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

Calibrating Sentence Repetition Tests

Joseph E. Grimes

Sentence repetition tests are potentially useful for surveying the distribution of different levels of bilingual proficiency in areas where a second language is part of the sociolinguistic picture. They are patterned on tests that have been used monolingually to investigate speech development and speech disorders.

In order to validate the use of such tests, it is necessary to give them to many subjects and compare the results with those of fine-grained qualitative tests such as the U.S. Foreign Service Institute's well-tried interview test or its spinoff for nonliterate subjects, the Second Language Oral Proficiency Evaluation.

A procedure described by Radloff (1991) for setting up a sentence repetition test for validation corresponds to a type of implicational scaling: test probes are chosen to maximize the likelihood that the rank order of correct responses will correlate well with the rank order of proficiency test outcomes.

This paper discusses a computer program that implements a procedure equivalent to Radloff's that may expedite the calibration of possible test probes. It also calls attention to questions that remain concerning the suitability of sentence repetition tests for assessing all aspects of bilingualism.

Variation in bilingual proficiency

Minority languages go out of use when their speakers become fluent in another language, provided they come to use nothing but that second language in all walks of life.[1] The sociolinguistic study of minority languages, then, often requires the study of how a second (or third, or fourth) language [2] is used by different segments of a community.

Good second-language research goes beyond simply listing what other languages are used. Three requirements for such research are widely recognized:

1. Different individuals in a community learn their second language with different degrees of proficiency, depending mainly on their personal need to communicate in it and their opportunity to learn it. If bilingualism is widespread, it is the task of the survey to characterize in social terms who is and who is not proficient.
2. Different sectors of a society have different modal patterns of bilingual proficiency. The proficiency of men often differs from that of women; different geographical or subcultural regions may differ widely in proficiency; differences in age, education, and travel experience usually go with differences in proficiency. The effect of these social divisions, of their combinations, and possibly of other distinctions in some situations, is what has to be investigated.
3. In a field situation, it is invariably the more bilingual people that an investigator meets first. He or she may therefore have to go to great lengths to find a sample that reflects the variability of proficiency throughout the society as a whole. This usually implies a fairly large sample that is stratified according to the likely lines of division mentioned in point two.

The size of the sample (typically 25 to 100 subjects in each community, depending on how segmented the society is) sometimes becomes a barrier to teams that need to survey bilingualism, because the higher precision tests that are available (Bruhn 1989, Summer Institute of Linguistics 1987) take a lot of time. So it is natural and legitimate to look for less time-consuming ways to get an accurate reading on bilingual proficiency.

A test known as the sentence repetition test is used routinely in speech pathology and speech development studies. It is based on the observation that once a sentence reaches a certain level of complexity, you can't repeat

[1]Bilingualism as such does not lead inevitably to language extinction. The second language has to take over all speech domains. Many bilingual groups settle into a stable pattern of using one language for certain functions and the other for other functions.

[2]I employ "bilingual" here to refer to the use of any language other than the speaker's mother tongue; "multilingual" would be more appropriate for many societies.

it if you can't understand it. Language is such a tightly knit web that if one strand gets loose it takes others with it.

Many people can do a fair job of mimicking a short utterance in a language they don't understand—two or three words is common in the teaching of phonetics. In a language they know only imperfectly they can sometimes mimic six or seven words successfully. But by the time a sentence gets longer than that—and especially if it is structurally complex like 'A loose nut was what we were advised to watch out for'—they have to control the syntax and the lexical valences rather well or they trip all over themselves. A native speaker or a fluent second language speaker, on the other hand, has little difficulty repeating that kind of thing.

A sentence repetition test is simply a list of natural sentences recorded on tape. Test subjects are asked to repeat each sentence. The list is ordered from less difficult to more difficult. It is usually preceded by warmup or training sentences that nearly anyone who knows the language at all can handle, but that are not taken into account as part of the score.

Each subject hears each sentence only once.[3] His or her task is to repeat the sentence exactly as it was uttered. There is a little leeway in the scoring that is explained later, but basically the subject either gets it right or misses it; scoring does not need to make fine discriminations that require expert judgments in order to quantify degrees of error. This makes it relatively easy to train people who do not know linguistics as testers.

The advantage of such a test is that it can be administered to a large number of people over a wide area in a relatively short time: twenty minutes or less per subject.

