DOCUMENT RESUME

ED 097 415		CE 002 112
AUTHOR	Fortune, Jim C.	
TITLE	Performance Test Development in Machine Shop: Appendix H. Final Report.	
INSTITUTION	Massachusetts Univ., Amherst. Center for Occupational Education.	
SPONS AGENCY	Massachusetts State Dept. of Education, Boston. Research Coordinating Unit for Occupational Education.; New York State Education Dept., Albany. New York Research Coordinating Unit.	
PUB DATE	Jun 72	
NOTE	39p.; For related documents see ED 060 218, ED 085 541, CE 002 111-117	
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE	
DESCRIPTORS	*Behavioral Objectives; *Criterion Referenced Tests; Educational Objectives; Industrial Arts; *Item Banks; Machinists; *Mechanics (Process); *Performance Based Education; Shop Curriculum; Test Construction; Trade and Industrial Education; Vocational Education	
IDENTIFIERS	ESCOE; Massachusetts; New York	

ABSTRACT

This is one of the outcomes of the work of the Massachusetts Evaluation Service Center for Occupational Education (ESCOE). The first part of this document is an overview of the Performance Test Development Project. The remainder of the document explores machine shop curriculum in terms of terminal behavioral objectives which were grouped by desired performance. Each performance group was synthesized into a single multifaceted objective (synthesized objective). Blueprinting was selected as the test item to be used in the initial field test which is described at length in terms of test description (general form and administration procedures), field testing, and revision recommendations. Tables and graphs supplement the report. (BP)

FINAL REPORT

Evaluation Service Center for Occupational Education

APPENDIX H:

Performance Test Development In Machine Shop

by

Jim C. Fortune

Submitted To:

Commonwealth of Massachusetts
Department of Education
Division of Occupational Education
Research Coordinating Unit
Boston, Massachusetts

State of New York
Department of Education
Bureau of Occupational Education
Research Coordinating Unit
Albany, New York

Evaluation Service Center for Occupational Education
Center for Occupational Education, School of Education
University of Massachusetts, Amherst

June 1972

This test is one of four tests published as separate appendices to the Final Report of the Evaluation Service Center for Occupational Education, Center for Occupational Education, School of Education, University of Massachusetts, Amherst, Massachusetts, June 1972.

TABLE OF CONTENTS

OVERVIEW OF THE PERFORMANCE TEST DEVELOPMENT PROJECT

Background

The 1968 Amendment to the Federal Vocational-Technical Education Act mandated the development of state-wide evaluation systems for the administration and operation of federally supported vocational education. Parallel to this mandate the Research Coordinating Unit director for the Commonwealth of Massachusetts was in the process of completing some predesign activities for the development of a vocational-technical education management information system. By 1969 the predesign of this system had moved into the feasibility stages and specifications of the system were being developed.

At this stage of development New York State, which already had a fine centralized testing program, became interested in the philosophy espoused by the Massachusetts system and joined in the funding of a more intense feasibility test, which eventually became the source of the Performance Test Development Project. The Evaluation Service Center for Occupational Education (ESCOE) was funded in July 1971 and was housed in Amherst, Massachusetts, to test the feasibility of systems development based upon the principles of (1) local control and development of vocational curricula, (2) data-based feedback based upon tailored performance tests, and (3) curriculum description through terminal behavior objectives. The following report deals with a subcomponent of the ESCOE system which was designed to develop performance tests as software support for the ESCOE program.

Whats and Whys of Performance Testing

Performance testing is more a new reality as opposed to a new concept in educational testing. The concept grows out of the need felt by educators to sample actual performances of trainees as opposed to merely measuring symptoms of desired (or intended) competencies through paper and pencil tests and then relying upon the predictive powers (i.e., previously established associations of paper and pencil test scores to some hypothetical or observed criterion of competency in performance) of the test to infer competency acquisition. This felt need has grown in part from the inability of standardized achievement tests to deal with the unique objectives of a specific educational program, in part from the reportedly low correlations between measured skills and on-the-job (or in-the-shop) performances, and in part from the lack of realism involved in the paper and pencil testing situation.

Hence the performance test can be conceived of as a criterion-referenced test, in that (1) it is objective or criterion-centered (in one-to-one correspondence with the extant component of a stated objective); (2) it seeks to ascertain a subject's possession of a specific competency rather than to complete a comparison of the subject's competency level to a previously measured norm group; and (3) it usually requires a dichotomous decision as to whether the competency has been demonstrated. The performance test can be construed to be a special case of the criterion-referenced test in that there is a definite attempt to establish fidelity between the sample observation of the performance test and the performance being sampled.

In the evaluation of instructional programs in vocational-technical education, the concept of performance testing is especially appropriate for several different reasons. First, performance tests can be hypothesized to produce more relevant and valid date concerning the instructional program output. Vocational program objectives tend to deal with competencies which require concurrent behavior changes across several domains of instructional objectives.

Hence the accomplishment of a vocational objective may depend upon the development of a psychomotor skill, the mastery of a cognate process, the acquiring of some fundamental facts, and the development of a particular attitude. Unlike paper and pencil tests, which emphasize the measurement of the cognitive aspects of the performance or observations which emphasize process and action components, performance tests possess the potential to measure the mixture of behavior domains appearing in the desired performance. The performance test can therefore be argued to offer a valid means of measuring intended outcomes.

Second, performance tests produce product records which can be studied by teachers to diagnose the place in the instruction where a weakness may have occurred, aiding considerably their ability to analyze their instructional methods. Since the teacher can determine what aspects of the competency are missing, he can trace the point in his instruction where his objectives were not met. Also, since the product is concrete it can be kept longitudinally to analyze pupil growth at different stages of a multi-year program.

Third, the nature of the data produced by performance testing contains the flexibility demanded by the information needs of an evaluation system. The tests are constructed in one-to-one correspondence to stated objectives, thus enabling selection of test components from a data bank situation in such a manner as to tailor the testing to the measurement of a unique set of program objectives. Since the tests are objective specific, comparisons of small aspects of an instructional program are possible. Since the tests are criterion-referenced, skill attainment in a particular area of interest can be ascertained; hence output of instructional programs can be described relative to percentage of skill development.

Restraints on Test Development

The design of the performance tests had to take into account both the philsophical and the operational structure of ESCOE. At times both of these structures served as restraining and occasionally frustrating hurdles for the test development team.

The philosophical nature of ESCOE provided the foundation of principles which are believed to have caused the performance tests to be unique. Since the objectives were generated by each local school, several very similar objectives appeared for a single behavior within a subject. Dr. David Berliner, now with the Far West Laboratory for Educational Research and Development, invented a process to state these similar objectives into a synthesized form accompanied by item changes providing for the unique characteristics of each objective. Thus, if enough objectives from different schools were collected to represent the curricula, by synthesizing those objectives one could arrive at a statement of all desirable behaviors within one curriculum.

The raw objectives based upon the curricula of each of the participating schools were synthesized to identify the major behaviors within a curriculum area. Hence, if the process worked ideally within a curriculum area a linear set of behaviors was produced. The degree to which this process failed to produce such a linear array of behaviors comprised the first major restraint. If a singular listing of behaviors could not be gained. then singular test items could not be written.

