DOCUMENT RESUME

ED 093 976 TM 003 832

AUTHOR Butt, Richard L.; Wideen, Marvin F.
TITLE The Development, Validation, and Use of an Arbitrary Implementation Scale (AIS) as a Basis for Ex Post Facto Curriculum Evaluation.
PUB DATE [Apr 74]
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974) For a related document, see TM 003 870

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Correlation; *Junior High Schools; *Models; Objectives; Rating Scales; *Science Curriculum; Statistical Analysis; *Summative Evaluation; Test Validity
IDENTIFIERS *Arbitrary Implementation Scale; Canada; Saskatchewan

ABSTRACT

Current curriculum evaluation models tend to cater to desirable or ideal situations and offer little help to the evaluator of ongoing curricular innovations, many of which are characterized by little assistance or control of implementation, little monitoring or supervision of operation, and haphazard postimplementation modification or development. This paper describes the conception, operationalization, validation, and role of an Arbitrary Implementation Scale (AIS) in an ex post facto curriculum evaluation model which facilitated the provincewide assessment of an inquiry oriented junior high school science curriculum in Saskatchewan operating within the context described above. The AIS is included in the appendix of this report. (Author/RC)

# THE DEVELOPMENT, VALIDATION, AND USE OF AN ARBITRARY IMPLEMENTATION SCALE (AIS) AS A BASIS FOR *EX POST FACTO* CURRICULUM EVALUATION



Richard L. Butt
McGill University

Marvin F. Wideen
Simon Fraser University

Presented at the American
Educational Research Association Conference
Chicago: April 1974

## CONTENTS

## ABSTRACT

Current curriculum evaluation models tend to cater to desirable or ideal situations and offer little help to the evaluator of ongoing curricular innovations, many of which are characterised by little assistance or control of implementation, little monitoring or supervision of operation, and haphazard post-implementation modification or development.

This paper describes the conception, operationalisation, validation, and role of an Arbitrary Implementation Scale in an ex post facto curriculum evaluation model which facilitated the province-wide assessment of an inquiry oriented Junior High School Science Curriculum operating within the context described above.

# BACKGROUND TO THE STUDY

## Introduction of the Curriculum[1]

Following a four year curriculum development period involving writing conferences, trials, pilot projects, and revisions similar in pattern to that which Grobman (1970, p. 4) describes, a new junior high school science curriculum was introduced in Saskatchewan schools in two phases in 1968. Firstly, existing pilot projects were expanded; with in-service education being accomplished through teacher interaction and cooperation. The second phase involved province-wide half-day workships, led by teachers trained by the Province's Science Curriculum Committee during a two day workshop. Subsequent implementation, in-service education, and local adaptations were the responsibility of local superintendents and their staffs within approximately sixty school units covering the southern third of the province - the northern two-thirds of Saskatchewan with its remote settlements, fly-in centres, and native Indian population is the responsibility of central government agencies.

## The General Evaluation:

Need: In 1972 the authors prepared a research proposal for the evaluation of junior high school science in Saskatchewan and commenced discussions with the Provincial Science Education Committee. The following reasons underlined the need for the study: i) By 1970 science education had undergone two decades of curriculum reform both in Britain and North America, ii) New thinking related to science teaching was creating pressures for change, iii) The curriculum had been in operation for four years and represented a significant departure from the previous curriculum. During this time scattered local formative evaluation had been taking place with some central initiatives but

[1] The reader is referred to the general and specific objectives of the curriculum contained in the Appendix.

not of sufficient magnitude and depth to be called a continuous program of evaluation. Certainly not sufficient enough to provide an ecological picture of the degree of implementation, province-wide, nor sufficient enough to give feedback on student and other criterion measures. It was hoped that this project would provide evaluative data for decision-making and be a spur for a continuous evaluation on a province-wide basis, both central and local.

Purpose: The purpose of the general project was, therefore, five-fold:

1. To assess the degree of implementation of the program,
2. To gain some measure of its suitability and worth,
3. To determine if such factors as costs, demographic variables associated with teachers and school organisation were related to implementation,
4. To compare student outcomes for classrooms that have implemented the program with those that had not,
5. To gain an ecological[1] picture of curriculum and instruction for junior high school science in the Province of Saskatchewan.

Setting: At the time of the evaluation it was expected that most junior high school classrooms would have introduced the curriculum, although the extent of actual implementation would vary considerably due to the following factors:

i) The junior high classrooms may have been in different school settings; either elementary school (for grades 7 and 8), junior high schools, or high schools (for grade 9).

ii) Rooms and facilities would vary considerably.

iii) Implementation required expenditure for equipment which ranged from $500 to $1,000 per class.

iv) Little guaranteed assistance with implementation (in-service, supervision, consultant help, etc.); the initiative for this would be a local prerogative of superintendent, principals, and school staffs.

---

1. See Barker, R.L. Ecological Psychology, Stanford Calif.: Stanford University Press, 1968.

v) Minimal central control over implementation procedures and pattern allowing for local adaptation.

vi) Minimal monitoring or supervision of curricula in operation.

vii) Haphazard or broken-front post-implementation patterns in modification and development.

It was suspected that these factors would result in a rather mottled ecological picture of implementation throughout the Province.

Evaluation Strategy and Model: In essence, the general problem for the total project was to develop a strategy and model for evaluating this curriculum, serving the purposes stated previously, and having implementation constraints and variations described above. Limited financial resources for the study in a Province with a population sparsely distributed throughout vast areas also had to be borne in mind.

The researchers were fortunate to be able to work as independent external evaluators when it seemed appropriate but also could involve the Provincial Committee, which developed the curriculum, very intimately at appropriate points, hopefully harvesting the best of internal and external modes of evaluation. This also enhanced the chance of subsequent recommendations being accepted and implemented.

An examination of existing curriculum evaluation models via general sources (Taylor, 1972; Worthern, 1973) revealed that they were not completely appropriate in an *ex post facto* situation although parts could be utilised; the same could be said for accepted research designs (Cambell and Stanley, 1963) - as Walbesser (1968, p. 54) states:

> "There is no universal research design that can be applied uniformly to all curriculum evaluations. In fact much of the 'text-book' research design has only limited application to the design of an evaluation for an ongoing project"

Tyler's (1951) simple and basic sequence of denoting observable objectives, specifying learning experiences likely to contribute to the attainment of the objectives, followed by comparing the outcomes to the intended objectives, has been mirrored and elaborated by Provus (1968) in his systems approach and discrepancy model. Contributions by Stake (1967) with his antecedents, transactions, and outcomes and similarly by Stufflebeam (1971) with the context, input, process, and product (C.I.P.P) model as well as Scriven's (1967) formative and summative evaluation have all served to discriminate different sets of elements within the domain of curriculum evaluation. However, it is seldom that a ready built model will suit a particular situation especially when complicated by the factors mentioned previously (p. 2); furthermore, a perusal of the literature reveals that little has been written concerning appropriate means for conducting *ex post facto* evaluation[1] where no facilitative steps were taken during curriculum development and implementation; so, the researchers were required to develop a generalized model for this type of situation from which to work.

