Florida Journal of Educational Research
Fall 1984, Vol. 26, No. 1, Pp. 21-27

Local Development of Subject Area Item Banks

Annie W. Ward
Ward Educational Consulting, Inc.

Gene Barlow
Dade County Public Schools

ABSTRACT. It is feasible for school districts to develop and use subject area tests as reliable as those previously available only from commercial publishers. This paper describes procedures used in a large school district to develop item banks in algebra, life science, career education, geometry, biology, and U. S. history. Districts who put a reasonable amount of effort and resources into such projects can develop tests valid for assessing specific courses.

With recent advances in computer technology and the development of latent trait test theory, it is feasible for even relatively small local school districts to develop and use custom-designed tests which are as reliable as those previously available only from commercial publishers. The use of local curriculum guides and local teachers as item writers is making it easier to attain curriculum validity of the items and tests developed using them. Furthermore, fieldtesting of the items can be done by random sampling of the total set. However, the great potential of such an undertaking should not blind one to certain pitfalls.

This paper describes three projects in local item development in a large school district. It was written with the thought that our experience could help other school systems decide whether to become involved in such projects and, if they do, to provide some suggestions which may help to make the experience successful.

Approximately two years ago, the district made a commitment to begin developing banks of test items for secondary subject areas such as algebra I, life science, biology, geometry, and United States history. This commitment, supported by the Superintendent and the Board, stemmed from the perception that neither the state minimum performance tests nor the district wide norm-referenced tests provide an adequate basis for insuring a quality program in secondary level, non-basic-skill areas. No matter how appropriate the basic skills and commercial survey tests may be for their intended purposes, they have very limited utility for assessing the quality of curriculum, instruction, or learning in specific content areas.

Initially, the program was conceived as a test development project. However, early in the planning it was refocused to item development in order to provide more flexibility both in item development and in use. The intent was to establish an item bank in each subject area that could be used to produce an infinite number of alternate forms of course exams. In addition, it was decided to develop item banks, sufficiently large that half of the items could be made available to classroom teachers for use in preparing their classroom examinations. The latter decision is recommended for two reasons: First, it is likely to improve the quality of classroom examinations and, second, teachers who work on the items probably take a more personal interest in the items when they know they will be able to use them in their own classroom.

Procedures

Three projects have been initiated thus far. The first involved only algebra I. The second involved life science and career education, a specially funded elementary program. The third project involved three subjects: geometry, biology, and U. S. history.

The first project, focused on algebra I, was conducted during the summer of 1982. Five algebra teachers were employed for a six week period during which they received training, developed a blue print, completed item specifications, and wrote approximately 500 test items. A beginning was also made toward

writing items for algebra II. The school system identified a person to be responsible for fieldtesting and for developing and maintaining item banks.

These items were reviewed by both content and measurement specialists and then sorted onto multiple forms containing approximately 40 items each. During the following year, each test form was field tested on a sample of about 250 students. The various forms were spiraled within classrooms in order to provide a random sampling within the field test sample.

The items were then analyzed and linked using Rasch equating procedures, and the item data were posted to the item files in the item bank. Items which performed acceptably on the fieldtest were then divided into two groups. One group was printed and disseminated to teachers along with the item specifications and fieldtest statistics. The second group of items was retained in the testing department as a secure bank for future use.

The remaining two projects were conducted during the school year. The training period was shortened to two days at the outset of the projects, and 21 teachers involved in the third project, geometry, biology and U. S. history, were all trained simultanteouly regardless of content matter area. Two school system measurement people provided guidance, review, and feedback. The training sessions covered the writing of both specifications and items. The purpose, components, and characteristics of adequate specifications were described and illustrated. Samples of specifications, varying from inadequate to good, were presented and discussed. Following this, guidelines for item writing along with good and bad example items were presented. A training manual was used in conjunction with the discussions.

The local curriculum objectives were inspected in order to set priorities and to identify those objectives that were covered in more than one section of the outline. Many skills were found to be subsumed under others, and some topics were judged to be of greater importance than others in terms of curricular emphasis. Supervisors of the various content areas were involved in these discussions and textbooks were frequently consulted.

Once priorities were set, teachers were assigned a group of objectives for which they were to write specifications and one sample item for each specification. As the specifications were completed, they were exchanged among teachers in the group for review which included writing one additional item for each specification. A schedule for review by participating teachers was prepared and posted. Reviewers consulted with the original writer about points of disagreement and if consensus was not reached, the entire group was asked to consider the matter and resolve it. At the end of the two-day training period, all specifications had been drafted with at least one item for each specification, and all specifications had been reviewed by at least one person other than the original author.

