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]

A Review of Walker's Research

Dale Savage[1]

For more than two decades there has been an awareness within SIL of the crucial role that language attitudes may play in the implementation of vernacular translation and literacy programs. Casad (1974) indicates an SIL interest in language attitudes related to bilingualism as early as the mid to late 1960s in Mexico. A recognition of the vital role of language attitudes in minority language planning spread widely in SIL during the 1970s and is reflected in the 1980s in three important conferences[2] which focussed significant attention on the issue of language attitudes in assessing the

[1]I wish to thank Bob Bodwell, a professionally trained statistician, with whom I discussed a number of the issues presented here before the actual writing was begun. I also wish to express appreciation to Eugene Casad, the editor, for his patient and helpful suggestions, and special credit is due my wife, Lorraine, who read drafts for coherence and who heard me rehearsing the central ideas more often than I care to think about.

[2]The first of these, the Sociolinguistic Survey Conference at Stanford University, November 27–28, 1982, (Huttar 1982) was organized and hosted by Charles Ferguson, Shirley Brice Heath, and John Rickford of Stanford, and brought together several participants from professional academia and SIL. Several participants discussed the difficulties of language attitude assessment during the relatively brief encounters associated with a sociolinguistic survey.

A second conference, the Stanford Conference on Vernacular Literacy, July 24–25, 1987, (Shell 1988) with a major attitudinal component was also hosted by Shirley Brice Heath and Charles Ferguson. Participants in this conference raised the issues of language attitude assessment and the implications of language attitudes for literacy program planning.

The third conference, the International Language Assessment Conference at Horsleys Green, England, May 24–31, 1989, (SIL 1989) included a large section of papers dealing with language attitude assessment.

prospects for vernacular literacy and translation programs among minority language groups.

It is within this context that Roland Walker has considered the problems of assessing language attitudes and vernacular literacy acceptance. A contributor to all three of the conferences mentioned above (Walker 1982, 1988, 1991), many of Walker's ideas about assessing attitudes and predicting the acceptance of vernacular literacy grew out of his field experience in Irian Jaya, Indonesia. His graduate studies at the University of California, Los Angeles, culminated in his 1987 dissertation *Towards a Model for Predicting the Acceptance of Vernacular Literacy by Minority-Language Groups.*

A general theme running through the works cited has been that the assessment of language attitudes is important for making good decisions for language programs, but that the direct assessment of attitudes is, for numerous reasons, too difficult to accomplish in a field survey setting. Walker's approach, then, has been to assess other factors and assume that language attitudes could be inferred from them.

Walker's Horsleys Green paper (Walker 1991) indicates that his research and thinking have had a significant impact on language survey design and interpretation in SIL Indonesia's Irian Jaya Program Committee. Through his publications, Walker's ideas about assessing language attitudes and vernacular literacy acceptance have the potential for gaining considerable currency within SIL beyond Irian Jaya. This is partly due to dual drives within SIL to assess its remaining task of providing vernacular translations and literacy programs as quickly as possible, and also to a desire in some quarters to limit the remaining task to a manageable size. Walker's approach, therefore, is attractive to some because it promises speed, and limits the task somewhat by identifying those language communities which have a low probability of accepting vernacular literacy.

Given its potential for wide influence, it is advisable to subject Walker's approach to careful scrutiny to ascertain if it is adequate in its present form to play a major role in decisions regarding initiating particular language programs. Several questions need to be asked: (1) Are the claims made about language attitudes and other underlying assumptions valid? (2) Is the sample drawn in a manner designed to render valid generalizations about potential literacy programs worldwide as it purports? (3) What is the nature of the data generated by the questionnaires? Is it reliable? (4) If the study meets the normal standards of reliability and validity, how good are the results as a decision-making tool?

Complexities of assessing language attitudes

Perhaps the critical questions to be addressed should be: (1) Does the model presented by Walker provide an adequate basis for assessing attitudes? (2) Are the claims made by Walker warranted that the results of this study have confirmed the approach of assessing language attitudes by evaluating the forces that shape them?

The importance of these questions becomes readily apparent when we consider the great shift that has occurred in thinking within SIL about how language attitudes (and other sociolinguistic factors) should be incorporated into program planning. John Bendor-Samuel (1982), participating in the 1982 Stanford Sociolinguistic Survey Conference, clearly enunciated the then generally accepted position that if comprehension (whether intelligibility or bilingualism) were low then work should be done. If, on the other hand, comprehension were high, then language attitudes should be assessed to determine whether a project should be undertaken anyway because of the negative attitudes toward the second language or its speakers.

By contrast, in recent years there has been a strong push toward the opposite position that even when there is clearly inadequate comprehension of a second language, putative negative attitudes toward mother-tongue literacy or other negative social factors may deter us from projects where translation might have been desirable 'in principle'. See David Bendor-Samuel (1991) for an exposition of this view.[3] Thus, in the current climate, questions of language attitude become crucial when we consider that reports of negative social pressures (including attitudes) may result in curtailed programs even for those minority language groups with clear comprehension needs. It is in this setting that we must address Walker's claim that this approach presents us with an effective means of assessing attitudes.

If we look at Walker's presuppositions and plan of research, I believe the answer to our second question above becomes evident. First, Walker assumes that attitudes toward language are shaped by sociolinguistic forces from within and without the community, and that these sociolinguistic variables [forces] can be observed and measured. On the other hand, Walker assumes that the difficulties of trying to assess attitudes, in their own right, in the field are so great that, "it would seem best . . . to set aside the study of language attitudes, and rather, to evaluate the sociolinguistic forces that shape language attitudes which are observable and measurable." (Walker 1987:49–50).

Walker then turns his attention to devising a questionnaire based on some of the factors implicated in the literature on language shift and death

[3]But note also Early (1991) and Hollenbach (1989) for two alternative viewpoints.

(Walker 1987:51). Rather than assessing attitudes, he has actually placed their assessment outside the scope of his study, focussing instead on establishing linkage between some of the factors related to language shift and the acceptance of vernacular literacy.[4] This is a perfectly legitimate research goal and strategy, and a very worthwhile investigation. However, Walker never returns, neither in his dissertation nor in any subsequent research, to validating the linkage between these factors and the assumed relation to language attitudes—the relationship of these factors to language attitudes remains merely an assumption just as it was at the beginning.[5]

Walker's insistence that this approach is a means of gathering attitudinal data is unfortunate for two reasons. First it draws attention away from the genuine contribution of this line of thought. There really does appear to be a relationship between the general factors Walker is considering and literacy behaviors in minority language groups. Its value is in no way enhanced by claiming it assesses language attitudes. The second consequence of claiming this is an effective (and adequate) means of assessing language attitudes is that such a claim may potentially undermine efforts to carefully assess the types of attitudinal data necessary to supplement this type of research for literacy program implementation.

There is no denying that trying to assess attitudes in emerging communities is full of difficulties. It is not, however, impossible, and there are numerous indications that static approaches which focus solely on observable social phenomena do not adequately account for the observed response set of social behaviors and the attitudes they imply.