Recent models developed for curriculum evaluation mentioned above provide a base from which to work, but were also somewhat inappropriate since they were designed to follow programs through the planning, development, implementation and final evaluation stages. Hence only partial use could be made of these.

The model developed for this study utilized the notion of discrepancy between expectation and reality. It was assumed that as the program had been in effect for five years, ample time had elapsed to allow reasonably

---

[1] This is understandably so in that the *ex post facto* mode is a less desirable form of evaluation; nevertheless, situations arise where it is most useful when compared to no evaluation. Therefore we should work towards the best *ex post facto* models.

stable and observable classroom practices to emerge. Furthermore, if such practices were occuring according to reasonable expectations, then it could be hoped that "implementation" had, in fact, been achieved. The degree to which a discrepancy existed between expectations and reality represented the degree to which implementation had not been achieved. Such a notion required the development of a set of expectations and a means of observing reality to see if those expectations had been met. Figure 1, overleaf, includes the total model and flow-chart of operations.

The expectations for implementation were essentially a description of those conditions that ought to exist in classrooms where the program was in operation. These conditions should logically follow from a careful consideration of the program objectives. Hence, the first stage of the evaluation model involved (a) clarification of program objectives[1], and (b) definition of expected conditions.

The second stage of the model involved the development of observational instruments. These fell into two broad categories - student outcomes and classroom transactions and conditions.

These outcomes, transactions, and conditions provided a basis for the collection and development of the data gathering instruments which were:

1. A general questionnaire containing some eighty items designed to gather data on facilities, costs, equipment, in-service training, teaching activities, teacher background, feelings, opinions, and attitudes.

2. Student measures including an interest inventory, attitude scale, cognitive test, and the student's perception of classroom.

3. The Science Classroom Observation Form which enabled observers to rate interactions in the classroom.

4. A guide for interviewing teachers.

---

1 See Appendix.

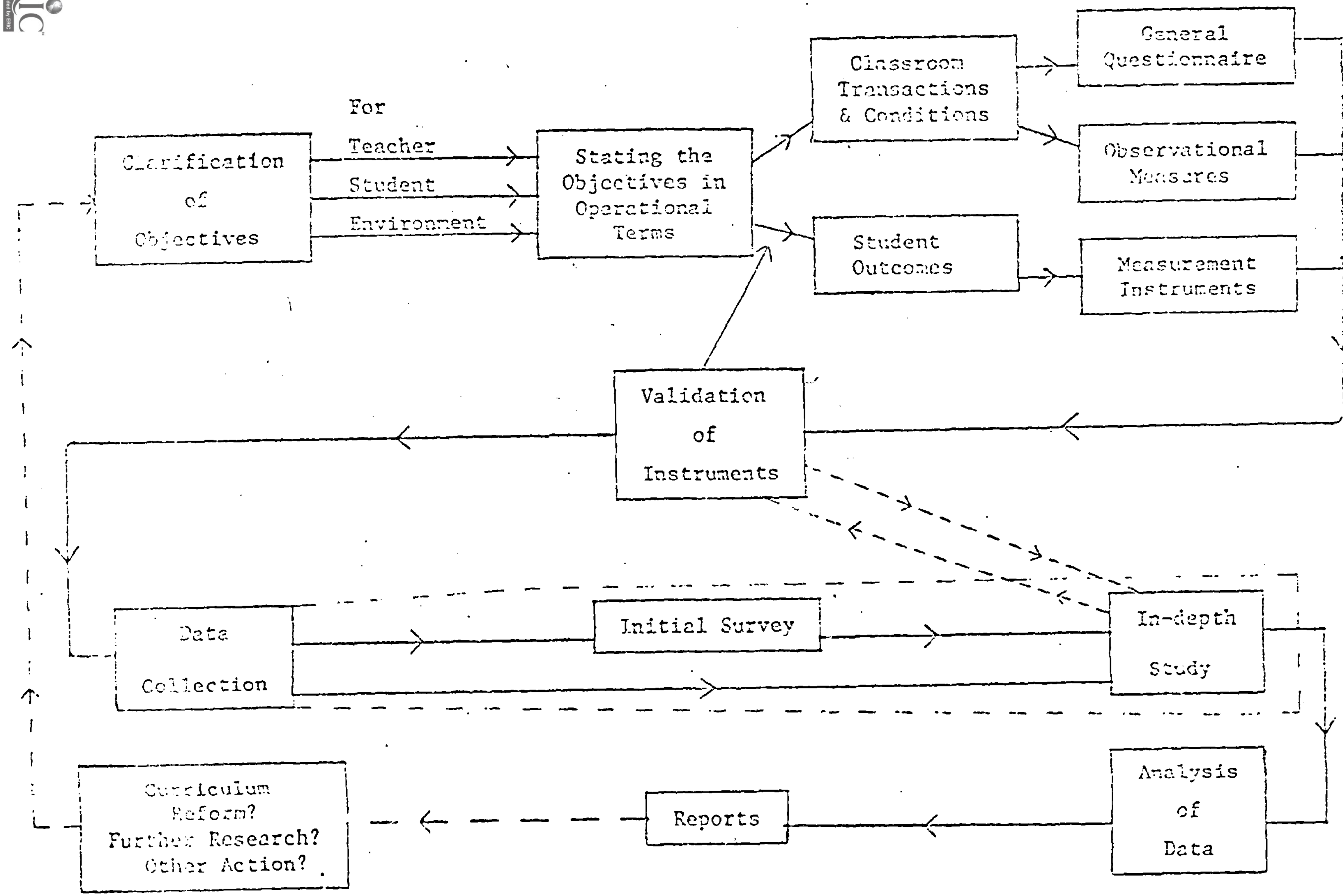



Figure I: Evaluation Model

The general strategy of the study was to utilise the general questionnaire to glean data from all the teachers of junior high school science in the Province which would serve purposes 1, 2, 3, and 5 (p. 2) of the study. The instruments mentioned in 2, 3 and 4 above for students and observers were designed for use in an in-depth study which would serve the same purposes as the general questionnaire, validating and/or qualifying its findings, but most importantly would serve to provide data for a comparison of student outcomes of classrooms which had implemented the curriculum to a large degree with those which had not (Purpose 4).