The disadvantages are twofold. As long as the surveyor is not tempted to draw conclusions that the test is incapable of giving by its very nature, neither disadvantage need stand in the way of using it. The disadvantages are:

1. Before the sentence repetition test for a particular second language[4] can be interpreted meaningfully, it must be calibrated against an independently validated test of bilingual proficiency. This calibration is a major effort in itself; to be convincing it involves administering

[3]Earphones can keep bystanders who are also potential test subjects from having the opportunity to practice ahead of time.

[4]It is an open question whether a test that is effective for native speakers of one language who also speak the test language is automatically effective for native speakers of another language who also speak the same test language—for example, whether a French test given to Arabic speakers who also speak French gives comparable results when administered to Italian speakers who also know French. The question merits investigation in its own right.

fine-grained interview tests to a large number of subjects representing all proficiency levels, preferably in each of several communities.

2. The content of the test is such that it does not probe all the areas of language that have proven critical in assessing higher levels of second language proficiency. It probes mainly sentence structure, some comprehension, and phonology. It does not probe command of the lexicon very extensively (the ability to choose words or phrases appropriate to the interaction), nor does it give much information about fluency or the ability to manage connected discourse or overall comprehension. It therefore permits only an a fortiori argument about proficiency: If a subject can't handle sentence structure and phonology and basic comprehension very well, or doesn't know even the 150 or so words used in the test, it stands to reason that he or she should have trouble with complex comprehension, the lexicon, discourse, and fluency as well, therefore is not very proficient. But the opposite argument cannot be made: A perfect score on a sentence repetition test still tells nothing about lexical command or discourse, and gives only a sentence-size reading on fluency and comprehension.[5]

Nevertheless, if a sentence repetition test is properly validated, and if it is used as a screening test to sort out subjects on the lower end of the proficiency scale without attempting to use it for discriminations at high levels of proficiency, it might form the backbone of the bilingualism phase of a number of surveys.

Calibrating a sentence repetition test

Calibrating a sentence repetition test begins by finding a pool of fifty or so bilingual people—preferably in each of several culturally important communities—whose proficiency in their second language has already been

[5]Thea Bruhn (personal communication) reports that lexicon, discourse, fluency, comprehension, and structural command are evaluated independently of each other by the Foreign Service Institute because they correlate poorly enough with each other that none of them has turned out to be a statistically impressive predictor of any of the others.

assessed by an external test such as SLOPE (SIL 1987).[6] Complete calibration requires a pool of people whose abilities range from practically nothing to practically perfect bilingualism.

The sentences used for testing can be drawn from any unconstrained text—newspapers, conversations, everyday life, stories. They should be long enough that mere mimicry breaks down. Only a few should be as short as five words; most should be nine words or longer. They should be sentences that can stand on their own two feet, clear without context. They should not represent nonstandard or low class speech; on the other hand, neither should they be literary or fancy. The more grammatical complications they include naturally, the better.

Arrange fifty or so sentences of this kind from short to long, from apparently simple to apparently complex, for the initial recording. Have several native speakers of the language make high quality test tapes of the whole list, preferably in a recording studio. If they stumble when they try to utter some of the sentences, edit those sentences out—if native speakers can't manage them, what will the others do?

Choose the best recording as a calibration tape. Run it past the pool of fifty or so bilinguals whose proficiency levels you already know. Have them try to repeat every one of the fifty or so calibration sentences with no replays. Score 3 for each sentence they get absolutely perfect, 2 if they make one mistake, 1 if they make two, and 0 if they make any more—Radloff (1991) tells how to set up testing sheets and what counts as a mistake.

Fifteen or twenty of the test sentences are to be chosen out of the initial list to form the actual field test. The ones that are retained show three characteristics:

---

[6]Where bilingualism is supported through schooling in a major language, the classroom-oriented interview tests of the Educational Testing Service and the American Council of Teachers of Foreign Languages may be useful. They are less precise than SLOPE when it comes to discriminating the higher proficiency levels. Another method for guessing intelligently at proficiency levels without really observing speech behavior has also been used for calibration, though it is more properly taken as another screening method in its own right. In the Reported Proficiency Evaluation described by Radloff (1991), a number of mother tongue speakers of the language in question are interviewed about their impressions of people they know who speak that language as a second language—for example, native speakers of standard Spanish might be interviewed about how well their Quechua-speaking acquaintances handle Spanish. The interviewer guides them through a set of criteria similar to the ones used in SLOPE, but there is no way to verify the interviewees' impressions against specific speech samples. The level of precision that can be attained is therefore more suitable for general screening than for calibrating another test.