In a long term research program which bears a great deal of topical similarity to Walker's research, Howard Giles and a number of associates articulated a theory of language in intergroup relations and ethnolinguistic vitality (Giles, Bourhis, and Taylor 1977). From the inception of the model, vitality was assessed by considering three classes of objective factors: status factors, demographic factors, and institutional support and control factors. The early operational procedures involved consulting "demographic, economic, sociological, and historical documents to arrive at as 'objective'

[4]Like Walker, I assume that there is a connection between language attitudes and the factors involved in language maintenance and shift. The operational aspects of his model, however, assume an isomorphic relationship between these factors and attitudes to which I do not subscribe.

[5]If Walker wishes to confirm this approach as a method of assessing attitudes, then he should actually measure the attitudes with a well validated, standard measure of attitudes and then do regression analysis of that result against the observable social factors. Lambert et al. (1960) provides a good description of the type of process involved in validating a new methodology (the matched guise technique) for attitude assessment. See also Huff (1954:74) for what he calls the semi-attached figure.

an assessment of a group's vitality as possible" (Bourhis, Giles, and Rosenthal 1981:146–47).

While the objectively assessed vitality appeared to provide "a useful tool for comparing ethnolinguistic groups in cross-cultural research," as early as 1979, objective factors alone were not considered sufficient to "account for group member's intergroup attitudes, skills and motivations for second language learning, attitudes toward language usage and use of code switching strategies" (Bourhis, Giles, and Rosenthal 1981:147). Instead in 1981, a parallel track of research was initiated to investigate ethnolinguistic group member's "subjective perceptions" to the same set of vitality factors in order to supplement the objective data.

> A combination of objective and subjective data may be extremely valuable in assessing the likelihood that ethnic minorities will survive as distinctive cultural and/or political entities in majority cultures. 'Subjective' vitality data may provide advance indication that a particular minority group is to mobilize in an ethnic revival phase not otherwise forseeable solely on the basis of 'objective' vitality information. (Bourhis, Giles, and Rosenthal 1981:147)

There are many other instances in the sociolinguistic literature which indicate that something beyond the observable social milieu is necessary to account especially for the behavior of subordinated or socially disfavored linguistic groups. As Ryan (1979) notes, once the legitimization of a dominant language,

> has resulted in universal recognition of the standard, one might expect the other varieties to disappear over a generation or two. However, many regional, ethnic, and social class varieties . . . have tended to persist for centuries, surviving strong pressures to succumb in favor of the standard dialects. (Ryan 1979:145)

This notion of persistence in the face of dominant languages (and their speakers) implies that even though there may be "universal recognition" of the legitimacy of the standard, in many cases there are also other forces

including attitudes at work which conspire to preserve group unity and identity in the face of pressures for assimilation.[6]

Susan Gal (1989) and especially Kathryn Woolard (1985; 1989) have pointed out, that in the case of linguistically dominated groups, quite different complexes of attitudes may lie beneath the surface of outwardly similar 'objective' circumstances. Woolard who did extended research on the politics of language and ethnicity (including language attitudes) during the period in which Catalonia achieved autonomy argues,

> We cannot read hegemonysaturation of consciousness directly from the institutional domination of a language variety. Just as nonstandard practices may accompany standard consciousness, so it is logically possible that standard linguistic practices may accompany or conceal resistant consciousness, as a form of accommodation to coercion rather than the complicity essential to the notion of cultural hegemony. The distinction is important, because accommodative behavior may be more easily dislodged and does not present the same problem for social change as does collaborative consciousness. (Woolard 1985:741)

Woolard, an anthropologist, gathered the data for her analysis through a variety of means including participant observation and interviews, but referring to the measurement of attitudes and the analysis outlined above,

[6]Unfortunately there are relatively few empirical or ethnographic studies of persistence. As Fishman (1990:5–10) reports in an important new thrust on "reversing language shift," as a result of "several societal and social biases," there has been definite skewing of research in the direction of shift rather than persistence. Sociolinguists (and other social scientists) have generally focussed most of their attention on processes of change within language groups. Accordingly we have a very refined taxonomy,

> with respect to the 'minus' side of the ledger (we speak of language attrition—shift—endangerment—loss—death and can itemise many studies of each way-station along this increasingly negative progression), while the 'plus' side remains rather gross and undifferentiated and studies of revival, restoration, revitalization and restabilisation remain proportionately few and far between. (Fishman 1990:6)

Fishman attributes the lack of attention to persistence to "our modern fascination with the dynamics of change *per se*," and points out that, "the forces and processes of change coexist, *in a single process*, with the forces and processes of persistence, and what most social scientists mistakenly call 'change' is really the by-product of the *interaction* of persistence and change" [italics in the original] (Fishman 1990:11).

If Fishman's call for researchers and "change-agents on behalf of persistence" is successful in attracting attention to the study of reversing language shift, we may begin to see more data and analysis into how and why linguistic groups successfully resist structural coercion and linguistic dominance.

writes, "this finding comes not from data on language use, but from what are called 'subjective reaction' tests. This is a form of empirical evidence on the social evaluation of language use, as important as evidence on language use itself" (Woolard 1985:741).

It is, in part, the social and psychological complexity of attitude structures that leads to the low correspondence "between attitudes and actual behavior" that is discussed in Agheyisi and Fishman (1970). We may agree with the notion that attitudes are "agendas to action," but there appear to be important situational constraints that mediate between various observable stimuli and attitude/behavioral responses.[7]

Thus, as Walker himself notes, a respondent may tell you he has one attitude with respect to an object of affect and then perform an act which runs counter to his verbal report. This does not mean that the verbal report was necessarily deceptive or inaccurate. Rather it is a function of the fact that attitudes are associated with a wide range of values, beliefs, and intentions with respect to objects of affect (including other languages and their speakers), and different social situational contexts necessitate differential normative patterns of behavioral response.[8]

This is in fact, the situation Labov (1972:292–96) describes in which nonstandard speakers endorsed the norms of the dominant group in the test situation, but did not wish to adopt those norms. In coming to grips with this apparent anomaly, Labov posited the existence of "covert norms" in support of the vernacular. In his own words,

> Why don't all people speak in the way they obviously believe they should? . . . Careful consideration of this difficult problem has led

[7]The whole notion of the seeming inexact match between attitudes and behaviors is one of the classic discussions in the attitude literature. Two major review articles of language attitude research, Agheyisi and Fishman (1970) and Giles et al. (1987), deal with the topic as either "intervening" or "mediating" variables between language attitudes and behavior. Brudner and White (1979) present research which highlights the language attitude/behavior problem with respect to Irish Gaelic. Ehrlich (1969), Fishbein and Ajzen (1975), and Wicker (1969) present discussions and theories about attitude/behavior problems from the perspective of mainline attitude research. An entirely new and promising approach to the problem of attitudes and behavior is being developed by several scholars using Catastrophe Theory, the recently articulated qualitative mathematical theory of René Thom (Anderson 1985, Ball, Giles, and Hewstone 1984, Flay 1978, Tesser 1980). This latter may eventually serve to make the notion of intervening variables obsolete.

[8]Given that this discussion is about social norms and attitudes, individual dispositions and deviance, though relevant to the study of attitudes and normative behaviors, will not be addressed here. Tesser (1980) provides a short, but interesting, entrée into the conflict between individual dispositions and social norms that may be useful to those involved in intensive participant observation.