We will not take time and space here to examine the detailed development, validation and use of the above instruments as it is documented elsewhere (Butt, 1973b; Wideen, 1973, 1974) and because the specific focus of this paper is the Arbitrary Implementation Scale which, finally, we can discuss.

## THE ARBITRARY IMPLEMENTATION SCALE

Purpose: After having developed the general evaluation model, the strategy mentioned above, and made beginnings on the development of the instruments, there still remained the problem of assessing, in some quantitative way, the degree of implementation in each classroom, together with a research design problem for analysing the data from both the general questionnaire and the in-depth study.

An Arbitrary Implementation Scale (AIS) was proposed to overcome the above problems; it was hoped that we could identify a scale of items based on teachers responses from the general questionnaire (it was unthinkable to visit every classroom in the Province) which would validly indicate degree of implementation in the classrooms within which individual teachers functioned, and would therefore act as a useful "independent" variable for research design and data analysis purposes.

Development

Items from the general questionnaire, which had thus already been through the preliminary validating (face and content) procedures of the judgement of a panel of experts (the research team and the provincial committee) pilot trials, and revision, were identified as being key indicators of successful implementation of the program. The final version of the AIS consisted of 28 items which fell into five categories reflecting pervasive constructions associated with implementation. These include:

1. The opportunity and extent of in-service education (two items).
2. The knowledge of, acceptance and agreement with the philosophy, aims, and objectives of the curriculum (six items)
3. The self-perception of teaching ability for the curriculum (5 items)
4. The extent to which certain factors helped or hindered in implementation (ten items)
5. Specific practices in teaching and evaluation (four items)

These items were picked because of their ability to discriminate degrees of implementation during the validation of the general questionnaire. The AIS scale is included at the conclusion of this paper.

For scoring purpose the items were treated as part of a Likert scale and a single score assigned to each respondent. The coding of items 1 and 2 was such that they were reverse scored. The wording of items 4, 7, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, and 28 was such that they were also reverse scored. Items were differentially weighted in order that each would contribute equally to the composite score. It seemed appropriate that missing data be treated as zero in the case of items 3, 4, 5, 6, and 7 while other items omitted were assigned the overall mean.

<u>Validation</u>: The "a priori" criteria stated to further assess the validity of the AIS included the following:

1. The AIS should correlate with the teachers' perception of how well the program had been implemented.
2. The AIS scores should show significant differences between samples of classrooms using the new program and the former program.
3. The AIS scores should correlate highly with the ratings of the research team done on a random sample of classrooms.
4. The AIS scores should show significant differences between classrooms rated by a science supervisor as having good implementation or poor implementation.

Data from the general questionnaire provided information relative to the first and second criteria. The responses to items which asked teachers in the three grades to state their opinion on the degree to which the curriculum had been implemented, were correlated with scores on the AIS. The correlations were significant ($p < .01$, $n = 618$) with values of <u>.53</u>, <u>.47</u> and <u>.34</u> respectively for seventh, eight and ninth grade teachers.

The second criteria stated that the AIS should show differences between different groups in the sample. Means for different groups classified according to grade and programs taught are provided in Table I. It can be seen that differences do exist. For example, "Life Science 7" group are approximately half a standard deviation above the "Earth 8" group. This was expected since evidence pointed to a better implementation picture in grade seven than the eighth and ninth grade. Teachers of the "IPS" group, a highly laboratory oriented program, exceeded all others by nearly one standard deviation. Of most significance is the mean of the "Science Activities" (former program) which is well below all other groups.

The data provided in Table II is related to the third and fourth criteria. As indicated earlier the three member research team independently ranked a random sample of completed questionnaires according to degree of implementation. This was done on the basis of a subjective study of each questionnaire without prior consultation as to prior criteria for ranking, and without prior knowledge of AIS items. The AIS scores from these questionnaires were later determined for the upper and lower ten. The mean difference between these groups as shown in Table II is highly significant.

The data for the second comparison shown in Table II was obtained by selecting a school system where variation in program implementation existed. The superintendent was then asked to indicate classrooms which had implemented the program and those which had not. It can be seen that the difference between the two groups is significant.

A reliability of 0.78 (K.R.) was obtained for the AIS scale. On the basis of the steps taken during the development of the general questionnaire and subsequently the AIS together with the results reported in this section, the investigators felt that the AIS was a reasonably reliable and valid measure for use as an indicator of program implementation.

TABLE I

MEAN VALUES OF AIS SCORES FOR DIFFERENT GROUPS IN THE STUDY

| GROUP | N | MEAN | SD |
|---|---|---|---|
| Grades 7, 8 & 9 | 28 | 110.54 | 16.56 |
| Life science 7 | 152 | 112.28 | 16.19 |
| Earth 8 | 54 | 103.69 | 14.60 |
| Space 9 | 74 | 104.36 | 12.02 |
| Science Activities 7 & 8 (former program) | 17 | 97.33 | 12.63 |
| IPS | 27 | 121.15 | 12.89 |

TABLE II

T-TESTS BETWEEN MEANS OF SELECTED GROUPS FOR TWO

COMPARISONS REPRESENTING IMPLEMENTATION AND NON-IMPLEMENTATION

| Comparison | Group | Mean | SD | | | | |
|---|---|---|---|---|---|---|---|
| Rating of Question-naires by Research Team | Upper Ten ($X_1$) | 126.40 | 7.94 | $X_1$-$X_2$ | df | T | Prob |
| | Lower Ten ($X_2$) | 91.73 | 11.19 | 34.67 | 18 | 7.72 | .001 |
| Supervisory Rating of Classrooms | Implemented ($X_3$) | 118.38 | 11.39 | $X_3$-$X_4$ | df | T | Prob |
| | Not Imple-mented ($X_4$) | 103.88 | 14.25 | 14.50 | 19 | 2.45 | .01 |

Collection of Data: For the first survey phase of the study, the general questionnaire was distributed to every teacher involved in the teaching of junior-high school science[1] in the Province via the local superintendents. The returns were mailed directly to the researchers by the respondents. Of approximately 960 junior-high school science teachers who received the questionnaire 612 returned a completed version representing a return rate of 64%; a very high rate for this type of survey and length of questionnaire.

Subsequent to the return of the general questionnaires the second in-depth phase of the study was conducted with a stratified random sample constructed on the basis of degree of implementation as judged from the general questionnaire. The sample had to be limited to 43 classrooms involving 1165 students due to limited resources. In each classroom all students responded to Attitude, Interest, Cognitive, and Classroom Perception measures; the science classroom interactions were observed and rated using the Science Classroom Observation Form, and the teacher was interviewed regarding aspects of the curriculum, its development and implementation.