Teachers were given contracts to complete the task of clarifying the specifications and writing representative items. The products were submitted to the district office by a specified date in order to be entered into the word processor and printed for review. Each subject area group produced approximately 500 items. The teacher/writers re-assembled on the specified date at the central office to review the specifications and items. At this time agreement was reached on the final form of the specifications and items.

Results

Reviewing the specifications and items produced by the teachers, it was discovered that the process used did not guarantee items of acceptable quality, and considerable additional editing was required by three groups: (a) testing and curriculum consultants, (b) persons sensitive to cultural and sex bias, and (c) individuals who have a firm foundation in the mechanics of clear, concise writing.

After an additional editing process, the items for all subject areas in the second two projects were fieldtested and are in the process of additional revision.

In the spring of 1984 items for algebra I and life science underwent a second field testing. Two 36-item forms for each course were prepared, each con-

taining some common items plus some unique ones. The tests were administered to all students enrolled in relevant courses, with the two forms spiraled within each classroom. Preliminary analysis of these data indicate that all of the items were of satisfactory quality. There were no negative biserial correlations on either form of either examination. Reliabilities were .79 for each form of the life science test and they were .86 and .87 for the two algebra tests.

The projects reported have met the original intent to create a large bank of items in specified content areas in a short period of time at minimal cost. For those item banks that have been fieldtested, the data analyses indicate that the item pools are of acceptable quality. In fact, the data on reliability and discrimination are equal to or better than those of most commercially available tests.

Recommendations

These projects have demonstrated that it is possible for local school systems to develop item banks which can be used to improve educational assessment. In order for this procedure to work, however, certain requirements must be met. They are the following:

1. The project should be conceptualized as item development rather than test development.
2. Before starting the item development process, the local curriculum objectives must be reviewed. It is essential that some thought be given to the relative importance of the topics as well as the overlap and interaction among them. Objectives do not have to be stated in behavioral terms at this point since the process of developing specifications serves to operationalize the desired behavior. The item specifications should include clarifying/descriptive statements for each objective.
3. Item specifications should be developed before the items are written. The specifications provide an operational definition of the skill and serve as a guide to item writers. This is especially important when items are being developed for an item bank which will be updated periodically, since future item writers will write items which measure the same skill.

4. Teachers who are to make up the item writing team must be carefully selected. There are at least two types of competencies they must bring to the process. First, they must have thorough knowledge of their subject and, second, they should possess prerequisite skills for writing items. Use of local teachers and a training program does not automatically produce good test items.

Professional item writers, provided with textbooks and teachers' manuals might do a better job. However, if items are not developed by local teachers, the relevancy to the local curriculum must be established and this process can be difficult and costly.

5. A training program for teachers is necessary but not sufficient to insure high quality items. The teacher-writers need early and continuing review and feedback during the entire writing process.

6. All items must have a final review and editing by persons not previously involved in the project. The review should either be done by someone trained in both the discipline and in measurement or by two people who have these competencies working together.

7. All items should be fieldtested.

Conclusions

The successful operation of a local test development project requires involvement of a team of individuals who have either subject matter, measurement, or writing expertise as well as those who possess a variety of other technical skills. The technical skills needed include word processing, graphics, managing data files, and performing technical analysis of item data.

The importance of adequate resources cannot be overemphasized. There are many tasks which must still be performed after items are written but before they can be effectively used in a testing program. For example, items must be typed, artwork added, and then prepared in camera-ready form for printing; fieldtests must be designed and conducted; data must be run and interpreted; poor items must be flagged, studied, revised, or discarded; an item file must be developed so that items can be selected to fit a prespecified blueprint; and provision must be made to update the

history of each item every time it is used.

Obviously local test development is not a procedure to be undertaken in order to save money on testing. Nor is it a procedure to be undertaken with the expectation that all that is involved is to recruit a group of teachers, provide a one-day training program, arrange for a few days of writing time, and assume that the test (or items) will meet reasonable standards of item quality.

However, if a school system will put a reasonable amount of effort and resources into such projects, the rewards can be great. Teachers will have access to a bank of items which have been tried out and which are generally better than an individual teacher can develop in isolation; innovations in the teaching of certain courses can be evaluated with tests which are valid for the specific course; deficiencies in courses can be identified and remedial procedures can be evaluated. Furthermore, the careful examination of the curriculum and the goals of teaching which the PROCESS of developing test items requires should have an impact on the teaching practices of the teachers involved.

AUTHORS

ANNIE W. WARD, Ward Educational Consulting, Inc., 2800 N. Atlantic Avenue, Apt. 1406, Daytona Beach, Florida 32018

GENE BARLOW, Supervisor, Dade County Public Schools, 1410 N.E. Second Avenue, Room 800, Miami, Florida 33132.