DOCUMENT RESUME

ED 156 691 TM 007 103

AUTHOR Baldauf, Richard B., Jr.
TITLE Evaluation Models and Instrumentation: Problems for Title I in America's Pacific Possessions.
PUB DATE Feb 78
NOTE 25p.; Paper presented at the Trust Territory Title I Conference (February 9-14, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Cloze Procedure; Cost Effectiveness; Decision Making; Elementary Secondary Education; *English (Second Language); *Evaluation Methods; *Models; *Norms; *Program Evaluation; Standardized Tests; Systems Approach; Test Construction; Test Selection
IDENTIFIERS *Elementary Secondary Education Act Title I; Local Norms Model; *Pacific Trust Territory

ABSTRACT

An overview of the system analysis approach to evaluation and the psychometric models currently proposed to evaluate Elementary and Secondary Education Act Title I projects is presented. It is argued that, in America's Pacific possessions, the sole use of the systems analysis approach to evaluation fails to provide adequate information to questions of central importance for English as a second language program development. In this particular setting, the proposed models, although well researched and considerably refined in recent years, are often inappropriate and difficult to implement correctly, and due to the cost factor they may actually detract from program quality. Tests designed for use in the continental United States are generally inappropriate for islands where English is a second language. Three models developed by the RMC Research Corporation for Title I evaluation are discussed: (1) the norm-referenced model; (2) the control group model; and (3) the special regression model. Two preferable evaluation models would be the use of standardized tests with local norms, and the development of cloze tests with local norms. (Author/CTM)

EVALUATION MODELS AND INSTRUMENTATION: PROBLEMS FOR TITLE I IN AMERICA'S PACIFIC POSSESSIONS[1]

Richard B. Baldauf, Jr.
James Cook University of North Queensland
Townsville, Australia

ABSTRACT

An overview of the systems analysis approach to evaluation and the psychometric models currently proposed to evaluate ESEA Title I projects is presented. It is argued that, in America's Pacific possessions, the sole use of the systems analysis approach to evaluation fails to provide adequate information to questions of central importance for English-second-language program development. In this particular setting, the proposed models, although well researched and considerably refined in recent years, are often inappropriate, difficult to implement correctly, and due to the cost factor may actually detract from program quality. A further complicating factor is the lack of adequate instruments (tests) for the models to be validly implemented. Reasons for these problems are detailed and an alternative model and some possible solutions to the instrumentation problem are suggested. The paper recommends that allowance be made for a wider variety of evaluation approaches in the final Title I guidelines.

---

1. Paper presented to the Trust Territory Title I Conference, February 9 - 14, 1978. This research was supported by a University Research Grant from James Cook University, and by funds provided by the U.S.O.E. to the Mariana Islands under ESEA Title III. The opinions stated in the paper are those of the author and no endorsement of the work by the U.S.O.E. is implied.

EVALUATION MODELS AND INSTRUMENTATION: PROBLEMS FOR TITLE I IN AMERICA'S PACIFIC POSSESSIONS

Richard B. Baldauf, Jr.
James Cook University of North Queensland

The systems analysis approach (McLaughlin, 1975) is one of a number of strategies which have been proposed for and used in educational program evaluation (i.e. Worthen and Sanders, 1973). It is particularly important however in the American context because "it has served as the major evaluation perspective in the United States Department of Health, Education and Welfare since about 1965 (House, 1977, p.1)." In recent years, a particular subset of evaluation strategies has been developing within the general systems analysis framework, and these models may soon become the required evaluation and reporting methods for all Elementary and Secondary Education Act (ESEA) Title I projects. These projects, which focus on compensatory instruction in the basic skill areas of reading and mathematics, involve large scale funding of programs in most of the 50 states and in America's Pacific possessions. Thus, the potential widespread use of the proposed models and their implications for program evaluation make them ones with which program managers and evaluators should be familiar.

