JOURNAL OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

VOL. 4, NO. 2

CONTENTS

JOURNAL
OF THE HP 3000 INTERNATIONAL
USERS GROUP, INCORPORATED

JOURNAL OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

TEST EVALUATION

Dr. Lloyd D. Davis, Director
Academic Computing Services
Univ. of Tennessee at Chattanooga
Chattanooga, Tennessee 37402

Introduction

Every teacher who constructs his/her own tests is faced with the dilemma of building a test that measures student progress and that also can be claimed to be objective. Such objectivity inherently has the attributes of validity and reliability, which will be discussed later in this paper.

Tests, of course, contain questions or items. These items can be examined in terms of difficulty, discrimination, and their individual contributions to the test as a whole. The end purpose of many tests is, of course, the assignment of grades; therefore, several statistics which deal with measures of central tendency and dispersion are discussed. With these test measurements and evaluation tools, the common classroom test can be improved and made more reliable, and the confidence with which the teacher administers the test can be increased.

Reliability

Any discussion of test evaluations includes the area of test reliability. Reliability (5:172) describes the extent to which measurements can be depended on to provide consistent, unambiguous information. Consistency of measurements means high reliability since chance or random factors are removed or controlled. The correlation coefficient can be used to determine the amount or degree of consistency.

Test reliability can be determined in several ways. Correlations may be computed using scores of students who take the source test at two given times, or who take two parallel test forms concurrently, or who take two different parallel forms at different times. Of course, not all these methods would be appropriate in all cases. The three methods - stability, equivalence, and stability and equivalence - are known as the techniques of reliability.

There is also a technique in the area of reliability known as internal consistency that may be applied either among test items or to a single test. For most tests of homemade construction these are the easiest to compute and evaluate. The primary statistical analyses for reliability are the split-half and Kuder-Richardson formulas.

Split-Halves

In the split-half method, each student has his responses divided into categories or halves. Although one might use the first and second half, the more common method is to use the odd numbered questions versus the even numbered questions. This technique is preferred since most tests are progressively harder as one moves through the test. Hence, for each student his/her score on the odd numbered questions and the parallel score on the even numbered questions are used as paired data; the data for the whole class is then correlated by computing the correlation coefficient (r).

The formula known as correlation coefficient is (6:80):

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x^2)}\ \sqrt{n\sum y^2 - (\sum y^2)}}$$

A well known fact is that (6:15) test validity is increased by increasing the number of test items. When r is computed as above, the length of the test is halved. Hence, r above must be corrected for its original length. The new R (to distinguish it from the original r) is known as Spearman-Brown and is computed by the formula (1:17):

$$R = \frac{2r}{1 + r}$$

As an example, if r from halves is .70, Spearman-Brown will be .82 and should be reported as the split-half coefficient. One caution is that the split-half method should not be used on pure speed tests since spuriously high reliabilities will be derived.

JOURNAL
OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

Kuder-Richardson

Unfortunately, one is never certain where the best split might be; the split halves compress the odd numbered versus the even numbered responses. The Kuder-Richardson tests overcome this difficulty by considering all possible splits. There are two such tests in common use, namely KR20 and KR21. Their formulas follow and may be found in (5:181) and (1:17):

$$KR20 = \frac{n}{n-1} \cdot \left(\frac{S^2 - \sum pq}{S^2}\right)$$

$$KR21 = 1 - \frac{M \cdot (n-M)}{nS^2}$$

where:

- n is the number of items.
- S squared is the variance.
- M is the mean.
- P is the proportion of students responding correctly on an item and,
- q is the proportion missing the item.

KR20 is probably most commonly used and hence the most widely known of the numerous Kuder-Richardson formulas. KR20 is also known as the Method of "Rational Equivalence" (2:340).

Reliability Comparisons

Both the split-halves (Spearman-Brown) and Kuder-Richardson formulas measure the internal consistency reliability, r, of a test. Neither should be used for a pure speed test. The Kuder-Richardson has problems when more than one trait or factor is being evaluated; hence, it should be avoided for multiple factor tests. All formulas yield values between 0 and 1 with high reliability being close to 1 and low reliability being close to 0. The following test (1:17) outlines the relationship, between reliability and decisions relating to it:

Reliability Coefficient	Decision To Be Made
$r < .60$	Revise using item/analysis
$.60 \leq r \leq .80$	Reasonably acceptable
$.80 \leq r$	Very good reliability

Validity

Validity (5:206) is the extent to which measurements are useful in making decisions relevant to a given purpose. With respect to testing, this may be restated as:

> *validity is the faithfulness of a test to consistently measure the trait or factor it was designed to measure.*

Validity implies consistent, reliable measurement abilities and also the accurate measurement of the factors the test was designed to measure. An example may be helpful here. Suppose an accurate scale is set five pounds too heavy in its measurements. The scale is not valid since it is always too heavy. However, the scale is consistent since it accurately measures actual weight plus the consistent error of five pounds. Test validity is normally established by content analysis performed by subject matter experts. These individuals apply both subjective and objective evaluations to test items (questions) and the corresponding responses to these items. However, a statistical evaluation for the validity of a test may be done by correlating scores from the same individuals on both a test of known validity and the test being evaluated. If the known test has high validity and the two tests are designed to measure the same factors, then the experimental test, to have good validity, also must have a high positive correlation with the known test.