A second philosophical principle which developed into a restraining factor was the decision to test only locally-maintained objectives within a specific program. This principle actually involved several implications for testing. First, a student would be tested only on the objectives maintained by the curriculum he was receiving. Therefore, the test items had to be described in a form indicating one-to-one correspondence with the synthesized objectives so that the local teacher could select only those items maintained for his course. This selection pattern, however, did increase the logical assumption that the tests possessed high validity in regard to the courses for which they were designed to measure outcomes. Second, each item had to be independent in its ability to be administered, since previous or adjoining items would not necessarily be administered with it. This item independence served as a restraint to test development in that objectives could not be clustered into tasks involving several test items.

The third restraint involved both philosophical and operational aspects in that two forms of scoring were preferred by the two cooperating states. Philosophically, the state coordinators differed on the location of scoring; this disagreement became a restraint to test development in that the items developed had to be scorable both in the local school and at a central test center. Three forms of scoring meeting this restraint were adopted, with choice of scoring form depending upon the nature of the individual item. Two of the forms are based upon meeting the restraint with a single scoring process. The third form requires two different processes in order to meet the dual scoring restraint.

The scoring approaches requiring only one process are (1) the caliper or mechanically scored form and (2) the selection of correct response form. In the mechanically scored approach, several measured settings can be placed in a test scoring kit; the student or teacher records by label which of the settings fits the final product. A key of correct setting labels can then be referred to, producing a dichotomous score for the product in terms of size tolerances. In the selection of correct response approach, correction keys can be applied directly to the students' responses. In both cases either a central office or an individual classroom teacher can use the keys.

The third scoring form is not as simple, since two types of scores are required to meet the dual-use restraint. This scoring form is necessitated by the many tasks in the vocational curriculum which require expert observer judgment for the determination of performance quality. The two types of scoring needed for these items are (1) structured criteria for observation, and (2) pictorial records (color-coded to facilitate central scoring). The structured criteria for observation communicate to the teacher what aspects of the product to check in order to judge the performance successful. These criteria would be used in class. In the pictorial scoring process, camera angles have been described which would allow Polaroid pictures to be taken as records of the finished product. Color-coding the criteria checks would enable observers in a central location to determine the quality of the performance.

Each of these three approaches provides a means through which credible and unbiased scores can be obtained. All of the processes can be scored by individual teachers and used within a classroom setting without the aid of a central scoring station. The fourth restraint to test development arises at this point, since it is impossible to arrive at an immediately usable set of norms through the current scoring system and the dichotomous item response without implementation of a program designed to gather enough data to norm the tests.

Two other restraints were present throughout the test development project, both operational in nature. First was the quality and quantity of the behavioral objectives themselves. Few if any of the curriculum areas were fully described, and the tests developed are limited to described curriculum. In two test areas, more items were developed and the synthesization process was repeated in order to sharpen the synthesized objectives. In these cases much curriculum had been left undescribed and the fill-in process aided considerably in explaining the descriptions. However, complete and multiple sets of items were not available from each school; therefore the test items may be lacking in content validity in cases of consultant-written items, may be representative of several behaviors, and may hence be difficult to test or represent only a small segment of the previously unwritten curriculum.

The second operational restraint was that of time. Although the budget was small, the seriously close deadlines in development work made time an even greater restraint. Creativity is sometimes especially evasive under deadlines and within the constraints of administrative conflict. Still, the time dimensions were met in terms of design. Since schools were closed during the critical month of June, illustrations of some items of the tests could not be produced; therefore only plans, item descriptions, materials descriptions and administration instructions could be developed.

A final restraint can be observed in the language in which the proposal was written. First, several terms apparently changed in meaning or in relevance to the project once development began. One apparent change occurred in the description of sixteen tests for four areas. One test for each level of a curriculum area cannot be developed so as to be equally relevant to all schools. Since the schools maintain different objectives, different items must be assigned to each school, even on the same level. Hence a more appropriate process becomes the development of an item bank from which tailored tests can be developed for each individual program. Second, the time restraints and the differences in nature of curriculum required different kinds of tryouts, making the language of the proposal seem sometimes inappropriate.

Purposes of the Test Development Project

The design of the test development project included not only the goal of producing tests as products but also the goal of establishing feasibility of the test development effort across a broad spectrum of vocational-occupational curricula. For this reason four different areas of vocational curricula were selected for test development. These four areas differed in hypothetical difficulty of test development. The areas chosen were machine shop, woodworking and carpentry, electronics, and automobile mechanics. The automobile mechanics area was hypothesized to be the most difficult since manufacturers determined the curriculum, which therefore differed across competing manufacturers.

The performance tests were hypothesized to be sufficiently flexible to fulfill many purposes of a comprehensive evaluation system. Because of their proximity to the desired outcomes, performance tests were hypothesized to serve as (1) student diagnostic and prerequisite instruments, (2) diagnostic instruments for the analysis of instruction, (3) criterion instruments, (4) measures of classroom achievement, and (5) program success indicators. Each of these uses has already been piloted to some extent.

The performance tests as developed have several application conveniences. First, since the test items are paralleled to synthesized objectives, computer selection of test items or "synob" comparison of items can be used as a methodology for tailoring tests to instruction. Second, since the conceptual frames of the tests can be described, each test has built-in potential updating or extension by the classroom teacher.

Problems Encountered

Problems occurred from three viewpoints. First was the problem of lack of known direction, a handicap which often occurs in the area of development. Second was the problem of lack of perfection or completion of the objectives used as raw materials for the development of test items. Third was the problem of contending with dual scoring requirements and with several different kinds of program emphasis and structure.

The first problem has been emphasized recently with the development work done on criterion-referenced testing. From a conceptual point of view, the criteria previously used to determine the quality of norm-referenced tests can no longer be used for criterion-referenced tests. Since the measurement strategy of the criterion-referenced test and the performance test is to determine the possession of either a skill or the capability to carry out an activity or process, the degree to which the test differentiates between subjects taking the tests does nothing to indicate test quality. Unlike the norm-referenced test, in which measurement strategy is to distinguish between subjects, the performance test cannot be hypothesized to produce large differences across subjects nor can any specific level of difficulty be expected. Hence, average levels of difficulty and large differences between subjects do not indicate quality of the performance test.

In performance testing, some concepts of reliability still appear useful, while others appear to have lost their relevance. Reliability over time, or test-retest reliability, is still meaningful as long as the time between tests did not include opportunity for the subject to acquire the skill in question. Since performance tests are designed so that each item does not necessarily refer to the same skill or activity, reliability indices dealing with homogeneity of the test no longer appear to be relevant criteria for test quality.

The degree to which the items of a performance test cover the skills of an area and approximate actual required performances operates in a similar relationship to the performance test as that of a prediction index to a norm-referenced test. This degree of similarity might be compared to the concept of fidelity so often used in the recording industry.

The second problem area involved the quality of the raw materials used. As should be expected, the synthesis process does not apply evenly to all areas and was not applied with the same consistency to each set of objectives. In the machine shop curriculum area, between 70 and 80 percent of the content was described by the objectives. These objectives possessed adequate depth across skill areas to enable the synthesis process to produce clear synthesized objectives describing unique performances. The creation of items parallel to the synthesized objectives and possessing the independence and flexibility required by the philosophy of the system was a straightforward process.