The Systems Analysis Approach

A system analysis approach to evaluation is based on relating quantitative output measures, usually test

scores, to program differences (House, 1977). This has most often been accomplished through the use of correctional techniques, but increasingly has involved the use of experimental design (i.e. Marco, Murphy & Quirk, 1976). The key concepts in the systems analysis approach as put forward by its leading proponent Alice M. Rivlin (1971) have been summarized by House (1977,p.8) as follows:

- Key decisions will be made by higher governmental levels.
- The end of evaluation is efficiency in the production of social services.
- The only true knowledge is a production function specifying stable relationships between educational inputs and outputs.
- The only way to such knowledge is through experimental methods and statistical techniques.
- It is possible to agree on goals and on a few output measures.
- There is a direct parallel between production in social services and in manufacturing. The same techniques of analysis will apply.

The aim of the evaluation process in the systems analysis approach is to provide generalizations that will hold in various circumstances. Large samples are required, both to detail the range of circumstances, and to rid the final production function of idiosyncracies. The final product will enable the major consumers of evaluation, the managers and administrators of governmental programs, to produce social services more effectively (House, 1977).

Background to the Title I Evaluation System

Title I of ESEA currently provides over $2 billion annually to educational agencies, serving approximately

9 million students in areas with concentrations of children from low income families, to provide remedial services in basic skills. Although the reporting procedures for such grants provided for appropriate objective measurement to be submitted on at least an annual basis, many of the reports submitted at the federal level have shown a "lack of comparability --- and often validity --- of the data in them (Anderson, 1977, p.2)." The need for better evaluation procedures was further documented by a literature search which

> encompassed some 2,000 projects, all of which had received some form of "official" recognition for success. Of the 2,000, only six could be found which, under close scrutiny, were able to meet the selection criteria of effectiveness, cost, availability and replicability established for this search. Most discouraging, however, was the fact that not one of the evaluations provided acceptable evidence regarding project success or failure. In all cases, problems in conducting and reporting the evaluations rendered the results inconclusive (Horst, Tallmadge & Wood, 1974, 1-2).

In an effort to insure a greater degree of accountability, the United States Congress passed section 151 of the Title I Act in August, 1974. Under this section the United States Office of Education (USOE) was required to implement a

> complete evaluation program: conduct evaluations, upgrade evaluation activities at other administrative levels so that reported data are comparable, use those data to --- among other things --- identify especially effective instructional practices, and disseminate information about those practices (Anderson, 1977, p.3).

Under this legislative mandate, a series of contracts were awarded to the RMC Corporation. The first provided

for a review of Title I program evaluation, recommended common reporting practices and checked on the feasibility of the resulting suggestions with a small sample of administrators and evaluators (Gamel, Tallmadge, Wood, and Binkley, 1975). The second involved all states and territories and some local school districts in a discussion of a prototype system and its implications for those settings (Bessey, Rosen, Chiang, and Tallmadge, 1976).

From these discussions have emerged three structured evaluation models, with the provision for the use of other models if they can provide comparable data. The system is expected to allow for "the aggregation of unbiased, project-valid estimates of the effects of Title I services, ... expressed in a common metric to make such aggregation possible (Anderson, 1977, p.11)." Besides providing for aggregation, the system is expected to be useful to identify especially effective educational programs, to facilitate the monitoring and guidance of programs, and to help upgrade evaluation activities, "Hence, our hopes for the system are that decision makers at all levels will find benefits from its use (Anderson, 1977, p.13)." Local education authorities (LEA's) are expected to begin implementation of these models in the 1979-1980 school year.

The Evaluation Models

The evaluation models first appeared in October 1974 as an RMC report entitled Measuring Achievement Gains in Educational Projects (Horst, et al, 1974). The Government Printing Office version is reported to have sold over 8000

copies during its first year in print (Anderson, 1977). As indicated in the preceeding section, three models have been selected from this initial report and further refined for use with Title I projects.

> These designs are: Model A, the Norm-Referenced Model; Model B, the Control Group Model; and Model C, the Special Regression Model. Further flexibility is afforded in that each design has variations to accomodate the use of either normed achievement tests (Models A1, B1 and C1), or tests for which normative data are not available (Models A2, B2, and C2) (Tallmadge & Wood, 1976).