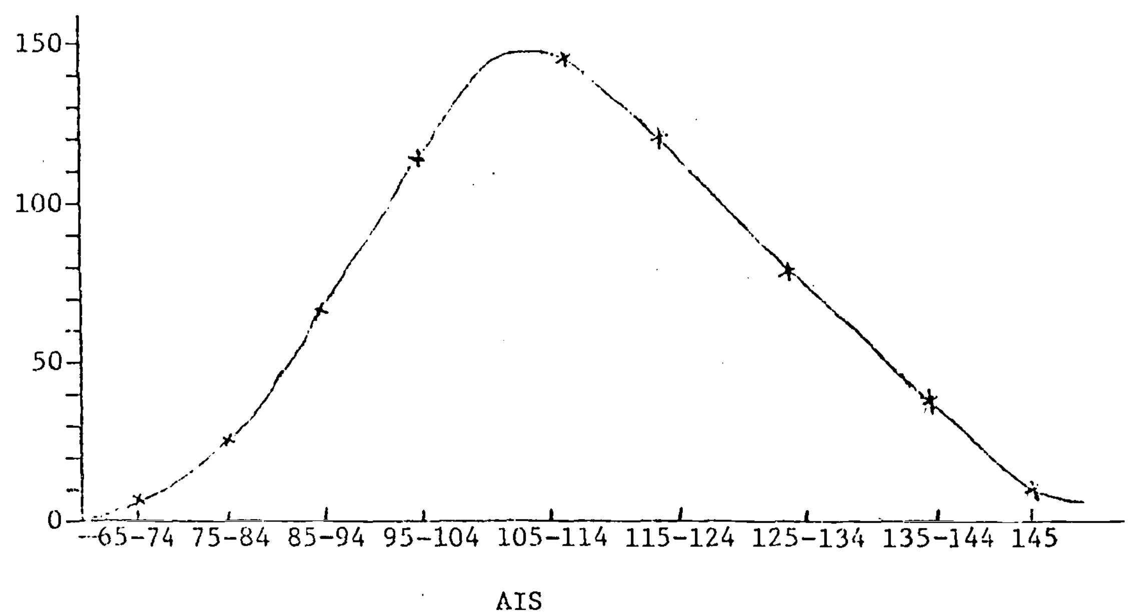
The comprehensive details of the total evaluation project are reported elsewhere (Butt, 1973b; Wideen, 1974). Here we will restrict ourselves to the data involving the AIS.

1 A compulsory subject in Saskatchewan.

## USES OF THE SCALE

Descriptive Statistics: For the respondents to the general questionnaire (n = 618) the AIS had a mean score of 108.7 with a standard deviation of 15.6; the scores ranged from 66 to 155. A frequency distribution of scores can be seen below in Figure II.

Figure II




Tables (VI, VII, VIII) of item means and standard deviations on AIS for all respondents, top scoring half and bottom scoring half splits, as well as item correlations with total AIS scores, are included in the Appendix.

An AIS as a Rating Scale for Curriculum Projects: Cyclic arguments, whose use is defended by Rozenboom (1966), would allow us to include here the results and data from the validating procedures of the AIS (p. 9 ) particularly with respect to differing degrees of implementation achieved by different courses or curriculum packages within the total junior-high school program as indicated by data in Table I. New elements of the program exceed the former curriculum in implementation scores; I.P.S. scoring highest. This suggests that a generalized implementation scale such as the AIS could serve as one rating mechanism for many of the new curriculum projects (both inter-project or intra-project).

<u>AIS and Correlation with Other Variables</u>: One objective of the study was to investigate the relationships between implementation and such factors as facilities, costs, and biographical characteristics of teachers. Table III details the correlations between the AIS and these variables. The relationships or lack of relationships and their implications are too numerous to discuss here but the data illustrate the usefulness of an AIS for teachers, curriculum developers and evaluators, as well as educational decision makers; for example, "In-Service" correlated highly with AIS (.44 $p < .01$) and while only 50% of teachers in the province had <u>the opportunity</u> to attend these series. In a sparsely populated province, Area (Urban - Rural) correlated negatively with AIS (-.26, $p < .01$) indicating the disparities between Urban and Rural areas which must be overcome. Costs, equipment, and space facilities relationships with AIS offer support for budget expenditures, in terms of implementation.

<u>AIS and Student Variables</u>: Table IV details the correlations between AIS and student and other variables. Again, these are just illustrative data and are not discussed at length; it is interesting to note significant correlations between AIS scores and both observers ratings and student perceptions. The other main use of AIS in the classroom phase of the study was as a main effect in an analysis of variance. The analysis of variance using AIS and SCOF as main effects and the student measures as dependent variables is shown in Table V. This analysis was [illegible] to previous correlational analysis but was conducted to determine if there were interactive effects between the AIS and SCOF as main effects.

It can be seen that in only one case is an F value significant. This was where the SPSC scores were dependent variables and AIS the main effect.

TABLE III

CORRELATIONS BETWEEN ARBITRARY IMPLEMENTATION SCALE AND SELECTED VARIABLES

| | Variable | Correlation with AIS |
|---|---|---|
| Teacher Characteristics | | |
| | Sex | .06 |
| | Age | .02 |
| | Area (Urban - Rural) | -.26** |
| | University Education | .22* |
| | Academic Science | .31** |
| | Teaching Experience | .07 |
| | Type of Teacher (Senior-Junior High-Elementary) | .10 |
| | In Service Sessions (If you had them (did they help?) | .44** |
| | Average Class Enrolment | .18 |
| | Number of Science Classes Taught | .24* |
| Teaching Activities | | |
| | Providing Notes | -.29** |
| | Field Trips | .23* |
| | Discussions | .10 |
| | Student Library Research | .16 |
| | Periodic Quizzes and Tests | -.12 |
| | Check Lists and Inventories | .18 |
| | Written Work (papers, etc.) | -.04 |
| | Interviews | .22* |
| | Subjective Assessment of Attitudes and Interests | .15 |
| Teachers Feelings, Perceptions, etc. | | |
| | Are you aware of Philosophy? of Program? (Yes-No) | -.22* |
| | Teacher Perception of Degree of Implementation of: | |
| | Grade 7 ) Validation | .53** |
| | Grade 8 ) Criteria of | .47** |
| | Grade 9 ) AIS | .35** |

Continued....