In the woodworking area, between 60 and 75 percent of the content was described. Unfortunately, the synthesizers of the raw objectives failed to produce synthesized objectives which dealt only with single performances. Instead, the raw objectives were synthesized by similar or related behaviors and the product of this process was a matrix of similar performances (rather than a single performance) with several form changes denoting differences in conditions and extents across schools. Since these products seemed usable, the decision was made to produce a matrix of test items generated in one-to-one correspondence to the performances included in each synthesized objective.

This decision was the source of some time lost due to the expanded number of test items which had to be written; however, this increase in items was accompanied by a large increase in test specificity, which increases the degree to which the performance test can be tailored to fit a given instructional program without any noticeable loss of efficiency of the item banking process.

Due to the variance of material and the limited scope of the objectives developed for the electronics curriculum area, a decision was made to rewrite many of the synthesized objectives. For more than one-half of the contract period two of the test development team members struggled to find a format within which the scope of the electronics curriculum could be described. By expanding the number of conditions it was found that classes of performance could be described by synthesized objectives. Hence, through considerable redesign and a small set of compromises of the synthesis process involving uniqueness of performances and allowance of performance form changes, subcollections of electronics objectives could be written which would allow test development along similar conceptual lines as those followed in the development of the machine shop test. Results of the test development effort again produced item banks, as in the two previous test areas, with the items possessing similar relationships to the synthesized objectives.

In the area of automobile mechanics, less than 50 percent of the content was described by the raw objectives. Many of the subdivisions of content were too sparse to allow for the development of synthesized objectives. In addition, the synthesis process applied seemed irregular across blocks and units. The level of abstraction of behaviors described by the raw objectives and the interdependence of the performances raise questions concerning the appropriateness of the synthesis process in this area. Certainly, the limited number of usable synthesized objectives and the necessary revisions of the existing objectives made the decision to rewrite the objectives essential. Revision of the curriculum descriptions were made in relationship to the job orientation of the curriculum. Test items were written around standard mechanics tasks as described in the automobile mechanics curriculum. In some of these items, synthesized objectives are tested in a format which includes a cluster of the objectives provided by the ESCOE system. In other items, only parts of ESCOE-produced objectives are included in the new synthesized objectives being tested. Once a test item has been constructed, the process can be reversed so that system capability as achieved in the other three test areas can be gained. Because of their time-consuming nature, tasks in the curriculum such as disassembly or reassembly of a motor or transmission were not included as complete test items. Instead, either sample tasks extracted from the large unmanageable task or written or pictorial selection items were created to test these phases of the curriculum.

The third problem area encountered was the difficulty involved in the existence of two separate scoring requirements and in the time limitations of the test development project. It was not always possible to produce useful in-class scoring of the performance item and credible, objective centralized scoring of the performance through application of the same scoring process. Therefore some items are suspected to produce more useful scores in the classroom than in a central scoring situation, while the reverse is suspected of other items. Only time and study of the tests can alter or affirm these suspicions. It is unfortunate that systematic refinement of the woodworking, electronics, and automobile mechanics tests is not planned to occur along the same lines as those applied to the machine shop test.

The following report includes development and field testing procedures, item bank descriptions, recommended analysis procedures and uses for one of the four test areas briefly described above.

Introduction

The machine shop test has often been referred to as the Booth test, in reference to Russell Booth, the original developer of the basic test format. The test is the pioneer performance test upon which many of the basic ideas of the ESCOE system were initially implemented. It was the first test developed because of the completeness of the objectives obtained for the machine shop curriculum area in the Massachusetts pilot test. This completeness of description by objectives was not replicated in the remainder of the ESCOE project, perhaps indicating some uniqueness of the curriculum area itself.

Work with the objectives in this area indicates that the skill aspects of the curriculum are easily adapted to the behavioral objective format. Performances are straightforward and can be described in terms of relatively independent activities. The performance of the activities requested by the objectives takes place in a somewhat standardized environment with a finite set of alternative behaviors. Thus each synthesized objective states a fixed desired performance which is to occur in a relatively standardized environment.

Conceptual Scheme for Development

Once each LEA had described its machine shop curriculum in terms of terminal behavioral objectives, the objectives were grouped by desired performance. Each performance grouping was then synthesized into a single multifaceted objective called a "synthesized" objective. The synthesized objective stated the unique performance and denoted the uniqueness of each LEA through the inclusion of form changes, which were recorded for both condition and extent. This synthesized objective was used for the conceptualization of the machine shop test, and a test item was created in one-to-one correspondence to each synthesized objective.

The process through which the test item was ultimately formulated was one of communication between a psychometrician and a consultant machine shop teacher. Beginning with the basic synthesized objective, the vocational educator offered a verbal description of the performance demanded. Expected time required to complete the performance and uniqueness of the required performance were then discussed. Following this verbalization the psychometrician converted the performance description into a test event, which the vocational educator translated into a form which could be used to communicate the task to students. In this case the form was blueprinting. After the test item had been agreed upon, criteria for successful performance were discussed and a scoring scheme was devised. This process was practiced for each synthesized objective until the curriculum was completed.

In the pilot run and in the basic first test developed, only about one-third of the curriculum was tested. Field tests of this pioneer unit showed promise and suggested some focus changes as well as the addition of detailed test administration directions. The past year's work expanded this test to include approximately 75 percent of the curriculum and established a field administration trial of the total test.

Test Description

General form. In addition to expansion of the original machine shop test, several teacher administration options were created. Figure 1 shows the item selection form, which allows the teacher to designate which items are to be taken by selected groups of students. Thus the teacher can elect to test students upon only a few of the items or upon all of the items, can elect to test the whole class or only part of the class, and can elect for the students to pursue the same testing program or for each student to complete a different selection of items. The item selection form provides a vehicle through which the teacher can report his testing intentions if the system makes this option available.

The total test is designed into operations performed on two pieces which can eventually be put together. Piece One covers 19 operations and Piece Two includes 16 operations, as illustrated on the form in Figure 1. Using standard grading practices, an additional 17 supervisory and grading operations can be built into the testing process. Therefore, 52 terminal objectives can be measured by the performance test; 35 of these performances result in products which can be carried across years or levels of a student's program, yielding visual product records of his growth. The 17 remaining performances, most of which fall toward the completion segment of the curriculum, can be kept in written record form. Figure 2 illustrates a potential record form for recording students' completion of grading items.

The total test was conceived to require eight to twelve hours for completion, and only the shop tools normally present in any machine shop instructional setting are needed for the administration of the test. In fact, it is felt that greater instructional validity is gained when the students are tested on the same shop equipment on which they were instructed. Hence no change of equipment or additional equipment is needed to conduct the test.