Sometimes a statistic called the Index of Reliability (2:349) or the Maximum Validity Coefficient (5:214) is calculated by the formula:

$$\text{Maximum Validity Coefficient} = \text{reliability}$$

Since the valid test is also reliable, this index allows one to determine the maximum likely validity for a test as a function of its reliability. For example, a test whose reliability is 0.49 would not have a validity greater than 0.70. Although this test places an upper limit on validity, the traditional methods for finding validity such as content or concurrent validity are preferred to this measure.

Item Analysis

The importance of item analysis in constructing and maintaini a valid test can not be minimized. Item analysis is the examination of the responses to a given item or question. Items may suffer from being (5:229-31) ambiguous, mis-keyed, too easy, too difficult, or nondiscriminatory. After inspection, items with an inordinately high miss rate should be examined for either obvious ambiquity or errors in marking the key. Items either too hard or too easy do not, in themselves, allow discrimination among those taking the instrument (test). For the most part, such a question does not give the evaluator sufficient evidence to warrant the item being included. If it is too easy, too many get it correct, so why give it? Similarly, if it is too hard, few get it correct, so why give it? Although this approach is simplistic, it is good practice not to have many either very easy or very hard questions but rather a few of both as well as a majority of questions of medium difficulty. This procedure will give a reasonable distribution of scores for most purposes.

Index of Discrimination

Given items to perform an analysis upon, the following procedure will produce an item analysis. First, score the entire test for all individuals. Second, order the tests from high to low on overall scores. Third, select the top 27% and the bottom 27% (4:17-24). Consider a test item. Score this item for the top group; score the bottom group on the same item. Calling the former T and the latter B, obtain the difference between the two (d). This difference, d, between T and B is the number more the top group had correct than the bottom group had correct on this item. By dividing d by the number n in either group, we get the index of discrimination (D).

Example:

1,000 students took a test; of the top 270 student scores, 210 had item K correct; of the bottom 270 student scores, 100 had item K correct. Hence, T = 210, B = 100, n = 270. Thus:

$$D = \frac{T-B}{N} \cdot \frac{110}{270} = .41$$

A chart of values and actions follows:

Index of Discrimination	Action
$D \geq .40$	Keep questions
$.39 \geq D \geq .20$	Revise question
$.19 \geq D \geq .10$	Discard question and rewrite
$.09 \geq D \geq .0$	Discard question
$.00 > D$	Discard question, bottom group does better than top group

There are many benefits accruing from computing the Index of Discrimination. Questions that negatively discriminate indicate the more successful group may be reading something into the question. Questions that discriminate well can be retained. Others that poorly discriminate are selected for possible revision. Although large numbers (N>370) are required for stability (5:236), many authors feel that the technique outlined is valuable even in smaller classroom meetings (N about 100) (6:84-6) or very small classes (N about 30) (6:86-8). One final comment about Index of Discrimination, D, is that the reliability of a test improves or gets larger with higher D values on items.

Index of Difficulty

The Index of Difficulty, p, or Difficulty Level (5:239) is defined as the proportion of students responding correctly to an item. That implies that the higher the difficulty level the easier the item. In fact, p varies from a minimum of zero where everyone missed the item to a maximum of +1 where everyone correctly answered the item. Since an item that is very easy or very hard has no discrimination power, there is an optional level of difficulty that is required for high discrimination indices. There are two approaches to the optional level of difficulty. The first is the Chance Score, (CS), which is the score between 0 and 1 due to chance. Hence, a multiple choice test with four items would have a chance score of .25. The second is a theoretical approach by Lord (3:181-94) which derives the optimal level for p by looking at distributions of item difficulties. A chart follows (5:240):

Number of Options per Item	Optimal Difficulty p Methods: Chance Score	Due to Lord
0 essay	.50	.50
2	.75	.65
3	.67	.77
4	.63	.74
5	.60	.69

The index of difficulty Is most useful in attempting to eliminate both very easy and very difficult questions. By having questions whose level of difficulty are in the range $.5 \leq p \leq .7$, the Index of Discrimination is optimal, giving much greater test reliability.

Summary Statistics

In describing a group of scores, two concepts seem most central in all descriptions. These are Central Tendency and Dispersion. Central Tendency is where the scores tend to cluster or fall equivalently where they are centered. Dispersion relates to how tightly they cluster about the center or how widely they are dispersed or distributed.

Central Tendency

The three principal descriptors of central tendency are the mode, the median, and the mean (2:27-35). The mode is simply the most frequently occurring score. For example, on a test with scores 10, 9, 8, 9, 7, 6, 9, 5, nine (9) is the mode, or the modal score is 9.

