Summer Institute of Linguistics and
The University of Texas at Arlington
Publications in Linguistics

Publication 110

Windows on Bilingualism

Eugene H. Casad
Editor

A Publication of
The Summer Institute of Linguistics
and
The University of Texas at Arlington
1992

Library of Congress Catalog No: 92–81102

ISBN: 0–88312–809–8

ISSN: 1040–0850

Cover design by Hazel Shorey

Copies of this and other publications of the Summer Institute of Linguistics may be obtained from

International Academic Bookstore
7500 W. Camp Wisdom Road
Dallas, TX 75236

[blank]

Surveying Language Proficiency

John Stephen Quakenbush[1]

Prior to discussing methods of surveying language proficiency it is helpful to define two key terms: survey and proficiency. The first term, survey, is the more easily defined. Cooper characterizes survey research as "research carried out with respect to an entire population, whether as small as a hundred neighboring households . . . or as large as a nation . . . " (Cooper 1975:29). A survey of language proficiency fits into the larger category of sociolinguistic survey, which Cooper also defines succinctly as an endeavor which "gather(s) information about the social organization of language behavior and behavior toward language in specified populations" (29). Language behavior includes such phenomena as proficiency, acquisition, and usage. Behavior toward language includes both attitudinal and implementational behavior, the latter being observable, the former only inferable. Sociolinguistic surveys have largely been motivated by the need for information of language policy makers and program planners. They can also be justified on the basis of their contribution to more theoretical concerns involving the interaction of language and society.

The second key term here, proficiency, is the more difficult term to define. A working definition for the purposes of this study has been "the

[1]This paper was originally presented at the Fifth International Congress on Austronesian Linguistics, January 1988, Auckland, New Zealand, and is to appear in *VICAL2 Western Austronesian and Contact Languages*, Ray Harlow, editor. The volume itself is to be a *Te Reo* Special Publication of the Linguistic Society of New Zealand. I would like to express my sincere thanks to Ray Harlow, the editor of that volume, and to the Linguistic Society of New Zealand for their permission to reprint this paper here [editor's note].

degree to which a language can be used successfully in face to face interaction." Proficiency in this sense involves primarily the skills of listening and speaking. Although it is possible to consider degrees of proficiency in the other two major skill areas of reading and writing, there are many situations for which these literacy skills are not relevant.

How have language surveyors traditionally measured language proficiency? In the absence of any standard method, a variety of techniques have been employed. The most obvious distinction between survey techniques has been between those methods that gather reports of estimated proficiency versus those that actually administer some type of performance test. The report method is by far the most common, with reports usually gathered by means of a written questionnaire. A further distinction can be made in the report method between those that gather a respondent's estimate of the proficiency of others versus those that collect self-report data. Self-report data are the kind most commonly collected, but in some instances it is helpful to gain an individual's estimate of the second language proficiency of an overall community.[2]

Self-report techniques usually ask respondents to rate their proficiency according to predetermined levels, such as 'fluent, fairly well, a little' or 'very proficient, adequately proficient, hardly proficient, not proficient'. The number of levels distinguished, as well as the descriptions of these levels, varies from survey to survey.[3] One large scale survey employed a slightly different method by asking respondents if they could handle a particular language successfully in a series of thirteen situations, each new situation supposedly more difficult than the preceding one (Polomé and Hill 1980:116). This yielded an oral proficiency score for an individual of anywhere from 0 to 13.

The second major kind of technique employed by language surveyors to measure language proficiency entails some sort of direct testing. Some surveys in effect test only comprehension, or listening proficiency, as they ask the respondent for some sort of response to a verbal stimulus.[4] The most fully developed survey procedure for testing comprehension is the dialect intelligibility test, described in Casad 1974. This type of test consists of a short, tape recorded story, followed by a series of simple content

[2]For an example of a survey where informants were asked to estimate the language abilities of a surrounding community, see Ladefoged, Glick, and Criper (1968:53).