<u>Administration procedures</u>. As previously stressed, the machine shop

MACHINE SHOP ITEM SELECTION FORM

School ______________ Class ______________

Testing 1 2

PIECE ONE OPERATIONS	Number Taught Objective	Number To Attempt Objective
Cut off stock		
Position stock in chuck		
Face off and center drill both ends		
Place between centers		
Layout various "steps"		
Straight turn all diameters		
Perform necking operations		
Cut chamfers		
Cut thread to be "chased" (tool bit)		
Reverse piece		
Shoulder turn		
Taper turn		
Cut thread with die		
Drill and tap hole		
Cut woodruff keyseat		
Inspection		
Repair center holes		
Cylindrical grind (taper)		
Inspection (taper and finish)		

BEST COPY AVAILABLE

MACHINE SHOP ITEM SELECTION FORM

School ______________ Class ______________

Testing 1 2

PIECE TWO OPERATIONS	Number Taught Objective	Number To Attempt Objective
Cut off stock (power saw)		
Plain mill (or end mill) to thickness		
Position stock in chuck of lathe		
Drill, bore and cut thread		
Inspection for thread		
Position stock in vertical m.m.		
Bore out the thread		
Cut keyway		
Numerical control		
Press piece on mandrel		
Turn OD and inspect		
Mount on index centers		
Cut gear teeth or cutter teeth and inspect		
Heat, treat and inspect		
Cutter grinder and inspect (if necessary)		
Surface grind both sides		

FIGURE 1

Name_______________

School_______________

Grade level_______________

MACHINE SHOP TEST 17.2302

Sequence of Operations (Piece #1)

	Day #1	Day #2
Finish		
Start		
TOTAL		

Check each operation to be performed

Operation #	Operation	Dimensions	Time Required for Each Operation	Circle One (1=excellent; 5=poor)	Comments
1	Face off both ends BLUEPRINT READING: Facing	6-1/8		1 2 3 4 5	
2	Center drill both ends BLUEPRINT READING: Center drilling	3/16 Dia.		1 2 3 4 5	
3	Straight turn BLUEPRINT READING: Turning, straight	1-1/8		1 2 3 4 5	
4	Necking BLUEPRINT READING: Necking	3/16x.553 5/16x.620 1/4x11/16		1 2 3 4 5	
5	Chamfer BLUEPRINT READING: Chamfer	3/64 and 5/64		1 2 3 4 5	
6	Shoulder turn* BLUEPRINT READING:	1/8x11/16		1 2 3 4 5	
7	Taper turning BLUEPRINT READING: Turning, taper	.875x.775x2.000		1 2 3 4	
8	Die-cut 5/8-18 BLUEPRINT READING: Threading, external	P.D. .589 $^{+.000}_{-.004}$		1 2 3 4 5	
9	Drill (for tap) DRILL PRESS: Drill	.201 (#7 drill)		1 2 3 4 5	
10	Tap (1/4-20) DRILL PRESS: Tap	P.D. .217 $^{+.000}_{-.003}$		1 2 3 4 5	
11	Cut external thread 3/4-10** BLUEPRINT READING: Threading, external	P.D. .685 $^{+.000}_{-.006}$		1 2 3 4 5	
12	Mill keyway TECHNICAL DRAWING: Milling, slot	.250 $^{+.000}_{-.002}$		1 2 3 4 5	

FIGURE 2

BEST COPY AVAILABLE

test requires only the normal machinery or equipment used in instruction. The only additional materials needed are (1) the pieces of stock, (2) the instructions for administration, (3) the blueprints describing the test items, and (4) the optional grading kit.

Figure 3 shows the administration instructions used during the field tests. Obviously, these instructions are highly dependent upon plans for data use and upon the system's requirements. Different instructions will be necessary if all students are to take the same test. Figure 4 records the instructions given to the students. Again, these instructions depend upon system requirements for much of their content. For instance, if the test were to be centrally scored or graded, the students would be instructed to use the grading kit rather than the instruction center procedure. Figures 5 and 6 illustrate the test kit used for the field tests.

Field Testing

The machine shop test was field tested in three schools over several levels of students (Levels 1, 2 and 3 in four-level programs; Levels 1 and 2 in three-level programs) at two different times. The field test was designed to (1) develop estimates of required testing times, (2) produce estimates of item difficulty, (3) produce some estimates of test-retest reliability, (4) try out administration instructions, (5) try out recording forms, (6) gain the reactions of shop instructors to the tests, and (7) indicate directions for future revisions. Although the sample size was greatly decreased by poor timing and by a somewhat reactive attitude in the field to the fact that the ESCOE system was being terminated, thus

The Commonwealth of Massachusetts
University of Massachusetts
Amherst 01002

SCHOOL OF EDUCATION

In recent years occupational education has acquired considerable stature. The rise in youth unemployment and underemployment, the shortage of needed personnel in technical, semi-professional and skilled occupations, the retraining and continuing education needs of workers, as well as the rising demand for new educational opportunities have all served to highlight the need for a re-examination of the field of occupational education.

The Evaluation Service Center for Occupational Education is designed as a prototype project with pilot schools, like yours, which participated in the development of a bank of behavioral objectives, defining your stated goals.

These program objectives were stated by the local educational agency (LEA), not prescribed by a central authority. Following the determination of these objectives (raw objectives) by many local groups, synthesizing of these objectives was performed (combining raw objectives having the same or similar performances into one larger objective).

Your assistance is requested at this time to help us field test the "criterion test" developed as a result of this process. It might be pointed out at this time that the purpose of this "field testing" is not to compare your school or program with any other. We will be testing to improve and refine the test itself. Even when fully developed, this test will not be used as a "norm-referenced" measure of your school or program to be measured against others. Rather, this is a criterion-referenced test which you may or may not use for improvement or modification of your present program. (For advantages to teachers and students, see enclosure.)

Specifically, we would like your assistance in these areas for two testing days (one following the other at an interval of approximately two weeks):

* Supplying us with six students with varying skills (i.e., two high, two middle and two low ability students) from each grade level involved in the machine shop program

* Setting up an "inspection area" in the shop manned by one of your better students (seniors?) during periods ... of each student's work. A guide ... supplied with the following inspection and measurement equipment made available:

  1) 6" and 12" steel rules
  2) 1", 2" and 4" micrometer calipers
  3) Vernier calipers
  4) Screw thread micrometer
  5) Go-no go thread ring gage (3 ...)
  6) Go-no go thread ring gage (5 ...)
  7) Taper ring gage, # Morse taper
  8) Go-no go thread plug gage (...)
  9) Hardness tester (Rockwell, Brinell, etc.)
  10) Milling cutter clearance gage

* Your comments, criticisms about how the test, testing, etc. could be improved

* Setting aside a small storage area for kit materials so that tasks may be performed on the test pieces as they are learned. (In the ideal situation a student receiving the piece as a freshman would work on it until he was a senior.)

* Getting together with the other teachers involved and selecting a time when all the students and teachers to be engaged in the project could be gathered together for a meeting with me, so that I could explain the purposes of the test and test procedures and so that you could answer their questions on specifics (as I have no experience in this area). A meeting should be arranged as early in the morning as possible as I will be able to spend only one day on each testing occasion in your school.

On the testing date each student will be given a kit containing:

One or two test pieces (depending on grade level)

A blueprint

An operations sequence with instructions

An inspection checklist

BEST COPY AVAILABLE

FIGURE 3

We do not expect [illegible] student to perform every operation. Some operations might not be taught in your program, some necessary machines may not be available, or the student may not have reached that goal in his [illegible] process to date. If there are operations which the students [illegible] taught, I would like to [illegible] the "test" [illegible] particular sequences.