The median is often given for tests. It is the score such that exactly one-half of all scores fall below it. In other words, it is the center score. If scores are ordered from top to bottom, the median is either the middle score if the number of scores is odd, or it is the average of the middle two if the number of scores is even. An example follows. Consider the scores 10, 9, 9, 9, 8, 7, 6, 5. The middle two are 9 and 8; hence, the median is (9 + 8)/2 or 8.5.

The mean or third measure of central tendency is the most useful of all the measures of central tendency. The mean is computed by adding all the raw scores and dividing by the'number of scores. The formula is: Mean = x/n. Since the median is the middle score, it does not reflect the actual values of the other scores and tends not to reflect very well scores that are distant from the median. On the other hand, the mean uses all scores in its calculation. Hence, the mean is influenced by all scores, including those extreme ones some distance from the mean.

Dispersion

The question as to how scores are distributed about the center of tendency (median or mean) leads to a discussion of several measures used for that purpose (2:43-56). The simplest such measure is range. Range is the difference between the largest and smallest score. That is, range is the maximum score minus the minimum score. Hence, if scores are 4, 4, 2, and 1, the range is 4-1 = 3 (three). Sometimes 1 (one) is added to the difference in computing range so as to cover the span of all possible scores.

A second measure of dispersion is the Interquartile Range. If one computes Q1 and Q3, the first and third quartiles respectively, then the Interquartile Range is Q3-Q1 or the range of scores for the middle half of the class. A Semi-Interquartile Range is (Q3-Q1)/2 and is known also as Quartile Deviation or Q. For formulas see Garrett (2:43-8).

The major dispersion statistic is variance or its derivation, standard deviation. Standard deviation, whose symbols are either S or σ depending on whether you are working with a sample or population, has many applications in statistics and is the most widely used measure for dispersion. The formula for Variance is:

$$\frac{\sum(x-m)^2}{n} \text{ or } \frac{n\sum x^2-(\sum x)^2}{n^2}$$

where:

X is the set of observations (test scores),
M is the mean and
n is the number of scores.

The formula for Standard Deviation is:

$$S = \sqrt{\text{Variance}}$$

Applications to Test Grading

Using the mean and standard deviation, one can build distributions for grades based upon the mean and standard deviation and concepts such as the normal curve. For example, to get approximately 10% each of A's and F's, 20% each of B's and D's, and 40% C's, one would use M-1.3 S, M-.5 S, M+.5 S, M+1.3 S as cut points where M is the mean and S is the standard deviation. Therefore, if the scores had a mean of 60 and standard deviation of 10, the cut points would be 60-1.3 (10), 60-.5 (10), 60+.5 (10) and 60+1.3 (10) or 47, 55, 65, 73 for D, B, C, and A respectively. Of course, other distributions could be developed as needed, based in part on the summary statistics from such grade distributions. One such is found in the text by Blood and Budd (1:201):

A Any score +1.5 S or higher
B Any score between +.5 S and 1.5 S
C Any score between -.5 S and .5 S
D Any score between -1.5 S and -.5 S
F Any score -1.5 S or lower

where S = standard deviation and scores are measured from the mean.

Some schools use the following scheme for either small classes or those with non-normal distributions. Using the measurement Q, Quartile, grades may also be assigned by:

A Any score higher than 2 Q
B Any score between Q and 2 Q
C Any score betwen -Q and Q
D Any score between -2 Q and -Q
F Any score below -2 Q

The use of Q eliminates the extreme scores from entering into the cut point selections with its obvious good and bad points. Each user must in the final process make his grade evaluations on a number of criteria including those listed.

Conclusions

All aspects of education from system wide evaluation to single classroom testing can benefit from test evaluation. Improving validity through content analysis and reliability through item analysis, the test constructor or administrator can eliminate nondiscriminators, ambiguous questions, and miskeyed items. The University of Tennessee at Chattanooga has a test scoring package that resides on the HP3000. This package, which has been in use in some form or other at UTC for over a decade, provides test scoring, item analyses, including indices of difficulty and discrimination, and coefficients of reliability. Student scores are ranked for ease in assigning grades. The associated summary statistics such as mean, median and standard deviation are produced. UTC's test scoring package depends on input from an OpScan reader interfaced directly into the HP3000. Test scoring is important; at UTC we believe it to be important enough to provide and encourage evaluation via software.

References

1. Blood, D. F. & Budd, W. C. *Educational Measurement and Evaluation.* New York; Harper & Row, 1972.
2. Garrett, H. E. *Statistics in Psychology and Education* (6th ed.) New York; McKay, 1971.
3. Lord, F. "The Relationship of the Reliability of Multiple Choice Tests to the Distribution of Item Difficulties" *Psychometrika*, 1952, 17(2), 181-194.
4. Kelley, T. L. "Selection of Upper and Lower Groups for Validation of Test Items" *Journal of Educational Psychology,* 1939, 30(1), 17-24.
5. Sax, G. *Principles of Educational Measurement and Evaluation*, Belmont, California; Wadsworth, 1974.
6. Wood, D. A. *Test Construction.* Columbus, Ohio; Merrill, 1961.