> us to posit the existence of an opposing set of covert norms, which attribute positive values to the vernacular... We have therefore some empirical support in positing the opposition between two sets of values as the normative correlate of stable sociolinguistic markers... We agree with Homans (1955) that *the proper object of study should not be behavior alone, or norms alone.* (Labov 1972:295–96) (emphasis added)

Trudgill (1984) notes a very similar situation in Norwich, England where a definite language change is underway in a nonstandard direction related to the concept of covert prestige. In Norwich, Trudgill found that there were expressed values about language that were consonant with conferring prestige based on standard language norms. But he also uncovered data which demonstrate that for certain sex and class combinations, "nonstandard speech is in a very real sense highly valued and prestigious," and the working class dialect is gaining speakers from the middle class. That is to say that in the face of a socioeconomic structure which confers prestige on more standard speech varieties, a language shift is taking place in which a stigmatized variety of speech is gaining speakers at the expense of the dominant variety.[9]

Trudgill's Norwich studies are highly instructive as we consider assessment. If we were to take a rather simplified look at the existing social, economic, and educational structures, we would find considerable negative pressure exerted on the non-standard working class urban dialect. Regarding attitudes as well, at the level of conscious awareness, there were overt expressions of dissatisfaction by subjects with their own speech and stated desires to "speak properly." If the assessment of social factors and attitudes stopped at that, we might be tempted to assume that the working class dialect would be doomed to disappear under the weight of outside social pressure harnessed with negative attitudes toward the vernacular. Trudgill, however, did find that there were "deeper motivations for their actual linguistic behavior than these overtly expressed notions of their own 'bad speech'" (Trudgill 1984:57). So, in this case at least, observable sociolinguistic forces and a cursory examination of language attitudes

[9]In a review of a large number of language attitude studies, Ryan (1979:152) found a general trend that, "both evaluative reaction and questionnaire studies have revealed that nonstandard speech varieties may have low prestige but are associated with other values of importance for an ethnic group."

would not be sufficient for us to predict the type of covert prestige-driven language shift that is occurring.[10]

In numerous studies focussed specifically on language attitudes, important differences in attitudes within a group have been discovered based on locally relevant social categories such as age, gender, social class, occupational groupings, etc. (Ryan 1979). Walker's approach does not take into account the fine-grained attitudinal differences which have been demonstrated to exist in various parts of the world.

The general drift of virtually all the attitude research literature in the mentalist tradition (to which Walker subscribes) indicates that an attempt to assess language attitudes without more internally-focussed discovery procedures (such as interviews or subjective evaluation methods) and careful, in-depth observation is not likely to yield the quality of attitudinal data which can be most useful in the planning and implementation of literacy programs.

At this point, we answer the first of our questions: a model which gathers only the more easily assessed "objective sociolinguistic factors" is too impoverished in itself to adequately account for the complex attitude structures which interact within dynamic social systems. A research model of this type does not assess the target population's subjective perceptions vis-à-vis the objective factors, nor is it equipped to probe for resistant consciousness to linguistic domination. The model can reveal neither the complexities of attitude structure and conflicting norms in various social contexts, nor the potentially meaningful differences in attitudes accruing to socially relevant subcategories within a linguistic group.

It is imperative, therefore, that we realize that the assessment of observable social phenomena (including patterns of language use) and the assessment of language attitudes are not fungible; rather they are complementary. When we have one without the other, we are unable to properly interpret the significance of either.

The criterion variables

Turning from questions about whether or not Walkers' research regimen assesses attitudes, we now look into the actual content of the research question itself. Walker first built indices of vernacular literacy acceptance by asking fieldworkers to assess community literacy acceptance in their

[10]This accords well with Ryan's (1979:154) observations that both "direct and indirect measures of language attitude appear to be critical," and "direct questions may not reflect the whole picture."

locales[11] according to four criteria: (1) the sale of vernacular literature, (2) reading ability, (3) the amount of informal reading, and (4) the usage of vernacular Scriptures in churches. Pearson correlations were performed to correlate the four criterion variables with 19 predictor variables drawn largely from the language shift literature.[12] Finally, multiple regression analysis was performed on those criterion variables which garnered enough significant simple correlations to entertain this analytical technique.

In implementing this regimen, only the first criterion variable, the percentage of the population purchasing vernacular literature and the fourth criterion variable, usage of vernacular Scriptures in churches[13] exhibited enough significant simple correlations to perform multiple regression analysis.

Within the test group then, a relationship is indicated between the predictor variables and the sale of vernacular publications (criterion variable 1) on the one hand, and between the predictors and the frequency of public reading of vernacular Scriptures in church (criterion variable 4) on the other. The more interesting criteria, reading ability (criterion variable 2) and informal usage of vernacular literature (criterion variable 3), which represent individual literate behaviors, washed out of the model.

There are legitimate questions as to whether the two remaining criterion variables in themselves provide a very useful way of characterizing vernacular

[11]Walker's study was limited to the single communities which the respondents knew best, not the entire language groups. His stated purpose in confining the study to single communities was to restrict the variability in the data because in a 1986 study "using the entire language group as the unit of analysis" tended "to average out the variation that could be explained by the predictor variables" (1987:71). While this is a legitimate restriction for the purpose of the study, the very variability he has chosen to restrict suggests that within many of the communities where literacy has not been accepted (as measured by Walker's criterion variables), alternate allocation strategies or program approaches might have produced more desirable effects assuming, of course, that the general thrust of Walker's model is valid.

[12]The actual questions used to assess the criterion and predictor variables from Walker's questionnaire are included in appendix G. A copy of the entire original questionnaire may be found in Walker (1987:238–45); a shortened, revised version appears in Walker (1988:41–45). Question numbers in this paper correspond to the numbering of the predictor variables found in appendix G.

[13]This criterion variable actually consisted of a weighted average for each church based on the percentage of the community attending each church (Walker 1987:78). In discussing this variable with a number of fieldworkers, many thought that effect of even a single small church using vernacular Scriptures could have a disproportionate influence in the further penetration of vernacular literacy into the community. If their views are correct, then an unspecified number of communities may be underscored on this variable. This point could, of course, be verified or disproven in the course of further field research.

literacy acceptance. That is, the sale of vernacular literature does not necessarily tell us much about the use of that literature; indeed, if we array literate behaviors along a continuum, simple sales would fall on the low end of the scale. On the other hand, the public reading of vernacular Scripture in church probably indicates a great deal of community acceptance of vernacular literature; it may, or may not, however, be accompanied by individual literate behaviors in the vernacular. So we are left with a multiple regression model based on two criteria separated by a broad behavioral gulf, and presenting little information about individual literate behaviors.

Since it is doubtful that these two criterion variables alone give an accurate index of vernacular literacy acceptance, more work needs to be done on the model either to develop additional indicators to reflect the mid-range literate behaviors which are lost through the failure of criterion variables two and three (reading ability and amount of informal reading), or perhaps other predictors could be found which would be effective with these criteria.

The sample

Randomness. One of the key elements in evaluating research which purports to draw inferences from a statistical base is an examination of the sample and how it was drawn. Statistical methods are built on probability theory and depend for validity upon each event[14] having an equal opportunity

[14]An event in this situation is a literacy program meeting the criteria for selection. The population from which the sample should be drawn consists of all the literacy programs worldwide which would meet the criteria for inclusion. In fact, however, all the places where literacy programs are in place only represent a sample of the possible allocation sites where literacy programs could have been or will be initiated.