Continued ......

| | | |
|---|---|---|
| Helps in Implementation | | |
| | Availability of Equipment | .48** |
| | Adequate Reference and Library Resources | .36** |
| | Administrative Support | .41** |
| | Favourable Student Reaction | .40** |
| | School Board Support | .37** |
| Parents and Students Reactions | | |
| | Since Implementation Student Interest has (Decreased ---- Increased) | .37** |
| | Impact of Curriculum on Parents (Unfavourable ---- Favourable) | .33** |
| Facilities, Equipment, Costs | | |
| | In What T·pe of Room Do You Mainly Teach? (Unmodified Classroom --- Fully Equipped Lab) | .37** |
| | Costs of Equipment Available | |
| | Grade 7 | 0.51** |
| | Grade 8 | 0.49** |
| | Grade 9 | 0.40** |
| Teacher Summarising Opinions of Particular Courses | | |
| | Grade 7 (dislike ---- like) | 0.31** |
| | Grade 8 | 0.20* |
| | Grade 9 | 0.27** |

* $p < .05$
** $p < .01$

TABLE IV

CORRELATIONS BETWEEN AIS AND STUDENT MEASURES,

AND THE SCIENCE CLASSROOM OBSERVATION FORM

| | | |
|---|---|---|
| Student Interest Scale | | .01 |
| Student Perceived Science Classroom | | .37*** |
| Moore-Sutman Science Attitude Scale | | .22 |
| Cognotive Measure | | -.03 |
| Grade | | -.10 |
| Science Classroom Observation Form | | |
| Environment | These are tentative Subscales | .27* |
| Pupil | | .23 |
| Teacher | | -.35** |
| Total | | .24* |

*** $p < .01$ )
** $p < .05$ ) Utilising the classroom as the experimental unit
* $p < .10$ )

= $p < .01$ )
- $p < .05$ ) Utilising the students as the experimental unit.

TABLE V

ANALYSIS OF VARIANCE FOR SCOF TOTAL AND AIS SCORES ON CLASSROOM AVERAGES

FOR II, SPSC, ATT AND COG MEASURES

| Source of Variance | df | II | | SPSC | | ATT | | COG | |
|---|---|---|---|---|---|---|---|---|---|
| | | ms | f | ms | f | ms | f | ms | f |
| AIS | 1 | 4.50 | .02 | 140.69 | 4.03* | 55.00 | 1.09 | .01 | .00 |
| SCOF | 1 | 23.50 | 1.52 | 87.25 | 2.47 | 3.00 | .06 | 3.46 | 1.88 |
| AIS x SCOF | 1 | 20.06 | 1.33 | 34.44 | .70 | 1.00 | .02 | 3.22 | 1.75 |
| Error (df) | | 15.46 (26) | | 34.91 (26( | | 50.23 (26) | | 1.84 (26) | |

* Significant at the .10 level

An examination of the cell means indicated an elevation of scores on the SPSC for the top scoring half AIS group; students felt better able to utilise materials, inquiry discovered in high scoring AIS classrooms. No other F values were significant indicating no effects for other measures and no interactive effects. Further analyses were run utilising AIS and grade; no differences were revealed in these analyses that had not been revealed previously. It is interesting that in this analysis the Students Perceived Science Classroom emerged as being the only criterion measure significantly related to AIS in view of Anderson's (1969, 1970) work on climate (see also Butt, 1973a) learning, and curriculum. To find even the one relationship was surprising to the researchers, bearing in mind the possible confounding factors in the mottled program patterns, especially when only a mid-line split in AIS scores could be utilised for analysis purposes. Student interest, attitude, and cognitive abilities were not detected as being significantly different for high and low AIS scores.

Alternative interpretations for these results include:-

i) As just stated, the "treatments" perceived as different by the students have no significant effect on other student outcomes (Cognitive, Interest, Attitude).

ii) The "treatments" are not sufficiently different to have a detectable effect on student interest, attitude, and cognotive outcomes.

iii) The AIS scale does not sufficiently discriminate implementation for the purposes of this analysis.

iv) Confounding variables present obscure "treatment" effects.

The results suggest that program implementation had been successful in terms of affecting students perceptions and observers ratings, but as yet has had little effect on other attributes measured in the study.

Further refining of the scale, selection of classrooms from the two extremes of AIS scoring range, a larger sample of classrooms, with separate replications for each of grades seven, eight, and nine, should guard against ii, iii, and iv above for future use of the scale.

Other Uses of AIS: Throughout this report, data generated utilising the AIS as an independent variable has served to build up an ecological picture of the current state of junior-high school science education in the Province of Saskatchewan. One may even build a visual map of the curriculum ecology of the Province utilising AIS or other criteria, drawing contours (if they emerge) on the basis of iso-implements (cf. isobars or iso-therms).

It has been possible to construct profiles of variables where the curriculum has and has not been implemented; it is anticipated that these, and the AIS in general, may be utilised as a backdrop and stimulant for local evaluation, curriculum improvement, and development.

## SUMMARY, DISCUSSION AND CONCLUSIONS

Current models in the field of curriculum evaluation have tended towards the ideal situation in terms of one or more of the following: the objectives of the evaluation; personnel available to conduct the evaluation; the scope of the evaluation; and the resources available.

While these conceptually and practically desirable constructs may act as useful inputs prior to curriculum development and innovation, for providing a sound and comprehensive evaluation model in school systems which have the requisite resources and personnel, there exist many ongoing innovative efforts which require directional decision-making on the bais of some sort of evaluation, for which these models offer very little pragmatic help.

Key factors which characterise these curricular innovations, whether by default or deliberate design, are:

i) Little assistance with implementation

ii) Little control of implementation

iii) Little monitoring or supervision of curricula in operation

iv) Haphazard post-implementation modification and development

These factors can and have resulted in rather mottled ecological pictures throughout a large area, province, or state, with regard to varieties and degrees of implementation. This presents the evaluator with a difficult task when existing models of evaluation and desirable research strategies are considered. This is usually compounded (especially latterly) by lack of funds and the like, which bar controlled longitudinal studies of sufficient classrooms or sufficient in-depth examination of a cross-section of classrooms.

The study from which the substance of this paper was drawn involved an extensive curriculum evaluation project characterised by many of the above factors. Six hundred and fifty teachers reported on factors pertaining to an innovative inquiry oriented Junior High School Science Curriculum, while 43 classrooms and 1165 students were observed and tested in an in-depth study of the curriculum in operation and student outcomes on cognitive, perception, attitude, and interest measures.