JOURNAL OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

LINKING TWO TECHNOLOGIES THE HP3000 AND THE QUADRITEK 1200 PRINTER

John Clark
Computer Manager
Warrnambool Institute of Advanced Education
Australia

A modern micro-processor driven phototypesetter, a Quadritek 1200 has been linked to a Hewlett Packard 3000 computer system as an alternative to the installation of a word-processing system. By combining the use of the text editing and mass storage facilities of the HP3000 with the high quality variable type output and text arranging/page design capabilities of the phototypesetter, the benefits of each technology have been made available to a wider range of users.

Introduction

Warrnambool Institute of Advanced Education is a regional college of advanced education which provides advanced, technical and further education to meet the needs of 170,000 people living in an area of 50,000 square kilometres in the south west of Victoria and the south east of South Australia.

Student enrolment has grown rapidly from 170 in 1970 to more than 2100 in 1980. Because of the large area serviced, the Institute's External Studies program offers people study at a distance, no matter where they live or work. A substantial range and variety of study materials is provided, including printed notes, audio tapes and slides.

Within the Institute, the Computer Centre operates as a service centre providing all the computer facilities necessary to meet the academic needs of students and staff, and handles the information processing for the Library, Accounts Department, Student Administration and other areas of the Institute. A variety of computer equipment is used to provide this service. The majority of the processing is performed on a Hewlett Packard 3000 Series III computer, currently equipped with one megabyte of memory and 290 million characters of online disc storage. This system, which operates 24 hours per day, 7 days per week, is used almost entirely for interactive processing via a wide variety of remote devices.

During 1979 an investigation was made within the Institute to determine whether word processing equipment would be of benefit in the preparation of external studies lecture notes. Because of the large quantity of material to be keyboarded and stored it was decided to use visual display terminals online to the HP3000.

To obtain the quality of printed output required and the variety of type fonts desired it was decided to investigate the feasibility of connecting a phototypesetter to the HP3000. In November 1979 tenders were called for suitable phototypesetting equipment capable of communicating with the HP3000. A Quadritek 1200 phototypesetting system made by ITEK Corporation was selected and installed in February 1980.

Quadritek 1200 General Description

The Quadritek 1200 is a software controlled phototypesetter typical of the many units found in commercial printing firms. In these situations it mainly operates in a stand alone basis with all text data and commands being entered via its own keyboard and stored on either cassettes or flexible discs.

The hardware is all integrated into one unit and consists of:

- typewriter-like keyboard similar to that found on most visual display units together with some additional special function keys. A total of 79 keys in all.
- a 12 inch CRT screen displaying 16 rows each of 64 characters.
- a microprocessor for controlling the operation of the entire system. This is a PACE unit manufactured by National Semi Conductor; it has a cycle time of 500 nsec.
- a main memory made up from 32768 x 16 static RAM boards. Up to 3 memory boards may be fitted.
- either dual cassettes or dual flexible discs for offline storage.
- a data communication interface (RS232C) for serial input and output, asynchronous with speeds from 75 to 1200 Baud.
- an optical system which selects from four fonts each containing 112 characters and generates the 38 possible output sizes by use of a zoom lens system.
- an output unit producing right-reading photographic paper or film positives in widths up to 8 inches.

Operation of the System

The usual sequence of events in processing work through the system is as follows:

- the author of the material to be typeset provides the source document containing the text in which is interspersed the appropriate typesetting commands e.g. type font to be used, printing size, line length, etc.
- the text together with the interspersed typesetting commands is then entered into the HP3000 com-

puter via any computer terminal keyboard. The standard HP supplied EDITOR is used for entry, corrections and saving of the material under any unique file name.

- this file is then processed by a program which checks for any illegal typesetting commands, e.g. requesting a type font that is not available, and prints a listing to enable the author to detect any typing or spelling errors.
- When it is determined that the file is correct, a computer operator uses the phototypesetter as a terminal to "log on" to the HP3000 system and down loads the file to the phototypesetter flexible disc. Commands then entered at the phototypesetter keyboard cause the information on the flexible disc file to be processed by the electronics and released to the photo unit. This images and projects the text via a zoom lens system on to photo mechanical paper or film. This paper is then fed from the film transport area into a light tight-output cassette, which is removed from the unit for external developing and processing. The final processed output is a high quality, right-reading, camera-ready positive suitable for proofing, paste-up, and plate making.

