place

Be able to describe in detail a particular place, such as a school, a store, or a park.

33. Explain a Complex Process of a Personal Nature

Be able to describe and explain in detail a non-routine process such as how to get a job, or how to apply to college.

34. Hypothesize About Probable Outcomes

Be able to discuss what could happen if something unexpected occurs.

35. Correct an Unexpected Situation

Be able to handle an unexpected outcome, such as receiving faulty merchandise.

[FOR SURVEY PURPOSES ONLY]

36. Change Someone's Behavior through Persuasion

Be able to persuade someone to do something he or she is not inclined to do, or to cease doing something which is annoying to you.

37. Describe Habitual Actions in the Past

Be able to describe people, places or things in the past, such as the work schedule you used to have or leisure activities you used to do.

38. Compare and Contrast Two Objects or Places

Be able to compare and contrast two objects, places, or customs.

STEP 4 ADDITIONAL COMMENTS

Please use the space provided in the three WRITE-IN AREAS on the back of the survey response sheet for any additional comments you wish to make regarding the oral language functions to be included on the TOPT-Spanish.

STEP 5 RETURNING THE SURVEY

Unfold the enclosed pre-addressed, stamped envelope. Insert the blue and white machine-readable survey response sheet into the envelope, being careful not to fold it. Return the machine-readable survey response sheet only as soon as possible, but postmarked no later than MAY 4, 1990, to:

Mr. Dorry Kenyon
Center for Applied Linguistics
1118 22nd Street, NW
Washington, DC 20037

Thank you for your participation in this survey.

RETURN BY MAY 4, 1990