[3]For various scales of proficiency employed in self-report surveys, see Aguilana 1978, Bautista et al. 1977, Cooper and King 1976, Whiteley 1974, and Olonan 1978.

[4]See Serpell 1978 and Barcelona 1977 for elicited nonverbal responses. Casad 1974 and Kashoki 1978 describe techniques requiring oral and written responses in the mother tongue to aural texts in another language variety.

questions to be answered orally. The questions are sometimes interspersed in the body of the text, and are usually asked in the respondent's mother tongue. The resulting scores are taken as an indication of the extent to which dialect B is comprehensible to speakers of dialect A. This type of testing has proved very useful in situations where the key factor is inherent intelligibility due to linguistic similarity, as opposed to learned bilingualism gained through social contact.[5]

Other surveys of language proficiency have concentrated on productive capacity.[6] Most of these tests have depended on single word or otherwise minimal responses. Apparently no large-scale survey has ever tested actual conversational ability, presumably because of the many difficulties and indeterminacies involved. There is, however, a recognized method for measuring overall oral proficiency, developed by the United States Foreign Service Institute and related agencies.[7]

The oral proficiency interview developed by the FSI consists of a trained interviewer conducting a more or less natural, yet highly structured, conversation with a respondent in an attempt to discover that respondent's overall strengths and weaknesses in a given language. Factors explicitly evaluated include accent, comprehension, fluency, grammar and vocabulary. Different factors assume prominence at different levels of proficiency. Possible levels range from 0–5, with 0 being no knowledge of the language and 5 being educated, native-speaker proficiency. Levels 0–4 may be further refined by the addition of a '+' or half point. The FSI interview procedure is a complex one. The logistics of training interviewers and the time required for conducting the interviews, quite apart from the task of convincing the general public to submit to being evaluated, will probably preclude its use in any large scale survey. The main point of this paper, however, is that the FSI method can be adapted successfully for use on the community level, and can be further adapted as a useful tool for gathering self-report data from larger groups of respondents.

[5]See Grimes 1986a for more complete consideration of the differences between inherent intelligibility and learned bilingualism.

[6]Serpell 1978, De Gaay Fortman 1978, and Bautista et al. 1977 utilized visual stimuli to elicit linguistic responses.

[7]See Adams and Frith 1979 for a detailed explanation of this method. The oral proficiency interview has since been adapted by the Educational Testing Service (ETS) for use by the Peace Corps in evaluating the language proficiency of volunteers, and more recently by the American Council of Teachers of Foreign Languages (ACTFL) for use in an academic setting.

A sample study

The Agutaynen sociolinguistic survey was conducted in 1984–85 in Palawan, Philippines under the auspices of the Summer Institute of Linguistics.[8] One of its main purposes was to investigate the extent to which mother-tongue Agutaynen speakers could use the second languages of Cuyonon, Tagalog and English. Over 200 Agutaynens were interviewed in three municipalities of northern Palawan province. All interviews were conducted by the present researcher exclusively in the Agutaynen language. Responses were tape recorded on a small, hand-held audio recorder.

The section of the interview concentrating on language proficiency consisted of a set of seventeen yes-no questions (see appendix E). These questions involve specific language skills, each one associated with a particular level of proficiency, as defined by FSI. A level 1 question, for example, is "Can you understand and respond correctly to questions about where you are from, if you are married, your work, date and place of birth?" A level 5 question is "Do you know as many words in X as you do in Agutaynen?" This particular set of questions was derived from a longer set adapted from FSI materials by Barbara F. Grimes, and then further adapted to fit local circumstances. The original, longer set contained 37 questions. It was found during pilot testing, however, that this was far too many to maintain the interest of the interviewee. Many of the questions also seemed singularly unreliable in that respondents invariably answered them positively. For these reasons, the number of questions was reduced to 17. One of the criteria for selecting a question for the shorter version was that it be answered negatively at some time during the pilot testing. Another criterion was clarity. The shortened set worked smoothly.