Communication Between the two Systems

The phototypesetter is located in the Computer Room and linked by a 3 wire cable to the HP3000 asynchronous terminal controller in an identical manner to all other terminals. To initiate the downloading of a file from the HP3000 the computer operater loads the communication translation program from flexible disc into the phototypesetter memory. The phototypesetter then functions as a terminal and the keyboard is used to initiate the downloading of the file from the HP3000. The data from the HP3000 file, in ASCII, is scanned by the ITEK program. Any translations required to suit the operation of the phototypesetter are made before the data is stored on the flexible disc. Typical translations are the conversion of a carriage return to the typesetting command EL representing end of line, form feed being converted to the typesetter EP, and paragraph. ITEK supply a standard translation table and this can be modified by the user at system set-up time to meet any special requirements. The communication interface software is also user programmable to operate at any of the following speeds: 75, 110, 134.5, 150, 300, 600 and 1200 Baud. The user also defines the stop-bits and parity required.

When the two systems were first linked up at WIAE a transmission speed of 1200 was selected. However, because of the occasional appearance of a software timing problem when the translation program was unable to handle a certain pattern of ASCII codes the speed was reduced to 600 Baud. It is hoped that the release of a new version of the data communication interface software by ITEK will overcome this timing problem and enable us to revert to 1200 Baud.

Uses for the System

As mentioned in the introduction, the initial reason for investigating the feasibility of acquiring a phototypesetter and linking it to our HP3000 was to assist in a preparation of our external studies lecture notes. The storage of the lecture note material on magnetic media with the facility of amendment as revisions are required in the future, together with the ability to select from a wide range of sizes and a variety of type fonts without any operator intervention and still produce an extremely high quality output ready for offset printing were the major benefits sought.

During the evaluation of tenders visits were made to a number of users of this type of equipment (mostly small commercial printing firms) and further possible uses for the equipment became evident. Our current usage ranges across:

- the preparation of lecture notes.
- the printing of advertisements where we wish to have control of the layout and final appearance.
- the printing of a variety of notices, programs etc. for recurring events, e.g. examinations, graduation ceremonies, etc.
- the preparation of the Institute Handbook. This 100 + pages book is published each year with changes to approximately 20% of the material from the previous issue.
- (usage by) design students from our Diploma of Visual Arts course in lieu of Lettraset for supporting informational text for illustrated and photographic display work.

Conclusion

I believe the original objectives we set out to attain have certainly been achieved and the linking of the two technologies has enabled a number of other worthwhile uses to be identified.

The connection of the phototypesetter to the computer has created an awareness within the Institute that a computer can be of practical use to people other than scientists and accountants.

For those already accustomed to using a computer it has added a new dimension in terms of output available and has, therefore, widened the scope of possible use in their work.

For those people not normally involved with computing, e.g. art students and typists, it has opened the door to another world, which, I am sure we would all agree, is the world we must face in the future.

JOURNAL OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

SELECTIVE LOCKS ON MANUAL MASTER SETS

B. Kohn, Systems Analyst
Head Wrighton Stampings Limited
Hartepool (Cleveland), England TS25 2BP

I noticed in the First Quarter 1981 Edition of the HPGSUG Journal, Mr. G. W. Davison expressed the need to hold selective locks on manual master sets.

In the design of a locking strategy for a production control system for a drop forge this problem was encountered. It was necessary to lock a job code for order entry purposes or for invoicing but not for the recording of production.

In addition there were some transactions proceeding over several VIEW screens during which time it was considered desirable to prevent other users from simultaneously updating the same code as was currently being transacted.

To get around the problem of holding permanent locks within a database, a second data base was opened. This second database was permanently empty because IMAGE maintains logical rather than physical locks.

The empty database contained three master sets representing the three types of permanent lock required by application users. The DBLOCK call requested a lock on the single code within the master and then determines if the condition word reply is 25 i.e. the code is busy. If it is the user is told this but is permitted to transact work on another job code.

The advantage of this approach is that within the production control database itself, locks are held for only brief periods and may be automatically unlocked after the database procedure. Thus, users wishing to generate a report using QUERY do not find they have to wait for a long period until the database is no longer busy.

INCREASING EXECUTION SPEED OF IMAGE PROGRAMS

Johannes F. (Hank) Van Leeuwen
BASF Canada Inc.
5850 Cote de Liesse,
Montreal, Quebec. H4T 1C1
Canada (514) 341-5411

As many of you are aware, IMAGE calls which use the list parameter will execute more quickly if this parameter is changed to the current list (*;) construct after the first call for a data set (see page 5-43 of the IMAGE manual for more information). There is a simple method of implementing this technique in most existing planned data base programs without disturbing the normal program logic. To do so, proceed as follows:

1. As soon as you have opened the data base, perform a reread (DBGET mode 1) on each data set to be accessed using explicit item lists (or the "@;" list construct). Ignore error 17 (no entry) since the only purpose is to set up the current lists for each data set.
2. Change all lists to "*;". The rest of the program need not be changed assuming that the current lists will not be changed during program execution.