The objective of the paper was to describe the conception, operationalisation, validation, and role of an Arbitrary Implementation Scale in an ex post facto curriculum evaluation model which facilitated the assessment of factors related to this curriculum and situation. It was concluded that the use of an implementation scale can enhance an evaluative study. This has been underlined by Charters and Jones (1973), subsequent to the conclusion of this study, who say that while it is becoming standard practice to utilise considerable resources for conducting evaluation of student outcomes for "experimental" and "control" schools or classrooms, it "is not the standard practice in evaluation studies to describe, let alone measure, how the program in "experimental" and "control" situations actually differ from one another - or even certify that they do!" Accordingly, they specify four levels of reality that may exist: firstly, Institutional Commitment, that is the formal announcement or introduction of the "innovation" by the administration; secondly, the Structural Context level, which includes the changes in formal arrangements and physical conditions (e.g. making the necessary equipment and facilities available); Role Performance, the third level of implementation, involves the actual necessary behavioral changes in teachers; while the fourth level, Learning Activities, involves the intended classroom transactions, which, hopefully, will enable students to reach the intended learning outcomes of the curriculum.

This evaluative study assumed that Level I had been attained by the formal announcement of the Provincial Department of Education (1968), but, of course this does not necessarily mean any significant change in classroom transactions! The scale, therefore, included elements of Structural Context, Role Performance, and Learning Activities. Some items which could fall under these categories were included in data collection but excluded from the actual AIS scale so that the process of implementation could be examined during data analysis. In essence, then, the AIS attempted to quantify on a continuous scale, as opposed to the four level approach of Charters and Jones, the degree of implementation of an innovation within individual classrooms on a Province-wide basis.

The results of this attempt to develop and utilise a valid AIS, which minimises the possibility of what Charters and Jones call "appraising a non-event", are encouraging.

Undoubtedly, this initial effort in the utilisation of an AIS has not exhausted all of its potential uses, nor, indeed displayed exemplary development and validation procedures. Further research is needed within the realm of implementation of innovations to assist in identifying items for less crude AIS scales.

While it is realised that efforts must be made to provide inbuilt evaluation procedures at the outset of curriculum development and innovation, it is hoped that perhaps the procedure used in this study might provide a basis for other pragmatic assessments of curricula operating in similar situations of uncoordinated change, whether as the result of poorly coordinated central initiative or widespread grass-roots initiatives.

REFERENCES

Anderson, G.J., "Effects of Classroom Social Climate on Individual Learning", American Educational Research Journal, 1970, 7, 135-152.

Anderson, G.J., Walberg, H.J., and Welch, W.W., "Curriculum Effects on the Social Climate of Learning: A New Representation of Discriminant Functions", American Educational Research Journal, 1969, 6, 315-328.

Butt, Richard L., "A Link Between Administration and Creativity?", McGill Journal of Education, Vol. VIII, No. 1, Spring, 1973, 95-98 (a).

Butt, Richard L. and Wideen, Marvin F., "Division III Science: Some Initial Findings of an Evaluative Study", Saskatchewan Journal of Educational Research and Development, Vol. 3, No. 2, Spring, 1973, 4-13 (b).

Cambell, D.T. and Stanley, J.C., Experimental and Quasi-Experimental Design for Research, Rand-McNally, 1963.

Charters, W.W. and Jones, J.E., "On the Risk of Appraising Non-Events in Program Evaluation", Educational Researcher, Vol. 2, No. 11, Nov. 1973, 5-7.

Cronbach, L.J., "The Logic of Experiments on Discovery". In: L.S. Shulman and E.R. Keislar (Eds.), Learning by Discovery: A Critical Appraisal, Chicago: Rand-McNally, 1966, 77 & 84.

Doll, R.C., Curriculum Improvement (2nd Ed.), Allyn and Bacon Inc., Boston, 1970, 21.

Grobman, H., Evaluation Activities of Curriculum Projects, Vol. 2, A.E.R.A. Monograph Series on Curriculum Evaluation, Rand-McNally: Chicago, 1970, 4.

Grobman, H., "Curriculum Development and Evaluation", Journal of Educational Research, July 1971, 64: 436-42.

Hurd, ., Address to National Association of Biology Teachers on October 24, 1970, Denver, Colorado. - Hurd proposed that the inquiry techniques developed in the 1960's are not sufficient for problems facing students in the 1970's.

Johnson, M., "Definitions and Models in Curriculum Theory", Educational Theory, April 1967, 17: 127-40.

McLuhan, M., Understanding Media, McGraw-Hill, 1966.

Provus, M., "The Discrepancy Evaluation Model" in Readings in Curriculum Evaluation by P.M. Taylor and D.M. Cowley, 117-127.

Rozenboom, William W., Foundations of the Theory of Prediction, Dorsey Press, Harewood, Ill., 1966, 213.

Samples, R.,"Science - A Human Enterprise", The Science Teacher, 1972, 39: 26-29.

Scriven, M., "The Methodology of Evaluation" in Perspectives of Curriculum Evaluation, Ed: R. Tyler, R. Gagné, and M. Scriven. A.E.R.A. Series on Curriculum Evaluation, No. 1, Chicago: Rand-McNally, 1967.

Shulman, Lee S., "Reconstructional of Educational Research", Review of Educational Research, Vol. 40, No. 3, 383.

Stake, Robert E., "The Countenance of Educational Evaluation", Teachers College Record, April 1967, 68: 523-540.

Taylor, P.A. and Comley, D.M., Readings in Curriculum Evaluation, W.C. Brown, 1972.

Tyler, Ralph W., "The Functions of Measurement in Improving Instruction", In: Educational Measurement, Ed. E.F. Lunquist, Washington, D.C.: American Council on Education, 1951..

Walbesser, H.H. and Carter, H., "Some Methodological Considerations of Curriculum Evaluation Research", Educational Leadership, Oct. 1968, 26: 53-64.

Wideen, Marvin F. and Butt, Richard L., "The Implementation and Use of a Science Program for Grades 7, 8 and 9 in the Province of Saskatchewan: An Evaluative Study", Research Monograph, PART I, S.S.T.A. Avord Towers, Regina, Saskatchewan, 1973.

Wideen, M.F. and Butt, R.L., "An In-Depth Study of Division III Science Education", Saskatchewan Journal of Educational Research and Development, (Forthcoming) 1974. (Research Report also available from S.S.T.A. Avord Tower, Regina, Sask.).

Worthen, Blain R., and Saunders, James R., Educational Evaluation: Theory and Practice, Jones Publishing, 1973.