Model A, the Norm-Referenced Model, assesses gain by comparing the pre- and posttest percentile status of students either directly with national norms (Model A1) or indirectly by using the equipercentile method to equate non-normed test scores with a normed test given at the same time as the pretest (Model A2). In addition, the model requires that (a) pre-posttesting must occur on the empirical normative dates, (b) the same level and form of the test must be used for pre- and posttests, and (c) pretesting must occur <u>after</u> project sample selection to avoid regression effects (Tallmadge <u>et al</u>, 1976).

Model B, the Control Group Model, requires that pre-posttest data be collected at the same time for both the Title I treatment and for the random or "random in effect" control groups. <u>Post hoc</u> matching procedures are not allowed, but either an analysis of covariance (assuming random assignment) or a standardized-gain-score method of adjustment (assuming different populations) may be used to adjust for pre-test differences. In addition,

all supplementary instructional services must be withheld from the comparison group. Although testing dates are flexible and different tests or levels of the same test may be used, the treatment group must take a nationally normed test sometime during the year (Tallmadge et al, 1976).

Model C, the Special Regression Model, incorporates two different evaluation designs, the regression-discontinuality design (Campbell and Stanley, 1963) and the regression-projection design (Horst et al, 1974). Pretesting must include the entire group from which the treatment and comparison groups will be formed. Students are assigned to the treatment group only on the basis of the pretest cut off score while comparison group students may not receive any special instructional services. Care must be taken not to remove slow or disruptive students as this may invalidate the evaluation process. Ideally about 50 treatment and 100 comparison pupils should be used to implement the model. The treatment group must take a nationally normed test sometime during the year. However, different pre-posttests may be used if they correlate highly (.40 or better) (Tallmadge et al, 1976).

The models discussed in the proceeding paragraphs are recommended for implementation "in a hierarchy based on technical desirability" (Model B, then Model C, then Model A), however "choosing a model will always involve making trade offs between technical and practical considerations (Tallmadge et al, 1976, p.19)." Figure 1 summarizes the models and indicates the key decision points for

model selection.

Insert Figure 1 About Here

Finally, regardless of the model selected, the results are converted to Normal Curve Equivalents (NCE's) to provide for comparisons across Title I projects. This last point is important because it emphasizes the underlying systems analysis basis of the evaluation system which is outcome rather than process oriented particularly at the project level. The system's originators however take the point of view that ...

> the data called for by the proposed system will do more than provide evidence regarding overall effectiveness of the Title I program. The system will permit analysis of project-level relationships among cost, achievement gains, hours of intervention, grade levels, instructor pupil ratios, and initial degree of educational need. It will, then, enable investigation of most of the major and minor concerns expressed by educational policy makers interviewed during Phase I of the study (Talimadge et al, 1976, p.12 - draft version).

Applying the Models in America's Pacific Possessions

The work discussed thus far represents an impressive job of synthesizing empirical evaluation designs and detailing models for a wide variety of evaluation needs. Undoubtedly, evaluation has developed to a point where attempts at synthesis should be encouraged (Gephart, 1977). Certainly, the need for a wider variety of valid and comparable evaluation studies is evident. However, to make the evaluation models discussed in the proceeding paragraphs mandatory for all agencies who wish to receive