If fieldworkers have been using similar (stated or unstated) criteria in choosing their allocations, then those allocations may be systematically unrepresentative of the remaining potential allocations in unforeseen ways. For example, if a field entity had a policy of initiating work in all its rural language groups prior to allocating teams among urban-based language groups, then a survey carried out in that entity before all the rural allocations were filled would be totally biased toward rural allocation. This is a patently concocted example, but, however subtle, the possibility exists that through common training, ideology, romanticism, etc., we may have exercised bias in choosing earlier allocations.

The previous paragraph discusses possible bias as reflected in choice of allocation sites between language groups. There is also "within group" bias that may be represented in the sample, i.e., bias guiding the choice of one community as an allocation site over others in language groups characterized by multiple communities. So we can see that the communities represented in the study comprise a sample of possible allocations for those language groups represented in the study sample.

to be selected in the sample. This notion is commonly referred to as randomness. In a simple random sample, every potential member of the study has the same probability of being selected, and the selection is independent, i.e., the choice of one item will not affect the choice of another item. There are a number of ways to randomize the sample, and most books on statistics for the social sciences contain discussions of the significance of simple random samples and how to draw them.

In Walker's study, the population about which he wishes to generalize is all potential minority literacy allocations. It is obvious, however, that only a subset of the population can qualify for inclusion in the sampling frame, i.e., those allocations which have been filled and which have had a qualifying minimum of literacy work conducted. This is a reasonable limitation in the study even though there may be some unidentified biases represented in these programs (see footnote 13). Given this limitation, the most reliable sample which could be drawn would be a simple random sample of all (SIL) minority literacy programs worldwide.[15]

As we examine the composition of Walker's sample (1987:xi–xii; 72–74) in (1), however, we see that the 54 cases in the sample are drawn from just eight countries: Mexico, Guatemala, Brazil, Cameroon, Ivory Coast, Philippines, Indonesia, and Papua New Guinea.

(1) Sample composition (Walker 1987)

Country	Number of cases in study	Percent of cases in study
Mexico	14	26
Guatemala	4	7
Brazil	13	24
Cameroon	4	7
Côte d'Ivoire	1	2
Philippines	5	9
Indonesia	3	6
Papua New Guinea	10	19
Total	54	100

At no point does Walker inform us how the countries that were included in the study were selected; the only information given in the dissertation is that letters were sent to the SIL literacy coordinators in the countries

[15]The reason we would want to draw the sample from all projects worldwide is that the generalization Walker wishes to make is to all potential literacy programs worldwide.

represented soliciting their cooperation (Walker 1987:74). Given that no description of a randomization process is presented, it does not seem likely that the countries were randomly selected from a larger pool. This means that the sampling frame is restricted to just those countries shown in the table. Whatever criteria (convenience of the researcher, willingness of branch administrations to participate, etc.) were used to select the countries included in the study, the exclusion from potential samples of programs not located in these countries introduces serious bias into the study.[16] In fact, "a sample can only be representative of the population included in the frame" Fowler (1984:19). If these countries were not chosen at random, then the practical effect of drawing the sample just from them is that any generalizations issuing from the data can only be valid for programs from those eight countries.

A further problem emerges as we examine Walker's description of the respondent selection process (1987:73–74). We can see clearly that within the countries included in the sampling frame, there is no attempt to choose a random sample. Instead, the questionnaires were distributed to as many people as possible who met the minimum criteria of two vernacular Scripture publications distributed and the ability to answer the questionnaire. The sample consisted of those who were simply willing to spend the "one to three hours to complete the questionnaire" and then return it. This is commonly referred to by terms such as an availability sample, a convenience sample, haphazard sample, etc. The problems inherent in availability samples such as this one are discussed in Fowler (1984:20),

[16]Perhaps the best known case of problems arising from a poorly selected sampling frame is the infamous *Literary Digest* poll of 1936. In that instance, *Literary Digest* sent out ten million questionnaires to prospective voters asking their preference in the upcoming presidential election. With a response rate of about 2.4 million, *Literary Digest* predicted a landslide victory for Alf Landon (Republican) over Franklin Delano Roosevelt (Democrat) by a margin of 57% to 43%. When the actual vote was counted, however, Roosevelt had won by a margin of 62% for Roosevelt to 38% for Landon.

How could the results from such a large sample be so wrong? Most of the error has been attributed to a biased sampling frame. The sample was drawn from sources such as automobile registration lists, club membership rolls, telephone directories, and magazine subscription lists. While this might be a reasonable sampling frame today, in 1936 the country was polarized politically along economic lines. Republicans tended to be much wealthier than the more numerous Democrats, and the sampling frame (which was biased toward those with larger disposable incomes) was loaded with Republicans far beyond their proportion of the actual voting population (McClave and Benson 1985:918).

This example underlines a profound, but obvious, point. Large samples and statistically significant results are meaningless if the sample does not represent the relevant population.

Bernard (1988:97–98), and numerous statistics texts for the social sciences. One of the more succinct statements, however, can be found in de Vaus:

> Availability samples . . . are the least likely of any technique to produce representative samples . . . Using this approach anyone who will respond will do . . . This type of sample can be useful for pilot testing questionnaires or exploratory research to obtain the range of views and develop typologies, but *must not be used to make any claim to representing anything but the sample itself.* (de Vaus 1986:69) (emphasis added)

At least one other potentially serious source of bias remains. Walker, as far as I can ascertain, does not discuss how many people were sent the questionnaire but didn't return it. He implies, however, that some did not "take [the] one to three hours to complete the questionnaire." It is unfortunate that we do not know the rate of nonresponse. If the nonresponse rate is significant at all, then it is likely that there is nonresponse related bias reflected in the data.

Fowler (1984) reports with respect to mail surveys,

> that people who have a particular interest in the subject matter . . . are more likely to return mail questionnaires . . . This means that mail surveys with low response rates almost invariably will be biased significantly in ways that are related directly to the purposes of the research. (Fowler 1984:49)

Additionally, Fowler notes that although we cannot know much about the bias of nonresponders, "it is seldom a good assumption that nonresponse is unbiased" (Fowler 1984:52).

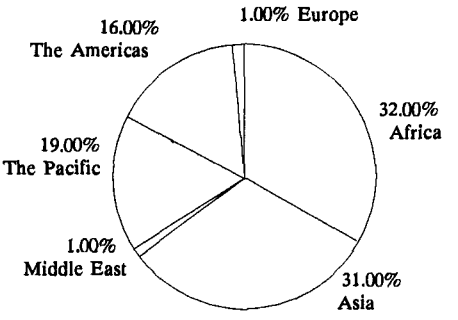
Representativeness. If the sample is not purely random, it should at least be representative of the population about which a generalization is to be made.[17] For a representative sample, the researcher attempts to predetermine (often through pilot studies or literature reviews) which natural categories may be relevant to the findings of the study, and randomly selects a portion of the sample from each of the categories.