APPENDIX

ITEMS USED IN THE ARBITRARY IMPLEMENTATION SCALE

1. Is science the subject you feel most capable of teaching? Yes ☐ No ☐

2. Did you have the opportunity to attend in-service session(s) on Division III Science? (Special workshop, institute, convention) N.B. Do not include University credit classes. Yes ☐ No ☐

3. If 'yes', how many? ____________

Estimate the amount of class time devoted to each of the following types of activities during the course of a year. (Note: Times do not need to add up to 100%.)

| | 0% | 1-10% | 11-20% | 21-30% | 31-40% | 41-50% | 51-60% | 61-70% | 71 and more |
|---|---|---|---|---|---|---|---|---|---|
| 4. (a) Lecturing and demonstrating | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. (b) Students conducting investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

Indicate the emphasis you give the following methods of evaluation in arriving at a final mark in Division III Sciences. (Percentages do not need to add up to 100%.)

| | 0% | 1-10% | 11-20% | 21-30% | 31-40% | 41-50% | 51-60% | 61-70% | 71 and more |
|---|---|---|---|---|---|---|---|---|---|
| 6. Student lab and project work. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Formal examinations | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

8 Indicate your level of agreement with the philosophy of Division III Science.

| Low 1 | 2 | 3 | 4 | High 5 |
|---|---|---|---|---|

Because of various factors (such as pre-service and in-service education, facilities, etc.) I feel that for the grades I teach my levels of competence in the following aspects of the Division III Science Program are as follows.

| | Low | | | | High |
|---|---|---|---|---|---|
| 9. Understanding the Division III Science philosophy and objectives .................. | 1 | 2 | 3 | 4 | 5 |
| 10. Knowledge of the content .................... | 1 | 2 | 3 | 4 | 5 |
| 11. Knowledge of materials and equipment required | 1 | 2 | 3 | 4 | 5 |
| 12. Ability to teach the program ................ | 1 | 2 | 3 | 4 | 5 |

Rate the following in terms of their usefulness in helping you to implement the Division III Program for the grade(s) you teach.

| | | Not at all | | Somewhat | | A Great Deal |
|---|---|---|---|---|---|---|
| 13. | Good space facilities .................... | 1 | 2 | 3 | 4 | 5 |
| 14. | Administrative support .................... | 1 | 2 | 3 | 4 | 5 |
| 15. | Belief in the philosophy of the program ... | 1 | 2 | 3 | 4 | 5 |

Rate the following factors in terms of how much they have hindered successful implementation of Division III Science for the grade(s) you teach.

| | | Not at all | | Somewhat | | A Great Deal |
|---|---|---|---|---|---|---|
| 16. | Lack of equipment ........................ | 1 | 2 | 3 | 4 | 5 |
| 17. | Inadequate reference and library resources. | 1 | 2 | 3 | 4 | 5 |
| 18. | Inadequate pre-service education .......... | 1 | 2 | 3 | 4 | 5 |
| 19. | Poor space facilities ..................... | 1 | 2 | 3 | 4 | 5 |
| 20. | Administrative constraints ................ | 1 | 2 | 3 | 4 | 5 |
| 21. | Adverse student reaction . ................ | 1 | 2 | 3 | 4 | 5 |
| 22. | Lack of school board support .............. | 1 | 2 | 3 | 4 | 5 |
| 23. | Lack of belief in philosophy of program. ... | 1 | 2 | 3 | 4 | 5 |

Give your opinion on the following statements with reference to the Division III Science Program for grades 7, 8, and 9.

| | | Strongly Disagree | | Undecided | | Strongly Agree |
|---|---|---|---|---|---|---|
| 24. | The approach used in the Division III Science Programs for grades 7, 8, and 9 is much the same as traditional science teaching .................................. | 1 | 2 | 3 | 4 | 5 |
| 25. | A teacher in each grade should stay very close to the subject matter for that grade .................................... | 1 | 2 | 3 | 4 | 5 |
| 26. | One major emphasis in the program is to use major ideas in relating different parts of the course. ...................... | 1 | 2 | 3 | 4 | 5 |
| 27. | The evaluation of students must be based largely on cognitive outcomes .............. | 1 | 2 | 3 | 4 | 5 |
| 28. | Generally, the Division III Science Program has had little effect in changing my teaching .................................. | 1 | 2 | 3 | 4 | 5 |

## The Curriculum

The general aspects of the program were to include the following characteristics:

1. The program should have a unitary flavour for the three grades and provide a three-year cycle. It was hoped that the grade 9 year would provide a capping effect upon the previous two years and that a student would touch upon the main content areas of life, earth and space or physical[1] science during his grade 7, 8 and 9 school career.

2. The program should be built around conceptual and behavioral schemes in science. Recent thinking in science education at that time seemed to suggest that a more fruitful way of organizing curricula was to utilize major ideas in science rather than subject matter content. In addition, certain behavioral themes were identified which would serve to give coherence to student experience. These themes were to provide an overall structure to which the three content areas would contribute. Specific concepts within each subject area were identified for possible study at the three grade levels.

3. The science program should emphasize student involvement with materials and the investigations should provide for an inquiry approach to learning. The implications of this facet of the program were that equipment and materials should be available. It was noted that the inquiry and investigation approach would require a major change in role for many teachers.

Following a discussion of the research proposal with original and current members of the Provincial Science Committee that developed the curriculum, they attended to the task of clarifying the objectives.[2]

---

1 Alternate approaches to grade 9 science were permitted (Space Science or Introductory Physical Science) as both were consistent with the philosophy of the junior high school science program and a number of teachers had prefered IPS during curriculum development.

2 The "clarification of objectives" detailed here is properly part of the evaluation model and research procedures discussed later, but is appropriate for inclusion here.

1. To develop in the student an appreciation for, and an interest in science.

2. To challenge the student to think and reason through scientific study.

3. To foster a spirit of inquiry.

4. To develop an understanding and appreciation of the methods by which scientists discover new knowledge.

5. To develop a deeper insight into the interrelationships that exist in science.

6. To develop and broaden the student's understanding of the big ideas of science.

7. To help the student acquire some of the skills and processes of science.

It was believed that these objectives could be achieved only in classrooms meeting certain minimum requirements. These included:

1. Adequate facilities and the addition of science equipment where needed.

2. The addition of new resource books and teaching materials.

3. A change from traditional classroom transactions. The teacher's role was to stimulate and challenge the student to think and to provide an atmosphere of freedom of operation. He was to provide support for inquiry and manipulation and investigation of materials, giving the student the opportunity to pursue problems and exercise some autonomy in interpreting results. Compared to the more traditional directive and expository role, a teacher was intended to be open, non-directive and to act as a resource person and guide to learning.