1. They can be ordered in such a way that any subject who repeats a later sentence correctly is almost certain to repeat all the sentences before it correctly as well. In the other direction, however, repeating an earlier sentence is no guarantee that any of the later sentences can be repeated. This kind of relationship among sentences is generally known as an implicational scale (Hatch and Farhady 1982) or a "Guttman scale" after Louis Guttman, who first developed the idea.
2. A subject's ranking on the implicational scale gives a clear indication of his or her ranking on the independent scale established via the use of a higher precision test of bilingual proficiency.
3. The sentences chosen give the clearest achievable discrimination across all levels of proficiency.[7] At this point a computer program called SENREP[8] simplifies accumulating the calibration test results, ranking the sentences by difficulty and the subjects by proficiency, and sorting out those sentences that fit the model of an implicational scale from those that don't. In addition, the same program can be used for administering the test widely (as opposed to calibrating it, which is merely the startup phase) by matching the response patterns of test subjects with those of the original calibration subjects.

The computer program goes through several steps:[9]

1. You give it an identification of the language being tested and the survey that the test is part of.
2. You give it the text and the order of the fifty or so candidate sentences used to establish the calibration, from which the actual working set is drawn.

---

[7]Because of the inherent inability of sentence repetition tests to probe discourse competence, lexical choice, range of comprehension, or fluency, my own recommendation would be to lump together all levels of proficiency from 3 to 5 on the Foreign Service Institute scale into a single range of "3 and above" in order to reduce the temptation to read high level discriminations out of a low level test. The reason for that cutoff level is that above level 2+ it is precisely the factors that the sentence repetition test cannot probe that are most important in making discriminations.

[8]Machine readable copies of SENREP compiled for PC-compatible computers are available for the cost of reproduction and mailing from the publisher of this volume. There is no guarantee that they will work. The function of a manual is fulfilled by a help file included with the program that can be called up at any point by pressing function key 1.

[9]You can stop at any point and store the data you have entered up to then in a file. Later on you can retrieve that file and continue from where you left off. A typical calibration for one community involves a minimum of fifty subjects and fifty candidate sentences, which is too much to type into a computer accurately at one sitting.

3. You give it the calibration test results for each subject: a coded identification of the subject (to safeguard anonymity), the proficiency level established for that subject by the independent test (FSI or SLOPE), and the tester's rating of that subject's response to each of the fifty or so candidate sentences in 2.
4. It gives you its analysis of each candidate sentence in the form of a chart. The chart is stored in a separate computer file that can be examined and printed by importing it into an ordinary spreadsheet program.[10] The columns in the chart correspond to the calibration subjects. Each column is headed with the subject's identification code, his or her proficiency level as established independently, and the cumulative score over all the test sentences. The subjects are arranged from those who got the fewest answers correct to those who got the most. The rows in the chart correspond to the candidate sentences themselves. They are ordered from easiest to hardest: a sentence that everybody repeated correctly would have a difficulty index of 0 associated with it, and one that nobody could handle would have a difficulty index of 1. In addition, each sentence carries a deviation index (Radloff's "discrimination index") that tells how closely that sentence fits an ideal implicational scale, with 0 deviation for a sentence that scales perfectly relative to the rest.
5. It gives you its summary of the scalability of the whole chart. There is an overall index of reproducibility or scalability, which in the calibration phase is likely to be unacceptably low because none of the candidate sentences that don't scale well have been eliminated yet. Two other indexes sum up the overall difficulty and deviation.

The table in (1) illustrates the calibration process in miniature. It consists of only seven candidate sentences instead of fifty, taken from an Urdu example of Radloff's (with hypothetical Urdu data). The sentence repetition ratings are given for only twelve calibration subjects instead of fifty, in order to keep everything on a single page.[11]

[10]Commonly used spreadsheet programs include Lotus 1-2-3 (TM), Quattro Pro (TM), Microsoft Excel (TM), or AsEasyAs (TM). The spreadsheet controls the printer and lets you look at the parts of the chart you want to see. You can remove sections of the chart that are superfluous. Other uses of a spreadsheet program in linguistics include putting discourse information into a Thurman chart (Grimes 1975) or into Longacre's (1989) band structure, or manipulating charts of tone patterns, affixes, or anything else that involves rows and columns. There are also the standard project management uses of spreadsheets: doing accounts, budgets, tax calculations, mailing lists, statistics, itineraries, bibliographies, progress records, and graphs of all the above.