As a very rough first estimate of how representative Walker's sample may or may not be, we can compare the geographical distribution of the world's living languages with the geographical distribution of the sample. Looking first to the figure in (2), we see the world's languages are partially

[17]In many cases it is advantageous to use a representative (or stratified) random sample rather than a simple random sample. See Fowler (1984:24–26), de Vaus (1986:57–59) or, especially, Babbie (1975:156–57) for discussions of the mechanics and advantages of stratified random samples.

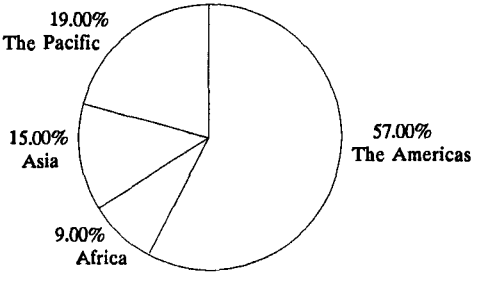
distributed in the following manner: Africa (32%), Asia (31%), the Pacific (19%), and the Americas (16%).

(2) Geographical distribution of living languages (source: Grimes (1984a:xvi))



When we compare (2) with (3), we see that the American continent with only 16% of the world's living languages is highly overrepresented, comprising 57% of Walker's sample. The Pacific is statistically represented just right if we only look at the percentages. If, however, we look beyond the numbers, we see that all the Pacific cases are from Papua New Guinea. It is an open question just how representative these language groups are of those found in other parts of the Pacific such as Polynesia or Micronesia. Looking to Africa we see an even more striking contrast. Africa which contains 32% of the world's living languages is represented by just 9% of Walker's sample, and that translates to only five languages from two countries in West Africa. East Africa is not represented nor is North Africa.

(3) Composition of Walker's sample (geographically distributed)



There are other factors besides geography—affiliation with which major religion (Islam, Hinduism, Buddhism, Christianity.), for instance, that we could use to examine the sample to see how representative it is, but it is unlikely that this sample could be construed as representative.

What this sample means for us is that we really have no basis for drawing statistical inferences from the dissertation study. More precisely, the results cannot be used to make generalizations about anything other than the programs it was drawn from. The statements of statistical significance are meaningless. It is not, however, a total loss; as a pilot test or exploratory research, a good deal can be learned from the data Walker gathered and analyzed. For instance, having to rely solely on criterion variables one and four as an index of vernacular literacy acceptance is not very satisfying (and perhaps only marginally valid). What has already been done can serve as a springboard for developing replacement variables for criterion variables two and three.

Statistical tests and the level of measurement

Given that the study may be repeated at a later date with a valid sample, it should be pointed out that some of the correlations are artificially high. This is due to the use of Pearson product moment correlations with dichotomous and ordinal variables (Walker 1987:94). In short, the Pearson's *r* makes strict assumptions about the data which ordinal and dichotomous data do not meet.

The way in which indicators are defined operationally in research also defines the level of measurement we attain in our data. This is important because the higher the level of measurement we attain, the more powerful statistics we may employ. Nominal variables are those in which the data are categorized into exclusive and exhaustive lists such as group membership, race, nationality, etc.[18] Dichotomous variables are a special subset of nominal variables which only have two categories, yes/no, gender, vernacular language/other language, etc.

Ordinal variables also produce data in categories that are exclusive and exhaustive, but the data are ranked as well. While the data are ranked, the distance between the ranks either has no meaning or cannot be ascertained. Scales such as low/medium/high, bilingual proficiency ratings, etc. produce ordinal data.

[18]The fairly ubiquitous nominal category, "other," is often used in social science research to fulfill the requirement that the categories be exhaustive. So, for example, a religion variable might have the categories: Christian, Jewish, Moslem, other.

Interval variables have all the properties of nominal and ordinal variables plus the characteristic that the intervals between values are equal.[19] For example, the ten degree interval from 60° to 70° is the same as the interval from 70° to 80°. Examples of interval scales are Fahrenheit and Celsius temperatures as noted, as well as population percentages, age, and weight.

Knowing the level of measurement allows us to choose the appropriate correlation coefficient for any two variables.[20] The figure in (4) illustrates the appropriate coefficients in tabular form. The different measures of association may themselves be ranked from those which make the least assumptions about the level of measurement of the variables to those which make the most assumptions about the variables.

(4) Appropriate coefficients based on data type (adapted from Fitz-Gibbon and Morris (1978:91))

		Variable 1		
		Dichotomous	Ordinal	Interval
Variable 2	Dichotomous	Phi coefficient ϕ		
	Ordinal	rank biserial r_{rb}	Spearman's rank order r_S	
	Interval	point biserial r_{pb}	Spearman's rank order r_S	Pearson's product moment r_{xy}

Phi incorporates the least assumptions about the data; it is accordingly the weakest of the measures of association. Pearson's r, on the other hand, incorporates the most stringent assumptions about the level of measurement; it requires interval data in both variables, and it is the strongest measure of association. If a coefficient is chosen which is based on weaker assumptions than the measures warrant, for example performing Spearman's r with two interval measures, then the result is still valid, but

[19]Ratio variables have all the properties of interval variables with the addition of an origin at absolute zero, but since there are no correlation coefficients which assume this higher level of measurement, they will not be discussed here. See Stevens (1946) for what is considered the modern "classic" treatment of the differences between levels of measurement.

[20]This discussion of the level of measurement is also relevant to the next step in Walker's analytical procedure, multiple regression on the criterion variables, since one of the basic validity assumptions of multiple regression analysis is that, "all variables are interval-level variables" (Loether and McTavish 1974:308).

it is understated. If, on the other hand, a coefficient is chosen that is too powerful for the measures being tested for association, for example using Pearson's *r* with two ordinal measures, then the result will be artificially high, and the reported correlation coefficient is not completely valid.

The table in (5) offers a categorization of Walker's predictor variables by their level of measurement and the appropriate coefficient for each when correlated with an interval level criterion variable. If Pearson's *r* were used on the dichotomous and ordinal level predictor variables, several of the correlation coefficients would be higher than they should be.

(5) Level of measurement and appropriate coefficients for predictor questions (scoring information from Walker (1987:87–94); table assumes the criterion variable is an interval measure)

Level of measurement	Predictor variables	Correlation coefficients
Dichotomous	11, 12, 13, 14	point biserial
Ordinal	4, 7, 8, 10, 17, 18, 19	Spearman's rank order
Interval	1, 2, 3, 5, 6, 9, 15, 16, 22	Pearson's product moment

Predictor variables 4 and 9 are of special interest. They appear to be interval data on the surface, but each is partly computed by multiplying a weighting factor, the bilingual level, which is clearly ordinal. The end result is a scale that appears to be distributed in even increments, but probably is not.[21] This is because the bilingual proficiency scale ranges from 0 to 5, and the increments are not generally considered equal. That is, the increment from 0 to 1 is not equivalent to the increments from 1 to 2, 2 to 3, etc.[22] The effect of multiplying a population percentage (which is interval data) by the bilingual level (which is ordinal data) is to create a composite number which serves to rank the communities after a fashion, but with respect to which the interval distance between scores is not interpretable in the same way as, for example, temperature, weight, age, and percentage.