4. In-service education to assist teachers to comprehend and implement the programme.

TABLE VI

MEANS AND STANDARD DEVIATIONS

OF AIS ITEMS

FOR ALL RESPONDENTS

| ITEM | MEAN | STANDARD DEVIATION | N |
|---|---|---|---|
| 1 | 1.48 | 0.50 | 603 |
| 2 | 1.50 | 0.50 | 585 |
| 3 | 2.26 | 1.41 | 606 |
| 4 | 3.93 | 1.56 | 601 |
| 5 | 4.04 | 1.76 | 556 |
| 6 | 3.97 | 1.57 | 567 |
| 7 | 3.82 | 2.05 | 530 |
| 8 | 3.50 | 0.92 | 536 |
| 9 | 3.27 | 1.08 | 599 |
| 10 | 3.71 | 0.92 | 603 |
| 11 | 3.43 | 1.02 | 604 |
| 12 | 3.39 | 0.87 | 599 |
| 13 | 2.67 | 1.24 | 571 |
| 14 | 2.92 | 1.28 | 584 |
| 15 | 3.53 | 1.00 | 567 |
| 16 | 3.03 | 1.25 | 587 |
| 17 | 2.97 | 1.22 | 590 |
| 18 | 2.97 | 1.22 | 572 |
| 19 | 3.00 | 1.30 | 580 |
| 20 | 1.76 | 0.97 | 569 |
| 21 | 2.08 | 1.02 | 583 |

Continued...

Continued......

| | | | |
|---|---|---|---|
| 22 | 1.90 | 1.02 | 563 |
| 23 | 1.96 | 1.00 | 569 |
| 24 | 2.01 | 0.98 | 585 |
| 25 | 2.22 | 1.16 | 590 |
| 26 | 3.77 | 1.01 | 578 |
| 27 | 2.95 | 1.01 | 568 |
| 28 | 2.46 | 1.12 | 568 |

TABLE VII

MEAN AND STANDARD DEVIATIONS OF AIS ITEMS

FOR TOP AND BOTTOM SCORING HALVES OF RESPONDENTS

| ITEM | TOP | | | BOTTOM | | |
|---|---|---|---|---|---|---|
| | MEAN | S.D | N | MEAN | S.D | N |
| 1 | 1.26 | 0.44 | 292 | 1.70 | 0.46 | 293 |
| 2 | 1.29 | 0.46 | 305 | 1.71 | 0.45 | 301 |
| 3 | 2.39 | 1.47 | 211 | 1.93 | 1.20 | 83 |
| 4 | 3.45 | 1.62 | 303 | 4.42 | 1.67 | 298 |
| 5 | 4.78 | 1.78 | 290 | 3.24 | 1.33 | 266 |
| 6 | 4.52 | 1.64 | 303 | 3.34 | 1.22 | 264 |
| 7 | 3.15 | 1.77 | 275 | 4.53 | 2.09 | 255 |
| 8 | 3.82 | 0.79 | 280 | 3.15 | 0.92 | 256 |
| 9 | 3.75 | 0.90 | 303 | 2.78 | 1.02 | 296 |
| 10 | 3.99 | 0.83 | 305 | 3.42 | 0.93 | 298 |
| 11 | 3.87 | 0.86 | 305 | 2.99 | 0.99 | 299 |
| 12 | 3.71 | 0.78 | 302 | 3.07 | 0.85 | 297 |
| 13 | 3.19 | 1.17 | 288 | 2.14 | 1.08 | 283 |
| 14 | 3.35 | 1.21 | 298 | 2.48 | 1.20 | 286 |
| 15 | 3.93 | 0.90 | 292 | 3.12 | 0.93 | 275 |
| 16 | 2.59 | 1.21 | 296 | 3.47 | 1.13 | 291 |
| 17 | 2.62 | 1.19 | 297 | 3.33 | 1.15 | 293 |
| 18 | 2.43 | 1.16 | 286 | 3.50 | 1.04 | 286 |
| 19 | 2.60 | 1.28 | 290 | 3.41 | 1.20 | 290 |
| 20 | 1.56 | 0.87 | 289 | 1.95 | 1.03 | 280 |

Continued.....

Contin·ed.....

| | | | | | | |
|---|---|---|---|---|---|---|
| 21 | 1.78 | 0.83 | 292 | 2.38 | 1.10 | 291 |
| 22 | 1.66 | 0.92 | 283 | 2.14 | 1.05 | 280 |
| 23 | 1.58 | 0.85 | 286 | 2.35 | 1.00 | 283 |
| 24 | 1.65 | 0.82 | 294 | 2.37 | 0.99 | 291 |
| 25 | 2.06 | 1.09 | 299 | 2.38 | 1.21 | 291 |
| 26 | 3.87 | 1.00 | 290 | 3.67 | 1.02 | 288 |
| 27 | 2.79 | 1.07 | 286 | 3.12 | 0.93 | 282 |
| 28 | 2.10 | 1.08 | 286 | 2.84 | 1.03 | 282 |

TABLE VIII

CORRELATION BETWEEN AIS ITEMS AND TOTAL AIS SCORES

| ITEM | CORRELATION |
|---|---|
| 1. Is science the subject you feel most capable of teaching | -.50 |
| 2. Opportunity to attend in-service sessions? | -.47 |
| 3. How many? | .22 |
| 4. Estimate time spent lecturing and demonstrating | .37 |
| 5. Estimate time spent when students were conducting investigations | .56 |
| 6. How much does student lab and project work feature in final evaluation | .43 |
| 7. How much does formal examination and project work feature in final evaluation | -.43 |
| 8. Level of agreement with philosophy of curriculum | .47 |
| 9. Understanding of philosophy and objectives | .56 |
| 10. Knowledge of content | .42 |
| 11. Knowledge of materials and equipment required | .52 |
| 12. Ability to teach program | .47 |
| Ratings of usefulness in implementation: | |
| 13. Good space facilities | .51 |
| 14. Administrative support | .42 |
| 15. Belief in philosophy of program | .51 |
| Ratings of hindrance in implementation: | |
| 16. Lack of equipment | -.46 |
| 17 Inadequate reference and library resources | -.38 |
| 18. Inadequate preservice education | -.55 |

Continued.....

Continued.....

| | | |
|---|---|---|
| 19. | Poor space facilities | -.39 |
| 20. | Administrative constraints | -.26 |
| 21. | Adverse student reaction | -.38 |
| 22. | Lack of school board support | -.27 |
| 23. | Lack of belief in philosophy of program | -.47 |
| Opinions: | | |
| 24. | New program same as traditional program? | -.43 |
| 25. | Teacher should stick closely to subject matter? | -.18 |
| 26. | Major emphasis is to use "big" ideas? | -.14 |
| 27. | Evaluation based largely on cognetive outcomes? | -.19 |
| 28. | New program has had little effect in changing my teaching | -.35 |