Thank you in advance for your cooperation.

Yours truly,

Philip Brooks
Research Assistant
School of Education

EVALUATION SERVICE CENTER FOR OCCUPATIONAL EDUCATION
85 North Whitney Street
Amherst, Massachusetts 01002

Advantages to the Student

1. From entrance to exit from each course, teachers and students know the specific objectives to be attained.
2. Specific objectives, stated in clear behavioral terminology, minimize misunderstanding.
3. Students know how each learning experience coordinates with his whole occupational education experience.
4. Knowledge of course objectives allows students to assume more responsibility for their own progress, in that:
   a. students complete objectives at their own rate.
   b. there is a minimum of unnecessary repetition.
   c. there is a minimum of waiting for the next assignment (wasted time).
   d. there is increased student motivation and a decrease in discipline problems.
5. Knowledge of course objectives and assumption of more individual responsibility can help students develop positive attitudes toward their occupational involvement, their teachers, their schools, etc.
6. There is increased opportunity for more flexible scheduling to meet individual student needs.

EVALUATION SERVICE CENTER FOR OCCUPATIONAL EDUCATION
85 North Whitney Street
Amherst, Massachusetts 01002

Advantages to the Instructor

1. Information in ESCOE data bank can assist in evaluation of existing course content, objectives, and goals.
2. More effective communications between teachers in each LEA, both within and between instructional programs.
3. Examination of course structures can lead to elimination of identical objectives within a student's program of study.
4. More effective evaluation of teaching methods and process.
5. Simplification of tests and testing procedures.
6. Feed back of all objectives from other LEAs' from the ESCOE data bank.
7. Feedback of test data based on tests developed from your own objectives as submitted to ESCOE.
8. Modification or initiation of LEAs' courses or programs based upon an analysis of the ESCOE data bank of objectives.
9. Feedback of student progress achievement through computer-based techniques.

General Advantages

1. Students, teachers, administrators, and other interested persons know clearly the stated goals of each program of study.
2. Establishment of an organized systems approach to curriculum development, analysis, and revision.
3. Psychomotor, Cognitive, and Affective Capability Taxonomy can be of value in assisting the counseling function of the LEA by describing desired student performance.
4. Provides for instructor participation in development of systems for Accountability in Occupational Education programs.

NOTE: Correction on earlier Technical Report #1, Page 47 -- C.1.1 is changed to C2.1

FIGURE 3 (Cont'd)

TO THE STUDENTS

You have been selected to work on a government-sponsored study of vocational education, specifically machine shop. This is a "field test" so that we can modify and improve the test before giving it to more students.

Your performance on these skills (in the lathe, drill press, and bench work areas) will not be counted toward your course work. A checklist of performance (included in your test kit) will be kept, as you will be bringing your test piece to the inspection center for checking after each step is performed. However, if the inspection area is full or the checker is not available, just go on to the next step.

Your test kit contains either one or two pieces to be worked with, a sequence of operations with instructions, a blueprint, and the checklist already mentioned.

You are not expected to be able to perform all the operations, and no one's work will be perfect. Do not do those operations you have not had experience with.

Are there any questions now, before you get your test kit?

(Pass out kit)

Any further questions?

Please keep this piece and work on it whenever you have completed any of the items.

FIGURE 4

NAME ______________________

GRADE LEVEL ______________________

	Day #1	Day #2
Finish		
Start		
Subtract Breaks		
TOTAL		

SEQUENCE OF OPERATIONS (PIECE #1)

[illegible] if sequence is to be performed

Sequence Number	Operation	Block	Unit	Dimensions	Time Required	Satisfactory	Unsatisfactory
1	Face off both ends	01	01	6-1/8"			
2	Center drill both ends	01	05	3/16 Dia.			
3	Straight turn	01	01	1-1/8			
4	Necking	01	10	3/16x.553 5/16x.620 1/4x11/16			
5	Chamfer	01	13	3/64 and 5/64			
6	Shoulder turn*	01	08	1/8x11/16.			
7	Taper turning	01	22	.875x.775x2.000			
8	Die-cut 5/8-18	09	13	P.D. .589 $^{+.000}_{-.004}$			
9	Drill (for tap)	03	04	.201" (#7 drill)			
10	Tap (1/4-20)	03	07	P.D. .217 $^{+.000}_{-.003}$			
11	Cut ext. thread 3/4-10**	01	23	P.D. .685 $^{+.000}_{-.006}$			
12	Mill keyway	02	07	.250" $^{+.000}_{-.002}$			

NOTES (PIECE #1)

* Piece has been reversed between the centers.

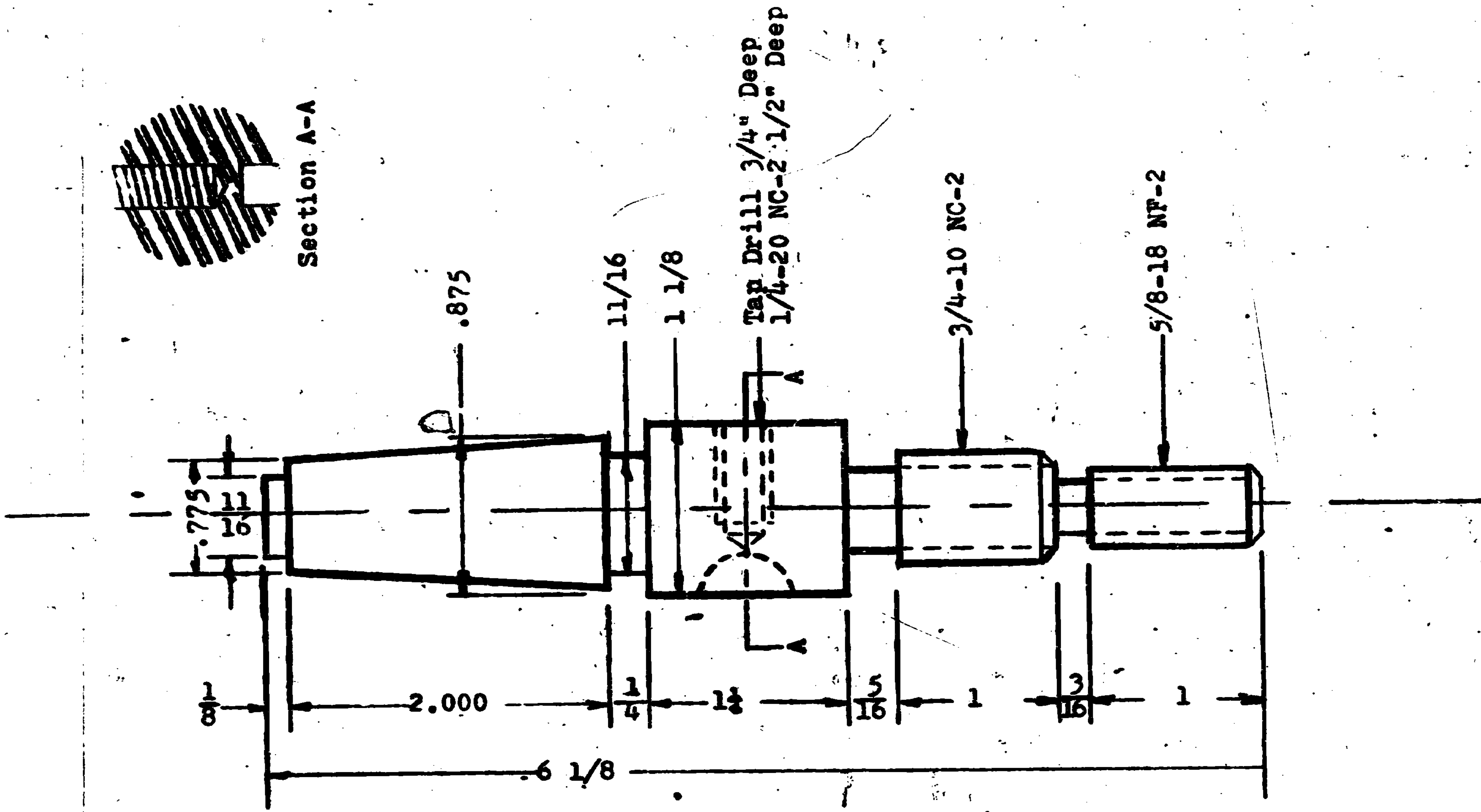
** If the freshman is capable of performing this operation, it will be the sixth operation in sequence.