[21]Essentially the same argument may be made concerning criterion variable 4, the usage of vernacular language Scriptures in churches. The scoring procedure for criterion variable 4 is similar to that of predictor variables 4 and 9 (Walker 1987:77). If one accepts that criterion variable 4 is essentially an ordinal variable, then (5) may be expanded using the information in (4) to create a column of coefficients for ordinal criterion variables.

[22]According to John Bordie of the University of Texas, Austin (personal communication), the increments implied by the FSI scale are probably more like the steps in a Fibonacci sequence. That is, rather than the increments being approximately equal, they are probably related in some (inexact or unspecified) geometric fashion.

In fact, the ranking of the composite number itself is not necessarily straightforward.

A hypothetical example illustrates the problem. For predictor 4, national language proficiency, the score consists of the average of the scores of each subgroup of the population (younger males, older males, younger females, older females). The score for each subgroup is determined by summing, "the products of the percentages of," the individual subgroups, "speaking the NL at a given level times that proficiency level," (Walker 1987:88). The table in (6) presents data for predictor 4 for three hypothetical communities.[23]

(6) Proficiency data and scores for three hypothetical communities for predictor 4.

	A		B		C	
Proficiency level	% Population	Products	% Population	Products	% Population	Products
0	40	0				
1						
2					50	100
3			100	300	20	60
4					6	24
5	60	300			24	120
Score		300		300		304

As we can see from (6), the three communities have very similar scores, but their makeup is very different. Communities A and B each have a score of 300. In community B, bilingual proficiency is moderate and is spread evenly throughout the community. Community A, on the other hand, contains two groups. About 40% of the population is nearly monolingual, and 60% of the population has virtual native proficiency in the national language.[24] The majority of community C is characterized by moderate bilingual proficiency, and a significant minority have native-like proficiency. (Rensch's paper "Community language profiles" (this volume) provides a Pakistani example of such variation).

[23]The subgroups have been collapsed to simplify the example. A more complete example with data on all subgroups would actually reinforce the point made here that the measurement resulting from predictor 4 is not readily interpretable. That is, with the additional data, there is potential for considerably greater complexity.

[24]Though a community like A seems improbable at first glance, something similar could occur in a situation of extreme impermeable social stratification such as a caste society.

Three observations may be made based on the hypothetical data. First, the composite score thoroughly masks potentially significant social dynamics within the communities. Communities A, B, and C have nearly the same score, but are radically different in their bilingual behaviors. Second, the level of measurement is probably ordinal and not interval. That is, the distance between individual increments of the scale do not appear equal. Third, while the level of measurement is probably ordinal, it is not clear that all the rankings implied by the scale are meaningful or interpretable. There is a clear ranking across the extremes of the scale from 0–500 (no proficiency to perfect proficiency), but it is not apparent whether closely clustered scores as in (6) may be meaningfully ranked.

Reliability of the data

The data for Walker's study are primarily drawn from the responses to a questionnaire circulated to SIL fieldworkers. In evaluating the results of the study, it is important to (1) examine types of questions asked and (2) make some judgment about the level of reliability we can expect in the answers. The following discussion will not exhaustively address the individual questions from the questionnaire but, rather, will serve to indicate potential sources of error and lack of reliability.

The obvious key to evaluating most of the questions is to consider what the fieldworker is being asked to estimate and how reliable the estimate is likely to be. A number of the predictor questions are relatively straightforward and are matters which any two observers familiar with the community and its history should be able to agree upon independently. Examples of such questions are:

1. How many hours travel is it to a town where the national language (NL) is widely used?
16. List the number of symbols in the vernacular language (VL) orthography that are not found in the NL orthography or which have different phonemic values.

This type of question which involves direct measurement (with a watch or simply by counting discrete graphic symbols) should yield fairly reliable data. Most of the questions which have a bearing on the substantive issues of the study, however, call upon the respondent to make highly subjective estimates of things like the level of proficiency in the national language for different portions of the population (predictor 4) or the percentage of homes in the community where one spouse is not a mother-tongue speaker of the vernacular language (predictor 2). In some very small communities

where the fieldworker intimately knows every individual, questions like predictor 2 may be answered with some reasonable degree of accuracy. In larger communities, however, there is little assurance that a fieldworker can estimate even relatively straightforward factors like the percentage of nonmother-tongue spouses accurately.

Part of the problem is that while the fieldworkers have the advantage of residing in or near the communities under study, they are not necessarily trained observers reporting the results of systematic observation or measurement. Some of them may actually be trained observers who have a prior interest in a portion of the data that Walker is asking for. The data from these fieldworkers on questions related to their prior interest will have a much greater probability of being accurate.

On the other hand, when we consider fairly complex behaviors such as estimating bilingual proficiencies of various subgroups of the village, as called for in predictors 4 and 9, it is not at all certain that any given fieldworker can make reliable estimates without conducting extensive testing.[25] In fact, following Huff's arguments, we should not even assume that the errors made by different fieldworkers will balance themselves out over the sample (1954:106).[26]

Looking a little more closely at predictor 4, we can see that there is a potential within some of the variables for compounded errors. First, the proficiency levels may be misjudged. That is, respondents may not be clear about what behaviors the categories imply. (As a variant of the problem, the categories are sufficiently vague that they may be conceptualized differently by the respondents). Second, estimates of the percentages of the populations at the various levels of proficiency may be wrong, possibly by large margins. (Contrary to daily experience, however, we might find some rare cases in which two wrongs do make a right). Predictors 9 and 15 and criterion 4 are subject to this potential of multiple errors of judgment, as well.

In evaluating the content of the questions to estimate the reliability of the data they return, we have to confront the fact that the attempt to "assess attitudes" without assessing attitudes is slightly fudged. Several of the predictor questions are, in fact, attitude questions. Note particularly that predictors 8, 10, 15, and 17 require the fieldworker to guess what the attitudes and motivations of others are with respect to diverse factors each of which are very likely to be sensitive to an array of social vectors.

[25]See SIL (1987) for an example of the complexities involved in obtaining a fairly accurate profile of community bilingualism based on the type of scale Walker proposes using.

[26]This is especially true in this case since we can't even make the assumptions that normally accompany a random sample.

Consider predictor 8, "how important do the people feel proficiency in the NL is to economic advancement?" No objective criteria are given by which the respondents may assess this "attitude" of the people. It is left, then, to the subjective estimate of the individual fieldworker. Respondent disposition such as personal optimism or pessimism will contribute to the outcome. The score for this item (and other attitude items) may be overly influenced by the respondent's more intimate acquaintances among the language community or by certain memorable events. The data, then, do not necessarily depend upon the range of response present in the community; the data depend upon the response raised in the fieldworker. When we consider the complexities of attitudes as presented above, it would seem that even a fieldworker familiar with a particular community would need to do specific investigation into attitudes to be able to provide reliable data on them. To answer these questions in the context of a field survey would undoubtedly require more than a few days of casual observation; it would require carefully crafted direct and indirect questions, interviews, and careful, systematic observation.