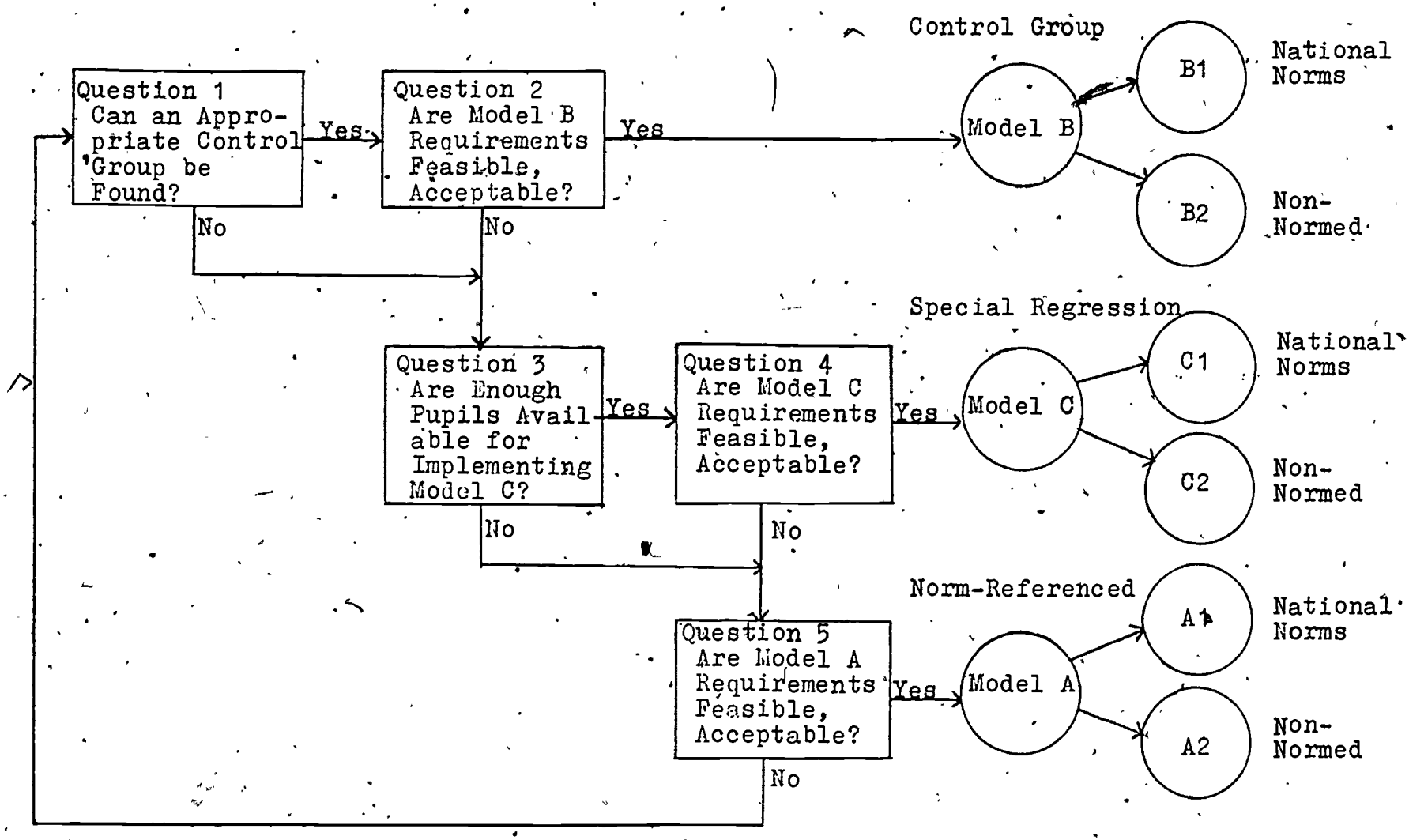

Figure 1. Decision tree for selecting an evaluation model. (Adapted from Tallmadge et al, 1976, p.20)

Title I funding, even if none of the models is appropriate, is not only bureaucratic mismanagement, but introduces error, in the form of invalid summary results, into the aggregation system which is central to the proposed system. This in turn undermines the usefulness and validity of generalizations obtained. The following sections attempt to validate the hypothesis that none of the models as they stand is normally appropriate for evaluating Title I English-second-language programs in America's Pacific territories.

Selection of A Model

In the Pacific island setting, where schools tend to be small, where communities even within a culture are relatively unique, where communication is difficult, and where the majority of students only begin to study and speak English when they come to school, the evaluator has few options when it comes to model selection. This statement becomes clearer when the five questions (see Figure 1) posed to aid in the selection of an evaluation model are reviewed.

1. Can a suitable control group be found? Because of community differences, cultural differences, different degrees of exposure to the English language and to western culture, and the fact that there are only a small number of schools, usually with one class to a grade, it is difficult to find control groups except in a few of the larger, mainly secondary schools.