Sizes of threads may be changed to accommodate available taps and dies.

Dimensions for external threads are given for thread micrometer measurements. If the LEA teacher wishes to test the three-wire system, the students will compute the measurement.

At the discretion of the local teacher a straight slot may be substituted for the Woodruff key slot.

FIGURE 5



Note: Chamfer and neck threads to Single Depth.
Leave .020" on Taper Dia. for grinding.

#5 Woodruff key (Section A-A)

Scale 1-1		Sept 19, '71	Dr. By
Clearance unless otherwise specified Decimal Dimension +- .005 Fractional Dimension +- 1/64 Angular Dimension +- ½ Degree			FINISH: Break all sharp edges.

NAME ____________________

GRADE LEVEL ____________________

Finish		
Start		
Subtract Breaks		
TOTAL		

SEQUENCE OF OPERATIONS (PIECE #2)

Sequence Number	Operation	Block	Unit	Dimensions	Time Required	Satisfactory	Unsatisfactory
1	Drill (for bore or ream)	02	02	31/32			
2	Bore or ream	02 02	01 03	1.000"			
3	Cut keyway	02	07	.250x.125			
4	Face off or mill to thickness	01 02	02 05	.270			
5	Straight turn (cleaning cut)	01	01	1-1/8			
6	Prepare program (N.C.)	08	01				
7	Prepare tape (N.C.)	08	02				
8	Drill and cntrsk (N.C.)*	08	03				
9	Cylindrical grind	17	01	3.750			
10	Set up index head	02	11	No error allowed			
11	Cut milling cutter teeth	02	09	1/2" deep			
12	Harden and temper	11	02				
13	Grind clearance angles	17	02	$+1^{\circ}$ -0°			
14	Surface grind	17	03	.250			

NOTES (PIECE #2)

* No decision has been made on inspection of the matching operations. The position of the holes may be checked by measuring over inserted plugs (pieces of drill rod), and this would necessitate adding a reaming operation to the procedure.

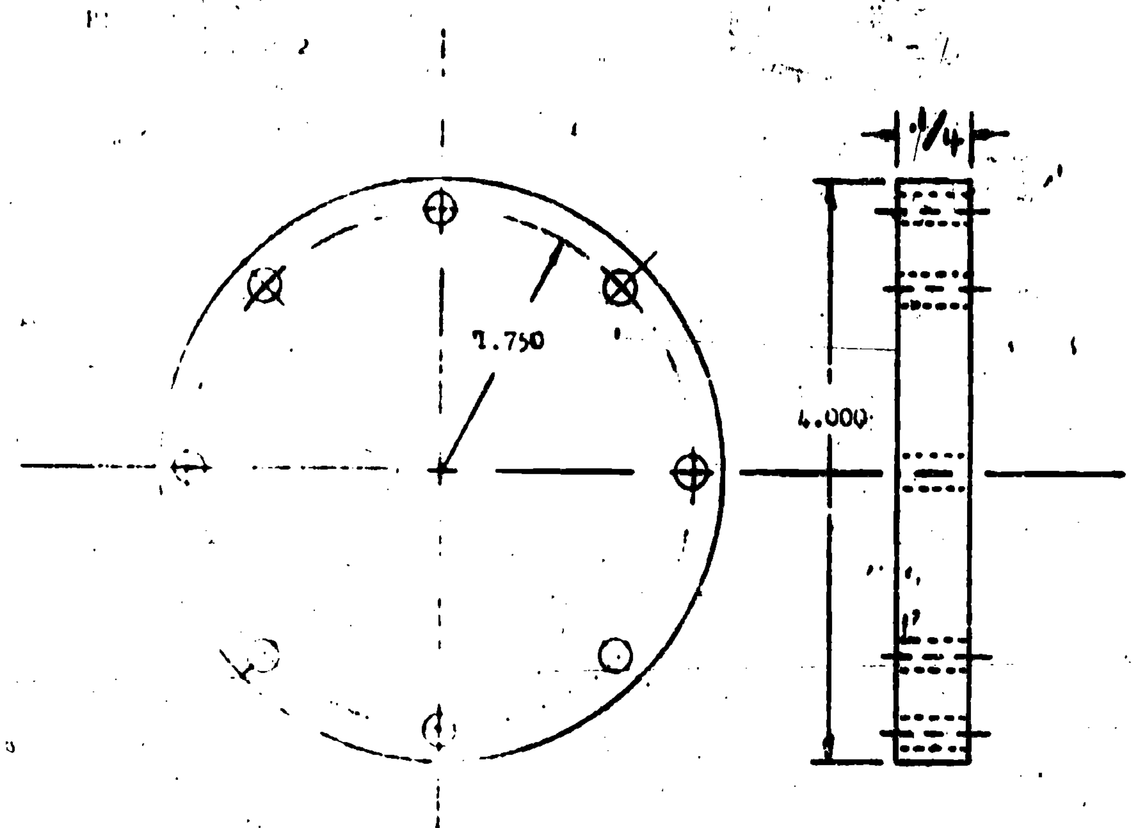
Because of the 1/4" thickness of the piece, it may be necessary to produce a backup plate.

Tolerances on dimensions will be determined by the LEA to conform to raw objectives.

Sheet #1 will be used if numerical control is tested. The holes will subsequently be milled out of the piece when the gear or milling cutter is made on the milling machine. It is suggested that the milling cutter be made because:

1) The 60° angle cutter is more likely to be available in the shop.
2) Objectives pertaining to cutter grinding can be tested.
3) The resulting cutter may be put to practical use in the shop.
4) The results of the heat treating will be more impressive when the cutter is used.
5) Objectives relating to gears can be effectively tested on written tests.