The point is that the data from the questionnaires should not be construed as reliably reflecting "objective" reality. Instead, for the most part, the data reflect the opinions of fieldworkers about factors which they have not measured. The only thing which these data may be said to reliably represent is the field worker's opinions about conditions in the village. In some cases these opinions may conform closely to what actual measurements of the real world phenomena under study would have been. In others they will not, but there is no way to judge the accuracy of the responses without careful, independent validation studies.[27]

When we realize the reliability of the data supplied by long term fieldworkers on many of the predictor variables is suspect, what then are the implications for surveyors who are conducting relatively brief and fleeting surveys? Only the very most obvious data, such as how long does it take to get to the next town, is likely to be accurate. Estimates, for example, of how many males between the ages of 10 and 25 speak the national language at an FSI level of three (part of predictor 4) are highly suspect from a local fieldworker if he has not indeed done a rigorous assessment of this group's bilingual abilities; as the subjective opinions of

[27]There is one attempt at validation mentioned in the dissertation (Walker, 1987:75), but since the validation consists of comparing questionnaires taken from three pairs of coworkers who presumably would have discussed many of the relevant issues during the course of program planning and implementation, we shouldn't assume the type of independence of response necessary to validate observer accuracy (or in this case, observer consistency). Validating observer accuracy, on the other hand, would require comparing observer response with actual measurements of the phenomena in question.

a surveyor, the estimates are virtually worthless. It is, in fact, a very complex and time-consuming process to accurately assess bilingual proficiency (and many of the other predictors, as well).

In planning (or evaluating) a survey or a study of this sort, there are several types of data which we might consider. The most reliable is that which is based on careful measurement of the population. For example, a thorough household census of the community should yield a fairly reliable picture of marriage patterns, professed language loyalty, etc. Likewise, thorough testing of bilingual proficiency throughout the community should give a reliable picture of the bilingual behaviors under scrutiny. The problem inherent in this type of data is that it is expensive and very time-consuming to gather.

The next most reliable type of survey data is that which is based on careful measurement of a random sample of the population. For larger communities, we can rarely perform measurements of entire populations. If the sample is random, however, we can at least state the statistical probability that our data are a reflection of the larger population. The point is, in our search for data, we are constantly confronted with the need to balance time and cost effectiveness against the degree of reliability necessary to inform our decision-making processes and program implementations. The figure in (7) is an attempt to graphically portray the trade-offs implicit in some of our choices.

There are many situations in which cheap data with a low reliability index (often referred to as "quick and dirty" data) are appropriate. Probably the key concern is what is the cost of making a wrong decision? If the cost of being wrong is low, "quick and dirty" data are probably the best choice. If, on the other hand, the cost of being wrong is high, such as bypassing a group that genuinely needs a program or filling an unnecessary allocation, then the cost of gathering more reliable data is justified.

(7) Cost-reliability ratio of various survey data sources[28]

+Expense +Reliability			−Expense −Reliability
	measure population	measure random sample	measure representative sample
	opinions of insider populations	opinions of insider random sample	opinions of insider representative sample
	opinions of informed outsider population	opinions of informed outsider random sample	opinions of informed outsider representative sample
−Expense −Reliability			opinions of surveyors

Predicting vernacular literacy acceptance

As noted above, given the sampling problems connected with Walker's research, any generalizations, and, therefore predictions, based on this data are ill-conceived. However, this line of research appears to hold a great

[28]This chart is only meant to provide a rough rule of thumb. It is a relatively easy matter to think of exceptions. I believe the topology to be generally correct, however, and of course, there are other data sources which could be integrated. For example, availability samples could be included in an additional column on the right. A third dimension, complexity of data required, could be included in a three dimensional array to account for the differential reliability in answering questions like Walker's predictor one concerning travel time to nearest NL town versus the bilingual estimates called for in predictor four.

The general assumption of the table is that surveyors are collecting the data throughout. Thus in the final row, "opinions of surveyors" refers to instances where the surveyor provides opinions in contrast to data collected in higher rows or without reference to data from higher rows. Note also that this table begs the question of careful participant observation which is not generally within the purview of a relatively brief survey visit.

deal of promise and will probably be refined in the future.[29] Therefore the role of this model (and others like it) in prediction should be briefly addressed.

The theoretical underpinnings of Walker's research derive from the literature on language shift, and there is an overt desire to build a model with predictive power to use in assessing the potential for vernacular literacy acceptance. However, as Fasold writes concerning prediction of shift,

> Just as we saw in the case of language choice, however, where the same factors were cited independently by many scholars, there has been very little success in using any combination of them to predict when language shift will occur. In fact, there is considerable consensus that we do not know how to predict shift. (Fasold 1984:217)

It is unlikely that this attempt to predict vernacular literacy acceptance will prove successful either. Walker himself admits that his, "model... has not matured far enough to accurately predict how readily a community will accept VL literacy" (Walker 1987:202). Some crucial knowledge is surely lacking in our understanding of the processes of vernacular literacy acceptance, but the problem of prediction seems more fundamental than simply our temporary ignorance.

In a second paper in this volume, I address the mechanical aspects of correlation and prediction, noting that in correlational analysis prediction does not mean guessing right; it means guessing less wrong. It is always tempting to hope that we can progressively refine our predictive models until we get it right, but as Gregory Bateson points out in a major work on the epistemology of science, "the generic we can know, but the specific eludes us... There is a deep gulf between statements about an identified individual and statements about a class... and prediction from one to the other is always unsure" (Bateson 1980:45–46). Bateson's point is that the view that "a little more knowledge and, especially, a little more know-how will enable us to predict and control the wild variables... is wrong, not merely in detail, but in principle" (Bateson 1980:44).

Somewhat less eloquently, social forecasters, Richard Berk and Thomas Cooley, note,

> There is no disputing that forecasts of social phenomena will almost inevitably be wrong. Social phenomena are either inherently

[29]Unfortunately Walker's current program (1988:35, 41–45) for gathering more data by inviting whoever will to send additional data to him is doomed to perpetuate the sampling and response bias problems noted above. It will not produce a sample from which valid statistical inferences can be drawn.

> stochastic or as a practical matter, must be treated as such . . . Most of the time forecasts will be wrong. (Berk and Cooley 1987:247, 263)

Walker's "model" is not a model in a technical sense. A technical model consists of an evaluation index derived from the regression analyses. That is, a model should contain a formula into which the predictor scores are fed to produce an index for vernacular literacy acceptance. A model is necessary to interpret the mass of data produced when a survey is undertaken to assess the prospects for vernacular literacy in a language community.

If we gather the data on the predictors, how do we know what it means? Some factors may appear positive; some may appear negative. How do we tell when the mixed positives and negatives mean that vernacular literacy acceptance is unlikely? Without a model and a way of interpreting it, there is no basis for using the material in decision-making. Walker writes concerning the use of this "model" in Irian Jaya, "at this point, we have not adopted a formula. Our predictions of VL literacy acceptance and decisions regarding priority language project status are still a product of subjective evaluation" (Walker 1991:86).

Essentially this means that there is no standard based on Walker's model for evaluating the scores a community receives as it is assessed. Language program decisions based on interpretation of survey data on these predictors are, then, *ad hoc*. This is because the "model" does not predict well enough for reliable decision-making. It also means that language groups which are subjected to an administrative evaluation using this data are not receiving a disposition based on objective criteria; rather, they are left to the vagaries of the subjective evaluation of whoever happens to be the evaluator at the time.