In order for an interviewee to rate a certain level of proficiency, he or she had to answer positively all questions for that level. If a respondent could also answer positively two of the questions at the next level, a '+' was assigned.[9] During the actual interviews an attempt was made to maintain the atmosphere of a naturally occurring conversation. Many questions were asked from memory. Nonverbal or paralinguistic cues were also taken into consideration. For example, an elderly woman obviously uncomfortable in discussing her Tagalog proficiency was not asked more than the most basic questions for that language. On the other hand, statements by the respondent to the effect that his or her proficiency was very high in a given language precluded asking the most basic questions for that language. No attempt was made to administer the questions in

[8]See Quakenbush 1986.

[9]'No' was considered a 'positive' answer for Questions 4-D and 5-D.

exactly the same order for each respondent, although questions did generally progress from easier to more difficult. The end result of this flexibility in the order and number of questions was a more comfortable interview for the respondent, and it would be hoped, a rating of proficiency that was correspondingly more valid.

This self-report method for measuring language proficiency was employed for the Agutaynen survey as a whole. It was considered superior to previous self-report methods for two reasons. First, it asked about specific language skills, rather than for an abstract appraisal of global language ability. This allowed respondents to give simple yes-no answers while focusing on specific behaviors, rather than forcing them to give self-evaluations in terms that could be more directly linked to core values and self-esteem. Secondly, this particular set of questions was based on a scale of proficiency that is still being found useful after years of sustained, careful scrutiny by professional language testers and those they evaluate. Still, the Agutaynen survey was on new ground. This particular method had never been tried in a community-wide survey. Therefore, some sort of test of the method's validity was desirable. It was for this purpose that a separate test was carried out in another Agutaynen community in Brooke's Point, Palawan, subsequent to the main survey.

The Brooke's Point test consisted of assigning proficiency ratings for a sample of 40 individuals by two methods—the self-report method utilized in the larger Agutaynen survey, and an actual oral proficiency interview conducted in the field. Assuming that the two methods were language independent, only the Cuyonon language was used for the Brooke's Point test. Two language evaluators were trained in the technique of the oral interview, specifically for this purpose. One interviewer was a 49-year-old woman, the other a 35-year-old man. Both were native Cuyonon speakers, college-educated, and elementary school teachers by profession. The combination of both sexes and different ages was part of a deliberate effort to insure that a broad range of respondents would be comfortable in being interviewed by this team.

The general sequence followed for an interview was for the present researcher to first interview the respondent according to the self-report method, out of hearing of the other two language evaluators. Certain biographical information was also collected at this time. One of the language evaluators would then interview the same respondent, with the other evaluator an active observer. Afterwards, the evaluators individually assigned proficiency ratings without discussing the respondent's performance. All interviews were taped so that any serious differences in ratings could be discussed later. In the end, none of the evaluations differed more than one level. In these cases, an average score was taken as the final

rating. In instances where the evaluators disagreed by only a half point, the lower score was chosen.

For the 40 interviews conducted, 8 of the ratings varied by one point, 5 by a half point, and 27 were exactly the same (see appendix F). The evaluators' ratings came more in line with each other as time went on. Had the interviewers been more experienced at the start, perhaps their scores would have been in even closer agreement. That the evaluations of novice interviewers agreed as much as they did is strong evidence for the reliability of the oral proficiency interview.

When the direct test ratings are compared with the self-report ratings, the results are favorable, if not overwhelmingly so. The table in (1) illustrates the two kinds of scores compared. Only 4 scores were exactly the same for the two methods. An additional 23 were within a half point. Four more were one point apart. In all, 31 out of the 40 self-report scores could be considered reasonably accurate (one point or less difference).

The mean and standard deviation of the two sets of scores are quite similar. The positive correlation between the two sets of scores, however, is only moderate.[10] The table in (2) lists the mean, standard deviation and Pearson product-moment correlation for the two sets of scores.