Example: A COBOL program accessing 2 data sets.

```
    PROCEDURE DIVISION.
    FIRST PARAGRAPH.
        CALL "DBOPEN" USING BASE PASSWORD MODE1 STATUS.
        IF DBERROR IS NOT EQUAL 0 GO TO DB-EXPLAIN.
        CALL "DBGET" USING BASE SET1NAME MODE1 STATUS SET1LIST
Extra      SET1AREA ARGUMENT.
Code!   IF DBERROR IS NOT EQUAL 17 GO TO DB-EXPLAIN.
        CALL "DBGET" USING BASE SET2NAME MODE1 STATUS SET2LIST
           SET2AREA ARGUMENT.
        IF DBERROR IS NOT EQUAL 17 GO TO DB-EXPLAIN.
        MOVE "*;" TO SET1LIST SET2LIST.
       .
       .
       .
     (rest of program)
```

JOURNAL OF THE HP 3000 INTERNATIONAL USERS GROUP, INCORPORATED

THE FIELD SOFTWARE COORDINATION PROCESS

Brian M. Perkin
Canadian Software Coordinator
Hewlett-Packard (Canada) Limited
2670 Queensview Drive
Ottawa, Ontario
Canada K2B 8K1

ABSTRACT

This paper will discuss the history of, and the future benefits of Hewlett Packard's field software coordination program.

Over the last few years, growth of the Hewlett-Packard 3000 customer base, through strong sales and successful installations, has caused substantial growth in Hewlett-Packard manufacturing divisions and in the field support operations. In addition, Hewlett-Packard has developed many new and innovative software and hardware products. Hewlett-Packard has (for various reasons) divided up manufacturing divisions into smaller, more manageable divisions and operations. These divisions are distributed throughout North America and Europe.

With this exciting growth, Hewlett-Packard is still very concerned about its customers. We want to be sure that a product will work currently and reliably the first time it is installed on a customer site. Many of these products have complex interactions with other products, making quality assurance a difficult task. In order to maintain Hewlett-Packard's high standard of reliability a new program has been created in co-operation with the Computer Support Division.

Software Coordinators have been appointed for each sales region. Their job responsibilities are very loosely defined. Generally, a software coordinator will use local and factory resources to identify potential problems and weaknesses, verify problem solutions, and to act as a integration and distribution point for new software. Software coordinators also have responsiblity for training and acting as a resource person to other Hewlett-Packard Support groups.

PAPER

In November 1980, Hewlett-Packard introduced a reorganized structure for its HP 3000 program. As a result of high sales, growth in the installed customer base and aggressive new product development, the old General Systems Division split into several divisions and operations:

Computer Systems Division (CSY)
Information Systems Division (ISD)
Data Communications Operation (DCO)
Manufacturing Systems Operation (MSO)
Boeblingen General Systems Division (BGD)
Computer Systems Pinewood (CSP)
General Systems Division (GSD)

With the exception of the new General Systems Division, the above divisions are primarily involved with the HP3000 product line (the new General Systems Division handles the HP 250 and its follow on products). To coordinate the marketing functions of the divisions and to ensure product compatibility, the Business Computer Group was created. A reporting structure from each division to the group exists.

Computer Systems Division has responsibility for all HP3000 hardware and operating systems as well as system utilities (like Listdir2) and non RS-232 device drivers.

Information Systems Division has responsibility for all compilers, language subsystems, data management and data entry tools, business graphics, laser printer support software as well as performance and optimization tools.

Data Communications Operation handles all synchronous and ansynchronous data communication products as well as terminal drivers.

Manufacturing Systems Operation develops application products for the HP3000. Their primary thrust is software products for the manufacturing industry.

Boeblingen General Systems Division (Germany) manufactures certain members of the HP3000 family for European markets as well as developing data capture software and customizing of Hewlett-Packard software to meet European language and business needs.

Commercial Systems Pinewood (England) presently has responsibility for our text formatting software as well as the development of future products for use in offices.

Each of the above divisions have their own labs, lab engineers, product management and marketing teams. See Table 1 (following) for the products belonging to each division.

Table 1
Product Breakdown by Division

Computer Systems Division:
MPE - Series II/III, MPE - Series 30/33, MPE - Series 44, any software in the System Utilities Manual.

Information Systems Division:
COBOL, COBOL II, BASIC, RPG, APL, SPL, KSAM, V/3000, Image, Query, DEL, DSG, XSORT, Scientific Library, Calcomp Plotter Library, Trace, Editor, Fcopy, Sort-Merge, IFS and IDS (Laser printer support software), Compiler Library, Performance Tools.

Data Communications Operation:
IML, MRJE, MTS, RJE, DS, CS, Terminal Drivers.