2. Can the requirements imposed by Model B be met and are they acceptable? In the few cases where a

treatment and control group selected in the same manner can be found, the use of Model B is often not appropriate because of the requirement that students in the control group may not receive any supplementary instructional services. In most cases, students in America's Pacific Region schools are involved in one or more compensatory programs. In American Samoa for example, all students in grades 8 - 12, except a few exempted first language students, were involved in Title I ESL programs. The similarity of bilingual and some Title III programs in method and approach to Title I further complicates the use of Model B. These group differences and the problems of program overlap make it difficult to meet the requirements necessary for the control group evaluation strategy to be implemented.

3. Are there enough participants and non-participants at each grade level to enable implementation of Model C? Adequate numbers are a problem to some degree with this approach particularly at the elementary school level. Since students must be assigned based on pretest scores to either the treatment or control group, students must come from one school so that a rearrangement of classes is possible. Small, one class per grade, schools are thereby effectively excluded from Title I programs and the problem of over testing and multiple program interaction is increased in the larger schools.

4. Can the requirements imposed by Model C be met, and are they acceptable? Although this Model C has been used as the best available in the circumstances for evaluation in the Trust Territory, it has some of the

same problems as Model B. In particular, it may be difficult to meet the requirement that all compensatory services be withheld from the control group.

5. Can the requirements imposed by Model A be met, and are they acceptable? Model A requires the administration of at least one appropriate normed-referenced test. There currently are no norm-referenced tests available for ESL students and so the Model can not be validly implemented. The question of valid tests is one which applies, of course, to all the models and because of its importance will be examined in detail in a separate section.

Implementation Problems

Although the use of one of the models is/will be a Title I evaluation requirement, it is only in rare circumstances that any of the models may be used in America's Pacific possessions validly without violating some major evaluation design assumption. Thus, without even considering the problem of what to do to obtain valid instrumentation, it is difficult to find an appropriate Title I evaluation design.

The use of the models and their proper implementation is further complicated by the fact that the Pacific territories have unique problems when compared to other Title I programs. These include vast distances, few trained personnel, high staff turnover, more limited access to external consultants, and higher charges for evaluation materials and services. These problems are compounded by the Title I programs' comparatively small

budgets, based on small numbers of pupils served. Madden (1976) has pointed out some of the potential implementation problems inherent in the models and has emphasized the need for greater training and support programs by the USOE.

Local Norms

It might be argued that many of the above problems can be solved by using tests with local norms. Although this is a step in the right direction in some respects, it does not help the evaluator to find a valid model since our previous analysis has shown that the models are usually invalid regardless of the type of testing selected. Furthermore, to use this approach it is necessary to give a "nationally normed test" in order to implement any of the three models (Tallmadge et al, 1976, p. 14). As we shall see, there are no valid nationally normed tests for use with ESL students, and so on the basis of current test availability, it is not possible to correctly implement any of the models using local norms.

An Alternative Model

For program evaluation to be valid, project students must be comparable to control students or to norms for students with similar characteristics. For the ESL students, this means they must be compared with students with second language backgrounds. There are few such students in American standardized test norming samples. Further, there is some evidence to indicate that ESL students in America's Pacific territories have weaker

English language skills than Native American ESL students in the States (Baldauf, 1978), due perhaps to the fact that there is less opportunity for and need to use English in Pacific cultural settings. Thus, Pacific island children's linguistic needs and problems are not comparable to the majority of Title I students.

These factors suggest the need for a new model which allows for these problems. I would propose the use of the Local Norms Model, which follows Model A and is the simplest of the evaluation models to implement. The Local Norms Model would use as a standard of comparison locally developed norms based on representative samples of the relevant population. Such an approach would allow evaluation model assumptions to be met and would remove the requirement to compare second language students to first language norms. Program evaluation would be greatly simplified and student progress would be measured against an appropriate standard.