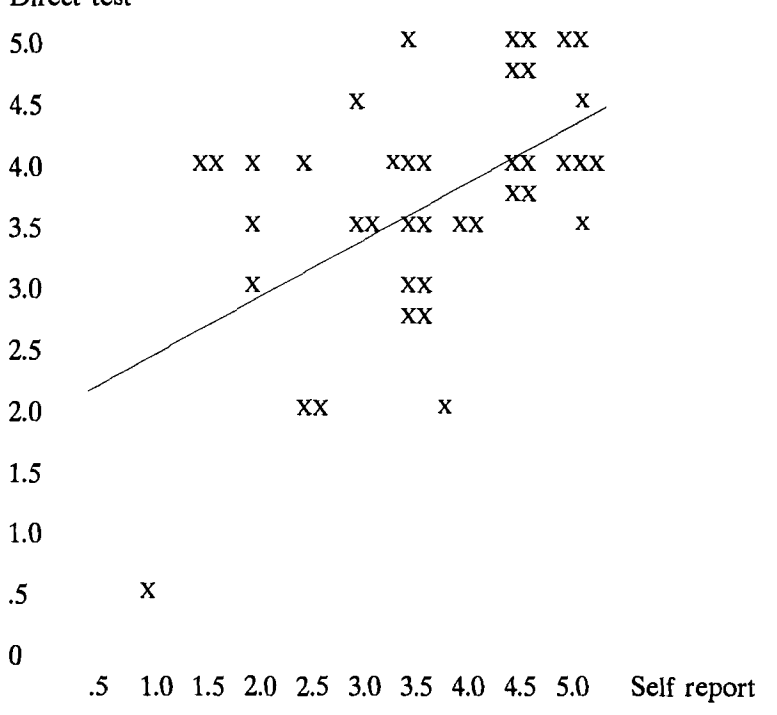
The moderate correlation may be spuriously low due to the restricted range of proficiency scores represented. A well-distributed sample of 40 proficiency scores would contain approximately 20 scores of 3.0 and above and 20 scores of 2.5 and below. As can be seen from (1), however, 31 of the 40 self-report scores are 3.0 and above, while 36 of the direct test scores are in the same category. Had the proficiency scores been distributed more evenly along the continuum, the positive correlation between the two methods of evaluation may have proved to be stronger.

The self-report method did not consistently yield higher or lower scores than the direct test method. To what extent, then, could the effects of the unreasonably low and unreasonably high self-report scores cancel each other out? The table in (3) shows that the differences are almost evenly split between scores that are too low and scores that are too high. This would minimize the importance of individual differences among a larger sample.

[10]Guilford (1956:145) gives the following interpretation system for measurement of correlation:

0.01–0.20 slight, almost negligible relationship
0.20–0.40 low correlation; definite but small relationship
0.40–0.70 moderate correlation; substantial relationship
0.70–0.90 high correlation; marked relationship
0.90–0.99 very high correlation; very dependable relationship

(1) Scatter diagram of Brooke's Point test scores. Regression line plotted.



(2) Descriptive statistics on Brooke's Point test scores.

	Self report	Direct test
Mean	3.60	3.74
Standard deviation	1.10	.95

Pearson product-moment correlation r = .56

A total of 9 out of 40 self-report ratings were off by 1.5 points or more. These 9 respondents represent a variety of ages and a range of educational and occupational backgrounds. There is, however, one striking similarity—7 of these 9 respondents were women. It would seem then, that Agutaynen women tend to understate their language proficiency under certain circumstances (6 of the 7 were understatements). The fact that they were being interviewed by an American male researcher may have been enough to produce this effect. At any rate, these women did not understate their actual language performance in conversation with the native Cuyonon speakers.

(3) Summary of differences between ratings obtained by self-report and direct-test methods.

Degree of difference	Self report lower	Self report higher	Total
0 points	—	—	4
.5	9	14	23
1.0	1	3	4
1.5	4	2	6
2.0	1	—	1
2.5	2	—	2
	17	19	40

Evaluation of case study

It is helpful to evaluate the Brooke's Point test from two perspectives. First, why did the direct-test method work as well as it did? Second, why did the self-report method not work any better than it did? We will then be ready to consider the implications of this test for future surveys concerned with the measurement of language proficiency.