[11]"Probe" is used instead of "calibration sentence" in keeping with general testing terminology, and to save space on the chart.

(1) Sentence Repetition Test calibration data

Calibrate probes by sum of subject scores

Subject→			E	A	I	B	F	J	C	G	K	H	D	L	
Proficiency			0+	0+	0+	1	1+	1+	2	2+	2+	3	3	3	
Sum of scores			1	1	2	3	6	7	9	13	14	16	17	20	
Probe	Diff	Dev													
1 (P6)	0.31	0	1	1	1	2	2	2	2	2	3	3	3	3	*talim hasıl kʌrke . . .*
2 (P7)	0.56	2	0	0	1	0	1	1	2	2	2	2	2	3	*kʌl tʃʌlte wʌqt . . .*
3 (P3)	0.56	2	0	0	0	0	1	1	2	2	2	3	2	3	*ɪs dʒʌdid dɔr . . .*
4 (P4)	0.58	4	0	0	0	0	0	1	0	3	2	3	3	3	*mæn ʌb mʌzid ta . . .*
5 (P5)	0.64	0	0	0	0	1	1	1	1	1	1	2	2	3	*ʊn ki talimat ke . . .*
6 (P1)	0.64	4	0	0	0	0	0	0	1	2	3	1	3	3	*tʊm ne hath pʌr . . .*
7 (P2)	0.69	0	0	0	0	0	1	1	1	1	1	2	2	2	*kjunka talim hʌm . . .*

Reproducibility (scalability)	0.95, should be >=	0.9
Minimal marginal reproducibility	0.62, should be <	0.9
Proportional reduction in error	0.87, should be >	0.6

The subjects are identified by capital letters. Each of them has been assigned a proficiency level by an independent test of known accuracy. The scores in each column (3 for a perfect repetition, 2 or 1 for minor mishaps, and 0 for any more serious flaw) are added up, and this sum is the basis for the left-to-right order in which the subjects appear.[12] The full text of each calibration sentence is given.

In (1) the sentences are no longer in the order in which they were presented to the calibration subjects (P1, P2, . . . , Pn). That tentative order was based on their length and an informal guess about their relative complexity. In (1), however, they are ordered by their calculated difficulty index and renumbered accordingly, carrying along the original presentation ordering (P) in order to make it easy to find them on the audio tape.

[12]The program also calibrates the probe deviations by the ranking of proficiency judgments. Superficially this appears less neat than the ranking of sentence repetition scores, but it is probably more realistic because the proficiency scores come from a nominal scale established by a set of definitions, not an ordinal scale based on higher or lower scores. Furthermore, the proficiency scale is nonlinear: the relative amount of effort required to get from level 0 to 0+ is only a small fraction of the amount of effort required to get from 3 to 3+, for example. The time projections for attaining proficiency given by Brewster and Brewster (1976:377) are very much like a Fibonacci series (0, 1, and the sum of the two preceding numbers), which is nonlinear. The sentence repetition scoring, on the other hand, is linear. In the long run I will probably recommend using the ranking of proficiency judgments for calibration.

From a printout of the calibration chart you can choose which candidate sentences give you the most information for the least work. Choose fifteen to twenty sentences out of the original fifty by these three rules:

1. Ignore sentences with a difficulty index less than .05 or over .95—the reasons why everybody might get everything right, or nothing right, are extremely complex but tell us almost nothing.
2. Choose a spread of difficulties that spans the spectrum evenly.
3. Prefer low deviation indexes.[13]

Bad sentences are the ones with high deviation indexes: they do not fit the implicational scale, therefore you cannot make accurate judgments from them. You don't want to choose sentences with difficulty indexes that are too close together to discriminate; you want your difficulties spread out evenly from .05 to .95, not bunched up. You may have to flip a coin to decide which of two equivalent sentences to use.