Meeting Additional Evaluation Needs

I believe a Local Norms Model would have additional benefits as well. ESL programs, such as those in the Trust Territory developed to meet the unique needs of small cultural groups, require detailed feedback on "how" and "why" a program has succeeded, or failed, if adjustments to the program are to be made and if the standard of education is to be improved. The systems analysis model, which Anderson (1977) and Talmadge, et al (1976) conclude provide useful results for program managers, does not, from my own experience provide the data

necessary to improve ESL programs. Mayer (1975) goes further than this to suggest that the standardized testing necessary to the systems analysis model has tended to divert attention from the theoretical problems in reading and lessen the effectiveness of compensatory education. The systems analysis approach has little concern for the theoretical issues on which learning depends. Rather, it concentrates on collecting production like results useful for governmental administrators who need to know "if" a program is working, "for whom", and at what cost.

I believe the Local Norms Model would provide the USOE with the comparable data necessary to measure the effectiveness of Title I ESL programs. In addition, the simplified data collection and analysis procedures and the ability to use relevant local tests would provide additional time and resources necessary to undertake the formative evaluation activities required for program improvement. However, the implementation of this approach to evaluation depends directly on development of adequate, locally normed tests.

Test Selection/Construction

The problem of instrumentation is a difficult one for program evaluation in general. Tallmadge and Horst, two of the researchers most responsible for the current Title I evaluation guidelines indicate that instrumentation, the selection of valid measures of achievement,

> ...is currently the weakest link in the educational evaluation chain. Until it

> is strengthened, the information available for making policy decisions will continue to be of marginal utility (1977, p. 14).

The problem is even greater in the Trust Territory because of the lack of valid ESL tests. A number of approaches to the selection and development of valid measures of achievement are possible and each of these alternatives will be examined in turn.

Standardized ESL Tests

Buros (1974) in Tests in Print II lists 14 English foreign language tests. Most are clearly intended as measures of university entrance English language proficiency. Although tests like the Michigan Test of English Language Proficiency (Division of Testing and Certification, English Language Institute, 1962) might be useful for evaluating upper level high school programs if appropriate local norms were developed (Baldauf, 1978), the tests are generally too specialized and too restricted to provide ready made solutions to ESL testing problems. Except as noted above, they probably are most useful as examples of what has and can be done in ESL testing.

Norm-Referenced Tests

A wide variety of norm-referenced tests are available from test publishers. The Gates MacGinitie (Gates and MacGinitie, 1972), the Stanford Achievement Test (Madden, Gardner, Rudman, and Kelly, 1964; Baldauf and Reupena, 1973) and the SRA Achievement tests (SRA, 1973) among others have been used extensively in the Pacific region to evaluate programs. These tests which were designed

for first language students studying a stateside curriculum, are too difficult for ESL students when given at grade level. The practice has been adopted of giving the tests at a higher grade level than the one for which they were normed. This practice has had several detrimental effects: In particular, curriculum content is usually inappropriate due to improper test level. There is also the problem of inproper comparison and interpretation of the test results, in which ESL Pacific Island Children are compared with inappropriate (in age, grade and curriculum) stateside test norms.

Stateside norm-reference tests do have a place in Pacific island programs. Properly used they can serve as a counseling guide for students wishing to enter stateside educational programs. They are not however, appropriate instruments to evaluate ESL programs. In fact many norm-referenced tests may not be valid evaluation measures for Title I program students. Doherty (1977) in a restandardization study of the California Achievement Test, concluded that the use of CTB scores led to improbable conclusions about ESAA students which were rectified when the test scores were rescaled. He further suggests that it is likely that these problems "will be encountered in the use of any norms that have not been based specifically on disadvantaged, low achieving students (p.31)."

Locally Designed Psychometric Tests

One of the ways to tackle the problem of the lack of appropriate tests would be to follow the example of

American Samoa and develop psychometric tests based on local curriculum objectives. This is a complex process. The details of design of the test development program are available in Baldauf and Dunn-Rankin (1973) and the results and sample tests are available in a series of University of Hawaii Reports (i.e. Chin-Chance, Norton, Rechebei and Bail, 1975). Another example of a test specifically designed for ESL students at the secondary level is the English Structures Test (Catling and Gobert, 1973). However, to be useful for program evaluation, these tests would need to have local norms based on a representative sample developed for them. A project is currently underway in Saipan to evaluate and norm the Marianas Test of English Achievement (MTEA) (Klingbergs and Dorn, 1977) to provide an appropriate ESL test for intermediate level ESL students (Baldauf and Annesley, 1977). Work in this area could be simplified by starting from the Samoan tests already available.