FIGURE 6

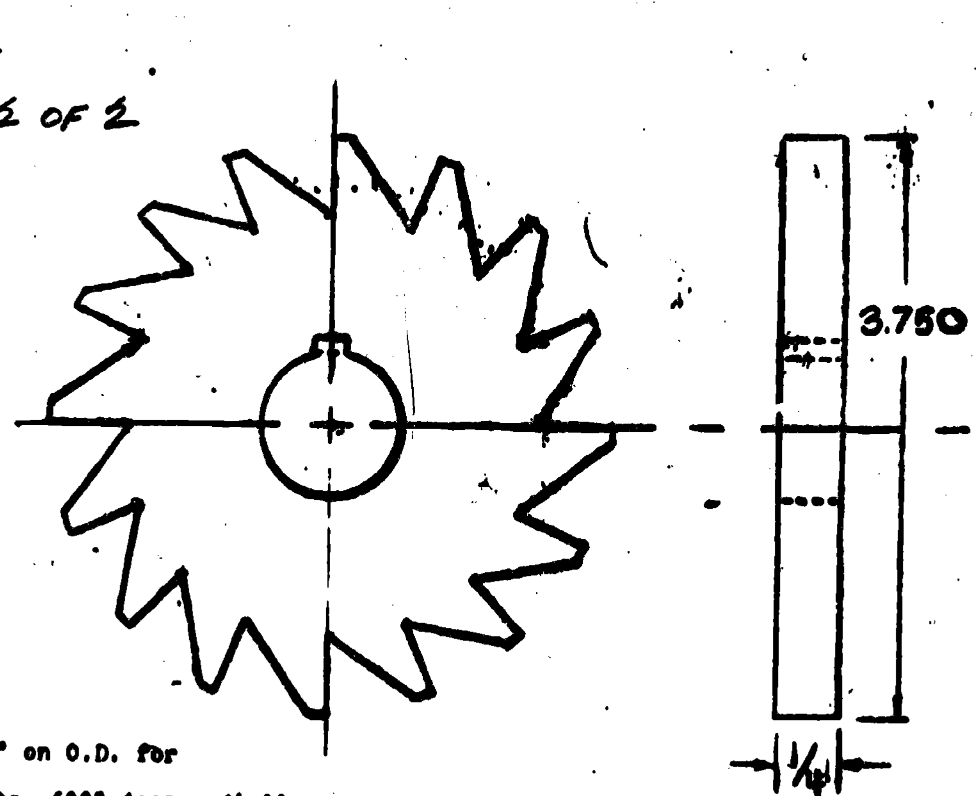


NOTE: ... holes equally spaced. Drill ... dia. and counter-sink ... dia., 60°. ... and width ...

Scale 1-1		Sept. 19, '71	Dr. By
Clearance unless otherwise specified Decimal Dimension +- .005" Fractional Dimension +- 1/64" Angular Dimension +- 1/2°			

Piece #2

Sheet No. 2 of 2



NOTE: Leave .020" on O.D. for Grinding.
Tooth face to .500" deep radially and 0 rake. Use 60° angle cutter.
Tool & Cutter Grinder:
Primary angle 8°
Secondary angle 16°
Width of land 1/16"

Bore and ream hole 1.000" +.0005 -.0000
Width of keyway .250"

Scale 1-1	Material SAE 10120	Oct. 24, '71	Dr. By
Clearance unless otherwise specified Decimal Dimension +-.005" Fractional Dimension +- 1/64" Angular Dimension +- 1/2 Degree			Finish: Break all sharp edges.

FIGURE 4 (Cont'd)

preventing the administration of some posttests in the reliability study, the field test was deemed a success on all seven purposes.

Table I represents the distribution of students cooperating in the test. Results from any one group are not reported individually since comparison of schools was not part of the field test. This table shows only participation for the testing and does not include separate item counts, which are reported in later tables giving statistics for individual items. It should be noted here that not all students took every item and that not all students participated in the second testing.

Table II records the average time required for completion across items as well as the number of students completing each item. This table also indicates both maximum and minimum times to completion by levels of students across testing sites. It can be seen from Table II that the variance of time needed to complete each item was greater than expected, causing the total range of required time per center to vary between eight hours and 16 hours--eight being the minimum estimated and 16 being four hours longer than the maximum estimated. The overage time of 9.8 hours is just under the predicted ten hours. These time estimates should be sufficiently accurate to provide the test administrator with completion time estimates for any combination of items. However, the field test did indicate a need to improve or standardize time-keeping procedures and to provide a better-organized form for recording completion time.

Table III represents the estimates of item difficulty calculated from the field test. Most of the percentages of correct response ranged from 60 to 80 percent--perhaps indicating that the items are not quite difficult enough. This indication is not of serious concern, however, because of the

TABLE I

SCHOOLS PARTICIPATING IN MACHINE SHOP FIELD TEST

School	Level	No. of Students
Diman Regional Vocational-Technical High School	9	6
	10	6
	11	4
Greater Lawrence Regional Vocational High School	9T*	6
	9RT**	6
	11T*	3
	11RT**	3
Nassau County BOCES	1	2
	3	2

* Test

** Retest

TABLE II

TESTING TIMES FOR MACHINE SHOP TEST ITEMS

(Average Time, In Minutes, Reported for Items with Maximum and Minimum Times Observed at a Test Site)

Piece	Item	Number Taking	Average Completion Time	Maximum Observed Site Time	Minimum Observed Site Time
1	1	37	12.2	22.8	3.6
1	2	36	5.4	17.5	3.2
1	3	38	41.6	88.3	15.0
1	4a	37	14.2	30.0	5.3
1	4b	37	14.1	22.0	5.5
1	4c	37	13.5	20.3	5.5
1	5	37	9.1	25.0	1.0
1	6	37	14.5	45.6	2.0
1	7	35	28.2	55.6	16.6
1	8	34	17.7	30.0	8.8
1	9	33	8.4	18.5	2.8
1	10	33	8.9	17.5	2.0
1	11	29	29.9	72.3	6.0
1	12	24	13.1	22.5	4.0
1	13	3	10.0	10.0	10.0
2	1	10	20.8	27.0	14.5
2	2	9	19.7	30.0	11.5
2	3	5	10.0	10.0	10.0
2	4	11	76.4	84.1	67.0
2	5	11	24.2	29.4	20.0
2	6	11	63.6	84.0	46.6
2	7	11	44.1	47.5	40.0
2	8	6	103.8	105.0	103.2

TABLE III

ITEM DIFFICULTY ESTIMATES FOR MACHINE SHOP TEST

(Indices Report Percent Correct)

Piece	Item	Number Taking	Number Correct	Percent Correct
1	1	37	26	72
1	2	36	24	67
1	3	38	27	70
1	4a	37	14	38
1	4b	37	15	40
1	4c	37	16	42
1	5	37	25	66
1	6	37	24	65
1	7	35	16	46
1	8	34	27	80
1	9	33	30	90
1	10	33	28	84
1	11	29	20	68
1	12	24	16	67
1	13	3	3	100
2	1	10	10	100
2	2	9	9	100
2	3	5	5	100
2	4	11	9	81
2	5	11	7	63
2	6	11	9	81
2	7	11	9	81
2	8	6	4	67

number of advanced students taking the test. Item 4 on Piece One appears to be the most difficult, and is perhaps the most desirable item from a difficulty level standpoint. No incorrect responses were recorded on Item 13 of Piece One or Items 1, 2 and 3 of Piece Two. However, only a few students attempted these items, and due to their nature there appears to be no reason for concern.