Once you decide which sentences form the best test, you invoke the computer again to record your choices. It shows you each calibration sentence from the easiest to the hardest, giving you its original number, text, difficulty index, and deviation index. Then you register whatever you decided: to use that sentence or drop it from the list. (The computer never really erases a sentence; you can get back to the original set at any time by going back into calibration mode.)

After you have registered your judgments about everything on the list, the computer recalculates everything using only the sentences you have decided to keep for the actual test. It gives you a new sum of scores for each subject, taking into account only what you found to be the most consistently useful sentences. It reorders the subjects by their new scores. It recalculates difficulty and deviation indexes and the three summary indexes and puts everything out in a new chart file. At this point you save the new working data base so that you will see only the sentences you want to use next time you retrieve the data.

The table in (2) is the recalculation for the mini-test given in (1). It is limited to three sentences instead of the usual fifteen to twenty. The sentences selected were the three with the least deviation from ideal scalability. Choosing those sentences and ignoring the rest improves the reproducibility scores noticeably. The test sentences in (2), however, cover the difficulty range only from .31 to .69, and cover it unevenly. This would

[13]Radloff (1991) reports that in a full scale calibration effort there are hardly ever any deviation indexes as low as zero. In any full scale testing, we have to settle for minimum distortion, and may never achieve zero distortion.

not be an adequate test in practice, but it illustrates how the real one is put together.

(2) Sentence Repetition Test calibration results

Calibrate probes by sum of subject scores

Subject→			I	E	A	B	J	G	F	C	K	H	D	L	
Proficiency			0+	0+	0+	1	1+	2+	1+	2	2+	3	3	3	
Sum of Scores			1	1	1	3	4	4	4	4	5	7	7	8	
Probe	Diff	Dev													
1 (P6)	0.31	0	1	1	1	2	2	2	2	2	3	3	3	3	*talim hasıl kʌrke...*
2 (P5)	0.64	0	0	0	0	1	1	1	1	1	1	2	2	3	*un ki talimat...*
3 (P2)	0.69	0	0	0	0	0	1	1	1	1	1	2	2	2	*kjunka talim hʌm...*

Reproducibility (scalability)	1.00, should be	>=	0.9
Minimal marginal reproducibility	0.68, should be	<	0.9
Proportional reduction in error	1.00, should be	>	0.6

Rating test subjects against calibration subjects

Radloff takes a straightforward approach to evaluating the performance of test subjects once the calibration is finished. She uses a linear regression model: that is, she observes that the scores of the calibration subjects plot out fairly close to a straight line when their sentence repetition scores are put across the bottom and their independently established bilingual proficiency scores down the side. A formula that generates that straight line can be calculated from the scores, together with a correlation coefficient that expresses how close on the average they come to the hypothetical line. From the regression formula it is possible to match any test score with a proposed bilingual proficiency score.

There are, however, two reasons to think that working from a regression formula could be an oversimplification. The first is that the test sentences usually do not fit the ideal implicational scale perfectly; their average deviation index for a full scale test is in practice greater than zero. This means that the error that is due to lack of scalability compounds with the error expressed by a correlation coefficient less than one, which indicates the extent to which the calibration data fail to fit the regression formula.

If the compound error were to turn out unacceptably large, something other than the regression equation would be desirable.[14]

The second reason for caution about simple regression has to do with the nonlinearity of the bilingual proficiency scale mentioned in footnote twelve. A difference in sentence repetition test ratings at the low end should correspond to a greater difference in bilingual proficiency ratings than a difference of the same magnitude at the high end, because of the way the proficiency scale is defined. A lot of work on test designs will have to be done before this statement can be either affirmed or denied: the point for now is that it is desirable to have an alternative to the regression formula as a way to interpret the results.

The SENREP program gives another way of matching test subjects (whose bilingual proficiency you are trying to estimate) with calibration subjects (whose bilingual proficiency you established by giving them a fine-grained, independent test). It compares each new subject with all the calibration subjects in order to be able to assign them a proficiency level indirectly.

In test mode you identify each test subject to the computer. The computer then takes you through the probes that have been selected—no longer the full list—in their order of difficulty and has you put in their scores as you did for the calibration.

The new chart file, when you bring it up in your spreadsheet program, begins with a copy of the current calibration chart for you to refer to. Then it shows you a comparison of each test subject with up to ten calibration subjects.