Experience has shown that the development of psychometricly based tests is an involved process requiring considerable outside expertise and local effort to complete. Depending on how it is done, it may require considerable expense and demands expert ESL and evaluation staff familiar with local conditions to produce adequate results.

Local Norms for Standardized Tests

Another approach to the problem would involve the selection of the most appropriate norm-referenced tests available for use as they are except that local norms

would be developed on a representative sample basis for the population to be evaluated. A study of high school students in American Samoa in which correlations between English classroom grades over four years and LTELP scores were examined suggests that this process may be a valid one (Baldauf, 1978).

It would be much cheaper and quicker than developing the psychometrically based tests mentioned in the previous section. Doherty's (1977) Restandardization Study provides an example of how it could be done. The irritating factor of having obviously inappropriate items in the test, in terms of culture, could be controlled by not including these items in the raw scores on which norms were based.

CLOZE Tests

A final approach to developing valid achievement measures for ESL students is the use of CLOZE tests. CLOZE tests, which are based on the concept of the student supplying words which have been deleted from reading passages, have been successfully used to assess reading ability of ESL students in Papua-New Guinea and in Singapore (Anderson, 1976), and with ESL entrants to American Universities (Hisama, Lewis and Woehlke, 1977). Baldauf and Propst (1978) have extended the use of CLOZE tests to lower Elementary school ESL children. In Australia, two tests based on the CLOZE procedure, the GAP Reading Comprehension Tests (McLeod, 1967) for use with Grades 4 - 7, and the GAPADOL Reading Comprehension Test, (McLeod and Anderson, 1972) for grades 8 and 9 have

gained acceptance.

Locally normed CLOZE tests could be fairly quickly developed and have the advantage of being easy to make culturally relevant. They have the disadvantage however of being rather specifically related to reading achievement.

Summary

This paper has briefly reviewed the problems of Title I evaluation in America's Pacific possessions. It has examined the systems approach to evaluation, the weaknesses of the models proposed to evaluate Title I programs and has suggested the adaption of a "local norms model" for program evaluation. Several approaches to instrumentation were also discussed as a basis for implementing the local norms model. It is suggested that specific decisions about the best instrumentation for the Trust Territory can only be decided in consultation with local educators and must be based on local needs and priorities.

REFERENCE NOTES

Anderson, J.K. USOE's Title I evaluation technical assistance efforts: Background and the evaluation system. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Baldauf, R.B.; Jr. and Reupena, M. A summary of Stanford achievement test results for May 1971 and May 1972 in Samoan elementary schools. Pago Pago: Department of Education, 1972 (Pacific Collection, Sinclair Library, U. of Hawaii).

Bessey, B.L., Rosen, L.D., Chiang, A. & Tallmadge, G.K., Further documentation of state ESEA Title I reporting models and their technical assistance requirements. Mountain View, CA: RMC Research Corporation, 1976.

Catling, D. and Gobert, D. English Structures Test. Pago Pago: Department of Education, 1973.

Chin-Chance, S., Norton, R., Rechebei, E., and Bail, F.T. Final Report on Article II (Testing, Measurement, and Evaluation) of the University of Hawaii - American Samoa Contract. Honolulu: Office of Foreign Contracts, University of Hawaii, 1975.

Doherty, W. Restandardization Study. Calif: Systems Development Corporation, 1977.

Gamel, N.N., Tallmadge, G.K., Wood, C.T. & Brinkley, J.L. State ESEA Title I reports: Review and analysis of past reports, and development of a model reporting system and format. Mountain View CA: RMC Research Corporation, October, 1975.