Manufacturing Systems Operation:
MMS/3000, MFG/3000

Computer Systems Pinewood:
TDP/3000

Boeblingen General Systems Division:
Datacapture Procedures

Testing a new product or an enhancement to an existing product is an interesting process. Long before any code gets written,a long planning stage occurs. In this process, several things are discussed. Funding and priorities are based on customer demand as well as the general marketing direction of Hewlett-Packard. At this time a document called an External Reference Specification is drawn up. This document contains a description of the product features, a targeted completion date, an estimate of the resources needed to complete the project as well as an assessment of what impact (if any) this product will have on other parts of MPE. Sometimes a seemingly simple enhancement will span several project teams and divisions. (An example of this is the implementation of Private Volumes, which required modifications to the File System, Loader, Disc Drivers, Console Interface, Command Interpreter, Initial, Sysdump and others).

Once this document is completed and the lab team is identified, the product is developed. If this is a product enhancement, testing is done to determine if the product enhancements work as specified and that the enhancements do not adversely affect how the product used to work. If this is a new product, more stringent testing occurs. When the lab team believes the product is almost ready for announcement, it is installed at what we call an Alpha test site. These sites are usually internal sites within Hewlett-Packard. Generally, these sites are located in one of our corporate EDP departments, or in a division which uses HP3000s to control their financial or manufacturing operations. The objective of this procedure is to give our products a good shake down in a production environment. This enables us to enhance or correct our documentation, remove any features, bugs or oversights from our software as well as to get comments on the overall quality of the new product. Since the communication loop is very strong between the lab team and the Alpha test site, a multitude of errors, omissions and oversights can be identified, and resolved in a very short time cycle. Once this aspect of testing is completed, we move on to the next stage of product testing.

Beta testing is when we take the resultant product from our Alpha testing and install it on carefully selected customer sites. The criteria for becoming a Beta test site include having an excellent product understanding, having the necessary resources to not only identify problems, but to help Hewlett Packard in documenting and solving them, having a strong Hewlett Packard field team nearby to assist in the above, and having enough flexibility in the production environment to be able to withstand the stress that an experimental software product may place on it. In the Beta test site, much of the same testing goes on as in the Alpha test site. The customer agrees that in exchange for being given a pre-release of some product, the product will be tested in as many of its aspects as possible. Hewlett-Packard has found this to be an excellent program. It enables us to release a relatively error-free product, with correct documentatation, on time. We also have a good picture of the resources necessary to support the product both in the factory and in the field.

Once we have gone through all this testing, we are now ready to announce the product. A tape known as the product tape is created with the software on it, the corrected documentation is sent to the printers, and support guides are written for the field organizations. When all this is done, the product goes on our price list and is announced.

Once we have developed a new product or an enhancement, we then integrate this to produce what Hewlett-Packard calls an "IT" tape. This Installation Tape is the tape from which customers' software is normally updated. It has been Hewlett-Packard's goal to produce 2 major updates and 2 minor updates per year, hence producing 4 "IT" tapes per year. To understand this process, I would like to digress to discuss what we call the "IT" cycle.

About 4 times per year, freeze dates are set in our labs, after which no further changes are allowed to be made to MPE or the subsystems. The system integration group takes all available enhancements, product tapes and the old subsystems and puts them together to build a new "IT" tape. This tape is installed on several systems in our quality assurance area for reliability testing. This software is run under heavy load with predefined scripts in order to determine if this combined set of new software will run reliably when it gets to a customer site. The decision point is at about 200 hours continuous running time without failure. If a failure occurs, the subsystem is identified, and the clock is set back to zero. The lab team whose product has the problem then has a decision to make. They can either remove the enhancement from this release or repair the problem if it is not too difficult. Once this decision is made a new product tape is resubmitted for integration. Then the testing starts again, from the beginning. Once this release successfully passes the complete Q.A. run, we then have a Manufacturing Release. It is important to note that the other lab teams are busy developing their next release of a product while this process goes on. As a result, a significant amount of parallelism occurs between the lab development groups and the Quality Assurance people.

It is easy to see that if one release spends a long time passing Q.A. and the next release passes Q.A. fairly quickly, then your account S.E. will have two updates for your system in a fairly short interval. Sometimes we will scrap the earlier release if that is the case.

We often have questions about why some Service Requests listed in the Software Status Bulletin take so long to be fixed. One of the major reasons is that most of these bugs are fixed in our lab by the lab engineer who is the most familiar with the product (and hopefully the problem). Considering the time delay between when a bug is discovered, verified and submitted and also the delays of Quality Assurance, shipping and installation in the field, by the time the lab engineer gets to look at your new bug, he may be working on a version of the subsystem that is 2 or 3 releases newer than the one you were using. As a result, your bug sometimes doesn't get fixed until 2 or 3 "IT" tapes later. However, part of this integration process ensures that fixing your bug will not create another problem. Also the Software Tracking and Reporting System ensures that the manuals will get updated to take note of the changes created by the fixing of your bug.

We have discovered that upon some occasions, we are unable to synchronize all of our labs to the "IT" cycle. When this happens, we release product tapes in addition to the 4 "IT's" per year. All of a sudden, we have all kinds of software coming in from all different directions. In the past few months we have seen several product tapes released for various reasons. Some have been brand new software products, software to support new hardware products and enhancements to specific products. Since many of these products interact with each other, it is not always clear whether all the products will work in conjunction with each other. This is where the Field Software Coordinator (FSC) comes into the picture.