It chooses the subjects to compare by looking for sums of scores that are as close as possible to the test subject's sum. Then it picks the five calibration subjects lower than that point and the five above it to make the comparison (fewer, of course, at the ends of the list of calibration subjects).

The subjects who are listed first are the ones most like the test subject. If they are all at a single rank, say 2+, then you can safely guess that your test subject may be at 2+ too. If the closest calibration subjects are at several levels, or if they really aren't very close, then you might be more tentative about classifying that test subject.

The measure of closeness has two parts. The first looks at the test subject's answers to individual questions one by one, and compares them with the answers each calibration subject gave to the same questions. It

[14]There are other association measures that might be equally appropriate. The cosine of the angle in k-dimensional space between the vectors that represent the *k* test and calibration answers is widely used in information retrieval for a similar function, and is easily calculated.

adds up the differences and reports them as "d"; d=0 indicates identical responses.

The other part measures differences in the shapes of the response patterns. It is the

From (3) it can be seen that there might be considerable leeway in deducing proficiency levels from sentence repetition responses—though a full set of fifteen or more probes would undoubtedly give less helter-skelter responses than the ones in the example. One would, for example, feel confident about assigning subject M to level 0+ (or possibly level 1); but subject N might reasonably match anywhere from level 1 to level 2+, though the match is more likely with 1+ or 2.

(3) Sentence Repetition Test matches on calibration subjects

Test subject M is closest to calibrator I, level 0+
M's scores are 2 0 0 Total 2

d=1	r= 1.000:	Subject I,	total 1, assigned level 0+	1	0	0
d=1	r= 1.000:	Subject E,	total 1, assigned level 0+	1	0	0
d=1	r= 1.000:	Subject A,	total 1, assigned level 0+	1	0	0
d=1	r= 0.866:	Subject B,	total 3, assigned level 1	2	1	0
d=2	r= 1.000:	Subject J,	total 4, assigned level 1+	2	1	1
d=2	r= 1.000:	Subject G,	total 4, assigned level 2+	2	1	1
d=2	r= 1.000:	Subject F,	total 4, assigned level 1+	2	1	1
d=2	r= 1.000:	Subject C,	total 4, assigned level 2	2	1	1

Test subject N is closest to calibrator J, level 1+
N's scores are 2 1 1 Total 4

d=0	r= 1.000:	Subject J,	total 4, assigned level 1+	2	1	1
d=0	r= 1.000:	Subject G,	total 4, assigned level 2+	2	1	1
d=0	r= 1.000	Subject F,	total 4, assigned level 1+	2	1	1
d=0	r= 1.000:	Subject C,	total 4, assigned level 2	2	1	1
d=1	r= 1.000:	Subject K,	total 5, assigned level 2+	3	1	1
d=1	r= 0.866:	Subject B,	total 3, assigned level 1	2	1	0
d=3	r= 1.000:	Subject I,	total 1, assigned level 0+	1	0	0
d=3	r= 1.000:	Subject E,	total 1, assigned level 0+	1	0	0
d=3	r= 1.000:	Subject A,	total 1, assigned level 0+	1	0	0

Test subject O is closest to calibrator K, level 2+
O's scores are 3 1 2 Total 6

d=1	r= 0.866:	Subject K,	total 5,	assigned level 2+	3	1	1
d=1	r= 0.866:	Subject H,	total 7,	assigned level 3	3	2	2
d=1	r= 0.866:	Subject D,	total 7,	assigned level 3	3	2	2
d=2	r= 0.866:	Subject J,	total 4,	assigned level 1+	2	1	1
d=2	r= 0.866:	Subject G,	total 4,	assigned level 2+	2	1	1
d=2	r= 0.866:	Subject F,	total 4,	assigned level 1+	2	1	1
d=2	r= 0.866:	Subject C,	total 4,	assigned level 2	2	1	1
d=2	r= 0.000:	Subject L,	total 8,	assigned level 3	3	3	2

Closing

The text of the SENREP program, as well as a disk containing it, which can be obtained from the publisher of this volume, is written in PDC Prolog (formerly Borland Turbo Prolog), a computer language that makes heavy use of recursively defined functions. As a consequence it carries no inherent limitations on the number of probes, calibration subjects, or test subjects that it can handle, nor on the length of specific pieces of data. It is, however, limited by the memory capacity of the computer it runs on. A slower version with much greater capacity could be developed by putting the larger data base components on a hard disk.