PROCEEDINGS of the HUMAN FACTORS SOCIETY—26th ANNUAL MEETING—1982

TESTING SMALL SYSTEM CUSTOMER SET-UP

James R. Lewis

International Business Machines
System Products Division
Boca Raton, Florida

ABSTRACT

Customer Set-Up is a proven approach to reducing service costs and providing products at lower prices to customers. To ensure the effectiveness of Customer Set-Up instructions and procedures, these instructions and procedures must be studied before being shipped with their associated product. This paper will address several points to consider when planning a study of a Customer Set-Up system, such as procedure, appropriateness of subjects, number of subjects, the iterative procedure, studies vs. test, and development of test criteria.

INTRODUCTION

The price of a computer system can sometimes be substantially reduced by designing the system to be set up by the customer. This procedure is referred to as Customer Set-Up (CSU).

Cost effective Customer Set-Up is accomplished by the proper trade-off of hardware and manufacturing cost against the resulting service cost and price. However, if the CSU instructions are too complex or not suitable to the customer audience being addressed, the customer will either be unable to or will refuse to set up his system. To ensure that CSU procedures are reasonable, they must be studied before they are shipped with a system. This paper will address a number of the recommendations for studying CSU developed as a result of the human factors effort with the System/23 Datamaster.

PROCEDURE

The procedure for studying CSU is a very simple observational method. Subjects are given the system to be set up, the set up instructions, and are asked to set up the system. Observers watch the subjects perform the set up task and record the occurrences of problems. The goal of CSU study is to identify and recommend alternatives to difficult portions of the CSU.

SUBJECTS

Appropriateness of Subjects

There is a great temptation in industry to test CSU on subjects whose primary advantage is that they are readily available, since they work for the company developing the product. In addition to availability, this practice alleviates the problem of security.

Unfortunately, this practice is reasonable only when employees selected as test subjects closely match the defined customer audience. Otherwise, any observations made concerning CSU are hopelessly confounded with differences between the subjects and the defined customer audience. For an audience defined as the "general public", employees of the company developing the product are not likely to be appropriate.

During the development of the System/23, the subjects for the studies were obtained from temporary employment agencies. Important subject characteristics, such as typing speed, age, and previous experience with computers were specified when arranging to hire from the employment agencies.

Number of Subjects

The binomial probability theorem can be used to determine the probability that a problem of probability p will occur r times during a study with n subjects.

For example, if an instruction will be confusing to 50% of the user population, the probability that one subject will be confused is .5. If two subjects are observed, then the probability that either one or both subjects will be confused is .75; and if three subjects are observed, the probability that at least one of them will be confused is .875.

The above example assumes that one occurrence of a problem will result in the problem being recognized as a problem. When interpreting the results of the study, one may not be satisfied with concluding that a problem exists based on one occurrence of the problem.

It is clear that more subjects must be observed if one wants the assurance of seeing problems twice before deciding that the problem must be examined. For example, if an instruction will be confusing to 50% of the user population, the probability that two subjects will be confused when two are studied is .25. If three subjects are studied, the probability that at least two subjects will be confused is.50. If six subjects are observed, the probability that at least two subjects will be confused by the instruction is .89.

From this reasoning, the recommended minimum number of subjects depends on the number of times a problem must be observed before it is regarded as a problem and the magnitude of the proportion of the user population for which one wishes to detect problems. If every difficulty that subjects experience is seriously treated as a problem, and one wishes to have a reasonable chance to detect problems that will happen to 50% of the user population, then at least three subjects should be observed. If all other conditions are the same, but problems have to occur twice to be regarded as problems, then at least six subjects should be observed.

During the development of the System/23, CSU studies involved the observation of at least five subjects, and a single occurrence of a problem was enough to indicate that the situation warranted investigation.

ITERATIVE PROCEDURE

An investigative technique that has proved to be very useful in product development is the iterative or 'test-fix-retest' procedure. This method has been successfully used in the development of a tutorial for the IBM 3277 Display Station (Al-Awar et al., 1981) and the IBM System/23 operator training as well as in the study of System/23 CSU.

The methodology of the iterative procedure is as follows:

1. Develop a product (or part of a product) to the point at which the iterative procedure will be implemented.

2. Observe members of the target population use the product.

3. Alter the product based on the results of the previous observations.

4. Repeat steps 2 and 3 until a satisfactory level of performance is reached.

Because "temporary fixes" can be easily incorporated into most CSU instructions, it is possible to effectively use the iterative procedure in the development of CSU.

Ideally, this procedure should continue until a large number of subjects complete the task with no difficulty. Practically, this may not be feasible due to time and economic constraints. Following the rationale presented in the preceding section, it is recommended that at least six subjects successfully complete the CSU with little difficulty before examination of the CSU is stopped.

STUDY VS. TEST

In this paper, a CSU study is defined as the procedure of observing the behavior of subjects under controlled conditions in order to identify operational difficulties associated with the CSU. A CSU test is visualized as a subset of the general category of study, and is characterized by the definition of criteria with which the CSU being assessed will be judged on a pass/fail basis. Tests are typically conducted toward the end of a product's development cycle (as in an IBM Systems Assurance User Oriented Systems Test), while studies should be conducted throughout the entire development cycle.

TEST CRITERIA

The problem of defining test criteria is an important consideration in the development of a CSU test. Some approaches to the solution of this problem are:

1. The test designer examines the tasks to be performed and recommends criteria.

2. Criteria can be defined as 1.5 (or some other number) times performance values obtained from experienced operators.

3. Criteria can be defined based on historical data obtained from previous tests in which the task was performed.

4. Criteria can be arrived at through negotiation with the groups who are responsible for the product.

Each of the approaches to the definition of criteria has its drawbacks. When a test designer recommends criteria (Approach 1), the definition has no objective basis. In addition to this, a criterion that the test designer believes is reasonable may not be viewed as reasonable by someone else in another development group. Approach 2 has the advantage of being objective in its utilization of the average performance of experienced operators, but fails in the same way that Approach 1 fails in that there is no objective basis for the selection of the value of the multiplicative constant. The use of historical data (Approach 3) is feasible only if such historical data is available. Even if the data is available, a number of additional assumptions must be made. For instance, the historical data will be applicable only if the data were collected under similar circumstances on a similar machine or system. Again the problem of objectivity must be confronted. How can one objectively assess the similarity of circumstances, machines, or systems?

For the present time, the best approach is probably Approach 4. It is not objective, but each group will have an opportunity to express its approval or disapproval of the criteria suggested by the test designer. One can use the average performance by experienced operators (as in Approach 2) to determine the lowest reasonable criteria if a machine or prototype is available. As an alternative approach, historical data from previous tests can be used to arrive at an initial estimate of the criteria (as in Approach 3) if such data are available. In lieu of these procedures, the test designer will have to make his best guess at reasonable criteria and assume that the negotiation process will result in the adoption of acceptable criteria.

REFERENCES

Al-Awar, J.; Chapanis, A.; and Ford, W. R. ‘Tutorials for the first-time computer user.” IEEE Transactions on Professional Communication, PC-24 No. 1 (1981), pp. 30-37.