Table IV indicates the results of the reliability study. Due to failure of two of the participating schools to provide for adequate retesting, the reliability study had to be based on a smaller than desirable sample. Two indices of test-retest reliability were computed. The first coefficient represents the percent of agreement between the pretest and posttest performances on each item. The second reliability estimate is the correlation of pretest completion time to posttest completion time for each item. A few items gave sufficient evidence of weak reliability to merit analysis and further study. Conceptual analysis of the four items indicating low reliability estimates failed to produce any reasons to suspect their consistency.

Item 4 on Piece One showed low but acceptable reliability in terms of replicated performance success, but showed little or no consistency in terms of completion time. This phenomenon is perhaps due to the difficulty of the item. Items 2, 7 and 12 on Piece Two indicated low percents of agreement (all near 40 percent levels). Items 2 and 7 also showed low completion time correlations. Perhaps their order of attempt could be an explanatory factor; further study should be conducted. The remainder of the items possessed acceptable reliabilities and percents of agreement (55 percent or above) and in most cases also showed acceptable correlations between completion times (60 percent or above). The time reliability should improve with improved

TABLE IV

TEST-RETEST RELIABILITY ESTIMATES FOR THE MACHINE SHOP TEST ITEMS (N = 11)*

Piece	Item	Percent of Agreement**	Correlation Between Completion Times***
1	1	56	.31
1	2	37	.45
1	3	100	.42
1	4a	55	.08
1	4b	44	.06
1	4c	44	.04
1	5	75	.27
1	6	42	.22
1	8	50	.71
1	9	100	.81
1	10	67	.73
1	11	62	.65
1	12	40	.98
1	13	--	--
2	1	100	.22
2	2	80	.65
2	3	100	.36
2	4	80	.46
2	5	67	.69
2	6	60	.70
2	7	60	.25
2	8	100	.67

* Not every participating student completed each item.

** Scoring same on posttest and pretest.

*** Time on pretest correlated to time on posttest.

timing and recording procedures.

The field test generated the following comments concerning the machine shop test, the testing instructions and the recording forms:

Diman Regional Vocational-Technical High School

1. Type of steel: Piece #1 is machine steel, cold-rolled.
   Piece #2 is Carlion tool steel (finished piece can be used).

2. Any operations may be altered to fit school (i.e., on Sequence #8 Diman Regional used 1/2-20 instead of 5/8-18. Also, different type keyway was cut as there were not enough machines to cut Woodruff key).

3. Sequences may be altered if machines are not available or other limitations arise.

4. Only one piece of stock is available per student. Any mistakes should alter only one other operation at most.

5. Checking can be done by the teacher at the end of the operation. There are too many operations to check after each one (each student should check his own after each operation). Estimated time to check each Piece #1 is 5-20 minutes. If tight, seniors may be used to check work.

6. Tools required have not been listed so that each school can better fit test to existing conditions (ex: after numerical control, Piece #2 may be strapped to table, put in vise, or have a special piece made up to hold it).

   Alternate view: Do not list tools necessary--may cause rigidity in viewing test.

7. Straight turning needs three time areas.

Greater Lawrence Regional Vocational High School

1. Perhaps an arbor piece could be made or provided for each school to cut flukes with dividing head.

2. Staggered start could be made on Piece #2, some doing numerical control.

3. Type of metal should be listed for Piece #2 (No. 11 special).

4. Make note on blueprint that center hole can be left in.

5. Could put a 5/16" ream at tapered end of Piece #1.
   a. Center drill
   b. Drill and ream

6. Set up the dividing head only once for all students.

7. Cannot do Piece #2 work on a milling machine (Block 2 sequences).

8. Change sequence from 1-2-3-4-5-6-7-8 to 2-3-8-1-4-5-6-7.

9. Diameter should not be different: 4" vs. 3.75" on some pieces.

10. Do only eight operations.

    General Comments: These tests are good for the students as they uncovered weaknesses--i.e., students who were proficient in certain areas were given projects or "jobs" covering those skills while others were weakened or left undeveloped. (It was too much like industry with the specialization that went on in the shop.)

Nassau County BOCES

1. Piece #1 test is tougher than state practical test for shop teachers, which is a five and one-half hour test. Students are in machine shop about two hours per day all year.

2. Piece #1 test is good for all grade levels (two grade levels are mixed in together for machine shop).

3. Make Piece #2 square so that lathe is not required prior to milling.

4. Make only one undercut on Piece #1.

5. Make tapered arbor out of Piece #1. Standard taper could be made instead of listed taper.

6. Eliminate the 3/4" thread and make the small thread a turn thread to "create a toolmaker rather than a machinist."

7. Many of the constructive criticisms [discussed prior to the field test] were eliminated from my mind once I received the well-written instructions on the testing procedures.

8. The objectives built into the drawings are excellent, with the possible exception of Piece #2 (Note #3), being actually used in the shop, due to safety reasons.

9. The objectives pertaining to cutter grinding can be achieved through the grinding of standard type cutters.

Revision Recommendations

The field test brought out the need for several revisions. Items 2 and 7 on Piece One and Item 3 on Piece Two should be redesigned. The grading items should be better structured and should be included as part of the total test. Improved instructions for timing and grading the performances must be written (perhaps by a machine shop teacher); possibly the instructions should be recorded on videotape.

Before closing this report two concerns must be discussed. The first is with the potentiality of creating different forms of the test by simply changing the metal provided for the test item operations. It is conceivable that parallel forms of the test can be created by changing stock or by changing machinery upon which the operations are performed. The change of materials is preferred, since the change of equipment may threaten the validity of the test in measuring the output of instruction.

The second concern is with the centralized scoring kit. There appear to be two ways such a kit could be constructed, with a third option being that of the scoring center used during the field test. The first possibility for centralized scoring is the use of plastic measures color-coded to facilitate reporting and to disguise the correct response measure. These plastic pieces would be inversely shaped to the performance test products. For each test item, several plastic measures would be included in the kit: a red piece at the lower tolerance threshold, blue at the upper tolerance level, yellow in the center, green slightly below tolerance, and white slightly above tolerance. The student would be instructed to try each plastic piece until one fit his product; the color of the piece fitting would be recorded and keyed for scoring at a central location.

The second possibility is similar to the first, except that it makes use of the tools normally used for measurement in the shop. A set of calipers would be customized to measure only the tolerances of each task. Each caliper would be numbered, and the student would report the number of the smallest caliper setting which fit the cut. Keyed caliper codes could be checked at the central office to determine that the measured size was within tolerance. Either of the two scoring forms described above would apparently be feasible, with the plastic keys being preferred because of the advantage of simplicity.

Future development of the machine shop test should investigate the low reliabilities in a controlled study and experiment with the centralized scoring models. However, prior to any further work a more detailed statement of mission should be formulated so that future studies can continue the development of the tests within those frameworks in which they will be most often used.