Conclusion

Walker's claim has been that this line of research provides an effective means of assessing attitudes. The discussion above has shown that the model in its present form is too impoverished to adequately reflect important intercommunity variation in attitude structures. Additionally, the sample was drawn in such a way that it has no external validity, and, therefore, generalizations and statistical inferences have no meaning for any cases other than those in the original study.

Even if there were no sampling problems or other potential problems with the reliability of the data from some of the questions, the model's inability "to accurately predict how readily a community will accept VL

literacy" (Walker 1987:202), and the lack of a technical evaluation model to assess the data for future surveys severely restrict any effectiveness it might have as a decision-making tool.

The technical criticisms of Walker's research should not be taken to mean that the research is of little worth or that it shouldn't have been done. On the contrary, the research is original and likely to prove quite valuable. Though the model should probably not be used as a forecasting tool, the information it generates may provide some of the levers language planners and vernacular literacy workers need for improving the acceptance of vernacular literacy in many minority-language communities.

Of particular interest is Walker's finding that orthography design and the involvement of community leaders can significantly improve the chances for vernacular literacy acceptance. While this may seem intuitively correct, the regression analysis has confirmed that it is, in fact, correct for the 54 programs in the study. If this observation proves generalizable beyond Walker's study, then there is reason to hope that vernacular literacy may be successfully introduced into communities with clear comprehension needs but poor prospects for accepting vernacular literacy. Likewise, Walker's finding that the "percentage of the community... who aim at living their lives according to the Bible," (predictor 15) significantly and positively affected the acceptance of vernacular literacy, holds out hope that spiritual change outside the direct responsibility of the literacy worker can lead to greater acceptance of vernacular literacy.

Walker has done us a considerable service by probing beyond our previous level of knowledge about the processes of literacy penetration into minority language groups. Hopefully, we have been provided with tools which can help us improve the acceptance of vernacular literacy in those communities where translation and literacy programs are needed.

Appendix G
Walker's Attitude Questionnaire

The following questions are from Walker's original questionnaire which was included as a part of his dissertation (Walker 1987:238–45). They have been renumbered here for ease of reference; the numbers in parentheses are the original questionnaire numbers. Responses are recorded as applying to the vernacular language (VL) and the national language (NL).

Criterion variables

1. (43) How many people have purchased (or wanted to receive as a gift) Scriptures (either the New Testament or Scripture portions)? VL = __ NL = __

(44) How many people have purchased (or received) other types of literature (i.e., not Scriptures)? VL = __ NL = __

2. (45) What is the percentage of the population who can read narratives with understanding? __%

VL = __% age 10–25; __% age 26–40 NL = __% age 10–25; __% age 26–40

3. (46) What percentage of the population spend time reading (any kind of literature) weekly in informal settings (i.e., outside church and school)?

VL = __% age 10–25; __% age 26–40 NL = __% age 10–25; __% age 26–40

4. (47) For each church in the community, are Scriptures read aloud in church meetings? 3 = every meeting; 2 = most meetings; 1 = some meetings; 0 = not at all

(list each church)	VL	NL	Average Attendance
_____	3 2 1 0	3 2 1 0	_____
_____	3 2 1 0	3 2 1 0	_____, etc.

Predictor variables

1. (8) How many hours' travel (by the commonest [sic] mode) is it to a town where the NL is widely used? ____ hours

2. (27) Intermarriage. Estimate the percentage of homes in the community in which one spouse is not a mother-tongue speaker of the VL. ____%

3. (21) Estimate the percentage of homes in the community where people live who are not native to the community and do not speak the VL. ____%

4. (19) Use the Rating Scale below to estimate PROFICIENCY IN THE NL for each of the categories below. Put the percentage of that category of the people in the boxes below the appropriate proficiency level. (See the Example—a situation in which 40% of the males age 10–25 are at level 0 and 60% are at level 1.)

Rating scale

Level 0. No ability.

Level 1. Can carry out minimal activities in daily living in the language.

Level 2. Can respond to opportunities and interact in routine social situations and limited work requirements.

Level 3. Can satisfy normal social and work requirements with sufficient structural accuracy and vocabulary to meet these limited needs.

Level 4. Can communicate effectively with vocabulary that is always extensive and precise enough to convey exact meaning.

Level 5. Native speaker fluency.

Example:

sex/age	0	1	2	3	4	5	
males 10–25	40	60					= 100%

Proficiency in the NL (% levels)

sex/age	0	1	2	3	4	5	
males 10–25							= 100%
males 26–40							= 100%
females 10–25							= 100%
females 26–40							= 100%

5. (30) What is the average number of years of formal education completed by adult males? ____

6. (29) What percentage of VL readers could read the NL first? ____%

7. (25) Economically, for the people in the community... (Check one)

( ) 0–most can earn a living as they traditionally have

( ) 1–a few are beginning to leave the community to find jobs

( ) 2–more and more are leaving to find jobs on the outside

( ) 3–many people leave the community to work for wages

8. (26) How important do the people feel proficiency in the NL is to economic advancement? (Circle a number) Not important – 0 1 2 3 4 – Very important.

9. (24) What is the % of the adult population needing the NL to carry out their occupation? (See no. 4 for proficiency levels) List common occupations:

____ % men need spoken proficiency at Level ____ _____

____ % men need written proficiency at Level ____ _____

____ % women need spoken proficiency at Level ____ _____

____ % women need written proficiency at Level ____ _____

10. (33) What is the prevailing attitude of local government officials, who are not VL speakers (e.g., schoolteachers or whoever is most influential in the community) to the development and use of the VL for literacy?

Negative – 0 1 2 3 4 + Positive

11–14. (9) Circle which language is most dominant in each domain for spoken use.

	Language		Domain[1]
	VL	NL	home
	VL	NL	community
11.	VL	NL	church/religion
12.	VL	NL	occupation
13.	VL	NL	school classroom
	VL	NL	government
14.	VL	NL	singing

15. (12) Estimate the percentage of the community (of any religious affiliation) who aim at living their lives according to the Bible. ___%

[1]The domains of home, community, and government did not receive sufficient response to be included in Walker's final model.

16. (22) List the number of symbols in the VL orthography that are not found in the NL orthography or which have different phonemic values.

Number	Items	Symbols
	consonants	
	glottal stop	
	vowels	
	nasalized vowels	
	vowel length (phonemic or ballistic [sic])	
	accent	
	tone	
	other	

17. (23) How difficult do people in the community view reading the VL? (Check one)

( ) 0–It is much more difficult to read than the NL

( ) 1–It is fairly difficult to read compared to the NL

( ) 2–It is about the same as reading the NL

( ) 3–It is fairly easy compared to the NL

( ) 4–It is very easy compared to the NL

18. (39) To what extent were community leaders involved in orthography decisions?

( ) 0–Actively opposed to the SIL produced orthography

( ) 1–Not involved, neutral

( ) 2–Involved and supportive of the orthography

( ) 3–Enthusiastic promoter/s of the orthography

19. (40) To what extent were community leaders involved in other aspects of the VL literacy program?

( ) 0–Opposed to it

( ) 1–Not involved at all

( ) 2–Involved to some degree

( ) 3–Actively involved

( ) 4–Enthusiastic promoter/s