The oral proficiency interview based on the FSI procedure worked well in the Brooke's Point test. This is somewhat surprising considering it was never intended for evaluating entire communities of sometimes marginally literate speakers. The FSI method was developed as a test for highly-educated individuals in the context of an intensive foreign language study program. Being interviewed is not optional for these individuals, but mandatory. Achieving a certain minimal rating is important to their careers. In Brooke's Point, in contrast, respondents had little obligation to submit to being interviewed. They were not involved in a formal language study program. Indeed, some had little experience in formal educational settings of any kind. Why, then, did the technique work? The answers to this question lie in the technique itself, and in the nature of the Cuyonon language evaluators and the Agutaynen community of Brooke's Point, Palawan.

The success of the oral proficiency interview can be attributed primarily to its natural and adaptable format. Although it may serve as a test for proficiency, on the surface it appears to be a natural communication event where information is conveyed between speakers in a socially meaningful way. It does not require a recitation of facts about language, a list of forms in a language, or answers to a series of multiple choice questions. The conversational format of the interview was such a strong factor that it apparently overshadowed any resemblance to a test situation. The actual

content of the interview may vary greatly from individual to individual, or from context to context. For example, whereas a foreign service officer may be asked to describe a political process, an Agutaynen farmer may be asked to describe a rice harvest.[11]

The success of the direct-test method in Brooke's Point also was due in great part to the personal characteristics of both the evaluators and the respondents. The Cuyonon language evaluators were willing to help, well educated, quick learners, good conversationalists, and friendly and unintimidating individuals. The Agutaynen respondents, on the other hand, were open to talking to outsiders—especially when one of those outsiders embodied the fascinating composite of an Agutaynen-speaking American. As a whole, they were also familiar with the idea of 'survey' and 'school project' (terms used to describe the present research) and were very willing to cooperate. Another factor which possibly contributed to the success of the Brooke's Point test was that the survey team had numerous personal acquaintances either in the Brooke's Point community or with relatives of community members. The surveyors' presence was further legitimized by two local guides who had been appointed by the chief political leader of the community to accompany the survey team.

With all of these positive aspects of the Brooke's Point situation, why did the self-report method not work any better than it did? Most likely, the main reason is in the very nature of self-report data. In reporting one's own abilities, concern for presentation of self may override concern for accuracy in either direction. That is, a respondent may overstate or understate an ability. In Brooke's Point, the majority of those who gave seriously misleading responses were understating their abilities, presumably in the interest of humility, but perhaps also in fear of being 'put to the test' and found wanting. It may also be the case that the correlation between self-report and direct-test methods would have been stronger had the sample not been skewed toward the higher ratings. In any case, 31 of the 40 self-report scores were accurate within one level, assuming the direct-test method yielded an 'accurate' standard for comparison.

Conclusions

What conclusions can be drawn from the above comparison of two methods for surveying language proficiency? First of all, it is evident that

[11]See Quakenbush (1986:277–287) for the materials used in training the Cuyonon language evaluators. Some minor adaptations in the procedure were made in the interest of cultural relevance.

assigning proficiency levels to individuals is not an exact science, no matter how these ratings are obtained. Proficiency data, therefore, and especially self-report data, must not be interpreted rigidly. Rather, they must be seen as indications of general trends. To the extent that a measure of proficiency is simply imprecise, it may reasonably be hoped that those scores which are slightly high will offset those scores which are slightly low, at least in part. Scores that are seriously off, however, will less likely cancel each other out. In the Agutaynen survey, for example, there was an apparent tendency for a proportion of women's self-report scores to be seriously underestimated. This leads to the second point, that self-report data on proficiency ideally will be interpreted in light of at least a subsample of direct testing measures.