Gephart, W.J. Evaluation reconsidered: Do we need a synthesis? Definitely! Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Hisama, K.K., Lewis, E.L. and Woehlke, P. A new direction in measuring proficiency in English as a second language. A paper presented at the Annual Meeting of the American Educational Research Association, New York, 1977.

House, E.R. Assumptions underlying evaluation models. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Klingbergs, I. and Dorn, D. Marianas Test of English Achievement. Saipan: Department of Education, 1977.

Madden, J.V. Summary of comments to the policy advisory committee on the proposed Title I evaluation models, Mimeo, March 1976.

Tallmadge, G.K. & Horst, D.P. Using the data from state and local ESEA Title I Reports. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Tallmadge, G.K. & Wood, C.T. User's Guide: ESEA Title I Evaluation and Report System. Mountain View, CA: RMC Research Corporation, 1976.

REFERENCES

Anderson, J. Psycholinguistic experiments in foreign language testing. Brisbane: U. of Queensland Press, 1976.

Baldauf, R.B., Jr. & Dunn-Rankin, P. A progress report on a plan for curriculum assessment in American Samoa. Pago Pago: Department of Education, 1973 (ERIC ED 074 200).

Baldauf, R.B., Jr. & Annesley, F.R. Item Analysis Report: Micronesian Test of English Achievement (MTEA). James Cook University, Townsville, 1977 (ERIC ED 138 641).

Baldauf, R.B., Jr. The validity of the Michigan Test of English Language Proficiency as a general measure of high school achievement in American Samoa. Educational and Psychological Measurement, 1978, in press.

Baldauf, R.B., Jr. and Propst, I.K. Preliminary evidence regarding the validity of a modified CLOZE procedure for lower elementary ESL students. Educational and Psychological Measurement, 1978, in press.

Buros, O.K. Tests in Print II. Highland Park, N.J.: Gryphon Press, 1974.

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental design for research on teaching. In N.L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Division of Testing and Certification, English Language Institute. Michigan Test of English Language Proficiency. Ann Arbor: University of Michigan, 1962 (Distributed by Folletts Michigan Book Store).

Gates, A.I. and MacGinitie, W.H. Gates-MacGinitie Reading Tests, New York: Teachers College Press, 1972.

Horst, D.P.; Tallmadge, G.K. & Wood, C.T. Measuring achievement gains in educational projects. Los Altos, CA: RMC Research Corporation, October 1974 (UR-243). Also A practical guide to measuring project impact on student achievement. Washington, D.C.; U.S. Government Printing Office, 1975. (017-080-01460-2).

McLaughlin, M.W. Evaluation and reform. Cambridge, Mass.: Ballinger Pub. Co., 1975.

McLeod, J. GAP Reading Comprehension Test (Forms B, R and Manual). Melbourne: Heinemann, 1967.

McLeod, J. and Anderson, J. GAPADOL Reading Comprehension Test (Forms G, Y and Manual). South Yarra, Vic: Heinemann, 1972.

Madden, R., Gardner, E.F., Rudman, H.C. and Kelly, T.L. Stanford Achievement Test. New York: Harcourt, Brace, Jovanovich, 1964.

Marco, G.L., Murphy, R.T., & Quirk, T.J. A classification of methods of using student data to assess school effectiveness, Journal of Educational Measurement, 1976, 13, 243-252.

Myers, M. Uncle Sam's a reading puppeteer: Guess who the puppet is? Learning, 1975, Nov., 20-23, 26.

Rivlin, A.M. Systematic thinking for social action. Washington, D.C.: Bookings Institute, 1971.

Science Research Associates, SRA Achievement Series (Multilevel Edition). Chicago: Science Research Associates, 1973.

Worthen, B.R. & Sanders, J.R. Educational evaluation: Theory and practice. Worthington, Ohio: Charles A. Jones, 1973.

Richard B. Baldauf, Jr. is Lecturer in Education at James Cook University of North Queensland. He has a PhD in Educational Psychology from the University of Hawaii, and has taught English-as-a-second language for three years in Sabah, Malaysia. He was Supervisor of Educational Testing for three years in American Samoa, and has been involved in Title I and Title III project evaluations in American Samoa and the Mariana Islands.