I will digress now, to explain what a Field Software Coordinator is and then explain what he/she does. Today, there are about 70 FSC's throughout the world. About 17 of them are in North America, one per Hewlett-Packard Sales Area. Generally, FSC's are among the most experienced System Engineers and Technical Support Engineers from the field. Most of them have been involved with the Hewlett-Packard 3000 program for some time in various capacities. Their varied experience includes performance consulting, developing and teaching various training courses, supporting major accounts, manning PICS centres, and acting as technical resources for their respective areas. In most cases, the FSC's have become very familiar with how both MPE and the HP3000 hardware work, as well as the finer points of the inter-relationships that exist between the two.

A Field Software Coordinator's job responsibilities are loosely defined. I like to think that my duties are to do anything within my power that will make a set of software that is being installed on a customer's site more likely to run reliably and be easier to support should something go wrong. There are two or three specific aspects I would like to touch upon.

One of the most important areas of a FSC's responsibilities is definition of "standard" Installation Tape. After the "IT" leaves the factory, there are sometimes unforeseen problems discovered. Some of these are of a critical nature, while others affect one or two sites worldwide. All of the software coordinators are kept abreast of these problems as they develop by maintaining regular contact with several key factory people as well as nearby Field Software Coordinators. At some point, each FSC decides which problems will be fixed in his "IT". An integration process then takes place creating a "standard" Installation Tape for his area. This process entails using a HP3000 in a stand-alone mode for a period of between a few hours and a few days. The factory release of software is installed upon it and various new modules, fixes and workarounds are installed on top. When this process is finished, the new "IT" is created. In parallel with this operation, a large amount of documentation is developed which is later made available to the various field support organizations. This new set of software is then run under test in several Hewlett-Packard Demo Centres in order to ensure that it will run as anticipated. If the FSC has any special concerns, he may install this software at selected customer sites for further testing, prior to making it generally available to the customer base. This field testing cycle takes as long as deemed necessary by the individual software coordinator. At the completion of this cycle, the FSC announces the software to be generally available and installable area-wide.

There are many other things to keep the FSC busy when he is not creating new sets of software. One of the activities is keeping track of what he has done. A set of documentation is created, which lists every change made to the software. The kinds of information kept here are subsystem names, problem descriptions, index numbers into the STARS system, an assessment of the severity as well as which combinations of software and hardware the problem occurs on. These records are distributed to various places including your account SE, PICS Centres, and the Technical Support Organization. Also these records become part of a permanent archive.

FSC's also act as resource people to the area. They assist in resolution of hot sites because they are often very familiar with the software run at that site. Because FSC's deal with various divisions and field offices, they come upon a lot of information that isn't always too useful to them, but may be to someone. I have developed some strategies to handle this information management problem. I write several newsletters per year containing the tips I have run into. These tips will often enable someone at Hewlett-Packard to do a better or faster job or to simplify an installation. Since so much of this information reaches me, I also maintain a telephone message device, which makes the newest information available to the caller without my intervention.

The last duty of the FSC, that I wish to mention is that of teaching and coaching. Since the FSC's generally have a very strong background on how the software and hardware products work, they often get involved in training throughout Hewlett Packard. This includes workshops held when a new operating system (for example) is released as well as substantial amounts of coaching over the phone.

I trust that the above paper has been informative, and that I have shown how Hewlett-Packard is preparing for the software needs of today and tomorrow.

The 1981 *JOURNAL* Publication Schedule, as given below, includes the deadlines for receipt in the Los Altos office of articles submitted for publication. Also given are the months of publication.

HP 3000 IUG JOURNAL PUBLICATION SCHEDULE

Article Submission Date	Quarterly Issue	Month of Publication
	1981	
November 15	Vol. IV. #1	January
June 15	Vol. IV. #2	September
October 15	Vol. IV. #3/4	December
	1982	
January 15	Vol. V. #1	March
April 15	Vol. V. #2	June
July 15	Vol. V. #3	September
October 15	Vol. V. #4	December

Send your contributions c/o Editor, HP 3000 International Users Group, Inc., 289 South San Antonio Road, Los Altos, California, 94022, U.S.A

JOURNAL

OF THE HP 3000 INTERNATIONAL
USERS GROUP, INCORPORATED

Journal
Administrative Offices
HP 3000 International Users Group, Inc.
289 South San Antonio Road
Los Altos, California 94022
U.S.A.

Ruth Ann Barrett, Association Manager
Tel. (415) 941-9960
ADDRESS CORRECTION REQUESTED
RETURN POSTAGE GUARANTEED

BULK RATE
U.S. POSTAGE
PAID
Permit No. 656
Los Altos, Calif. 94022

A01648 RH/P1 SP OCT 810057
Linford Hackman
Xybion Medical Systems
38 Garwood Terrace
Denville, NJ 07834