The purpose and extent of a language survey will ultimately determine whether it is more beneficial to rely on a self report or direct measure of language proficiency. In the Brooke's Point test, the purpose was to assign proficiency ratings for 40 individuals in one second language. The self-report method took five minutes or less per respondent to gain the necessary information. The direct method generally took a manageable, but much longer, fifteen to twenty minutes per respondent. The overall Agutaynen survey, in contrast to the Brooke's Point test, examined proficiency for over 200 respondents in three languages. It would have been impractical, to say the least, to attempt direct testing as the sole method in such a survey. Depending on the purpose and extent of a survey, it may be advisable to sacrifice some precision in the interest of time, effort and expense. Nevertheless, future language surveyors who are concerned with the more precise measurement of language proficiency would be best advised to at least attempt the more time-consuming direct interview method when this is possible. Many circumstances can work against its successful utilization in the community, but as the Brooke's Point test demonstrates, it can also work surprisingly well.

Regardless of the particular instrument used in a language survey, the Agutaynen example demonstrates that the FSI levels of proficiency can provide a meaningful, standard framework for eliciting and interpreting degrees of oral language proficiency. The use of the FSI scale of proficiency should be encouraged in future language surveys, not only to ensure more comparable, and comprehensible, results, but also to test the usefulness of this scale in a broad range of speech communities.[12]

[12]Barbara F. Grimes, editor of Ethnologue, is compiling proficiency profiles on minority language communities using the FSI scale (see Grimes 1984b). Frank Blair (personal communication) has expressed reservations about the applicability of such a scale to nonliterate societies.

[blank]

Appendix E
Proficiency Questions Used in Agutaynen Survey

S-0+ Can you speak x just a little bit?

S-1 (A) Can you understand and respond correctly to questions about where you are from, if you are married, your work, date and place of birth?
(B) Could you explain the way from here to the high school to someone who did not know?

S-2 (A) Can you describe in detail your present or former work?
(B) Could you give a brief account of your lifestyle and plans for the future?
(C) Could you hire someone to work for you, arranging his wages, qualifications, hours, and responsibilities?

S-3 (A) Sometimes do you not know how to say something in x?
(B) Do you debate well in x?
(C) Can you listen to and give a brief summary of conversations in x on topics that you are interested in?

S-4 (A) If x-speakers are debating, are you always able to say to them whatever you want?
(B) Do you speak x well even when you're angry?
(C) Can you accomplish whatever task in x, just as if it were in Agutaynen?
(D) Do you make mistakes in x?

S-5 (A) Can you use as many words in x as in Agutaynen?
(B) Sometimes is it easier to think in x than in Agutaynen?
(C) Do you speak x as well as an x-speaker?
(D) Do people know that you are not an x-speaker by the way you speak x?

Appendix F
Proficiency Scores For Brooke's Point Test

Respondent	Self report	Direct test	Tester one	Tester two
1	4+	5	5	5
2	5	5	5	5
3	4+	5	5	5
4	2+	4	4	4
5	4+	5	5	5
6	3	3+	4	3+
7	5	5	5	5
8	1+	4	4	4
9	1+	4	4	4+
10	3	4+	4+	5
11	3	3+	4	3
12	4	3+	4	3
13	3+	3+	4	3
14	3+	5	5	5
15	4+	5	5	5
16	2	3+	4	3
17	3+	3	3	3
18	5	4+	5	4+
19	5	3+	4	3
20	2	4	4	4
21	3+	3+	4	3
22	4	3+	4	3
23	5	4	4	4
24	3+	4	4	4

Respondent	Self report	Direct test	Tester one	Tester two
25	3+	3	3+	3
26	3+	3	3	3
27	3+	3	3	3
28	2+	2	2	2
29	3+	4	4	4
30	4+	4	4	4
31	4+	4	4	4
32	1	0+	0+	0+
33	3+	4	4	4
34	4+	4	4	4
35	4+	4	4	4
36	5	4	4	4
37	5	4	4	4
38	2	3	3	3
39	2+	2	2	2
40	3+	2	2	2+