

Development of a Testing Service System

Catharina C. van Thiel and Michel A. Zwarts

Institute of Educational Measurement, The Netherlands

The development of an integrated system for the storage of items and the construction and analysis of tests is described. The system is being developed both as a general facility for the Dutch Institute of Educational Measurement and as a support system for the use and maintenance of item banks in schools. The methodology of developing the system is described with attention to the system architecture and to the results of the first stage of the system development.

A testing service system (TSS) or item bank system is an integrated system for the storage of items, the construction of tests, and the analysis of scores. For developing such a system it is not sufficient to have access to a computer, knowledge of a test theory model, and good programming skills. It is essential that a system development methodology be used.

This paper describes the first stages of the systematic development of a TSS by the Dutch Institute of Educational Measurement (CITO). In the first section arguments are presented for a systematic development of information systems such as TSS and for the choice of ISAC, the particular system development approach used in the project. Then, the different stages (or areas, as they are called in ISAC) and the resulting TSS system development process are described, primarily from a systematic

(i.e., computer science) perspective. The concluding section describes the influence of the method chosen on the development and the results.

Choice of a Method

The development of systems like TSS can take place in many different ways. Such systems are essentially information systems (i.e., people cooperating in the processing and transmission of information). For the construction of information systems, several systematic approaches have been developed over the past 35 years. The use of such an approach has many advantages:

1. The process of development of the system is controllable. The entire process is divided into steps, the end of each step giving the opportunity to monitor progress, efficiency, and economy.
2. During the development, documentation is generated more or less automatically. Documenting the system is not a separate activity, but an integrated element of the development.
3. A standardized approach is important for ensuring successful and efficient communication between the parties and persons involved in the development.
4. The possibilities for extending the system as well as the flexibility of the system increase; in other words, the system will be of higher quality, which results in longevity and low

costs for maintenance and extension.

5. The risk of design errors is reduced. Repairing design errors is, as a rule, more and more expensive as system development progresses.
6. The design activities can be divided into manageable parts (decompositions). This allows the designers to control the complexity of the system.
7. Initially, the development of the information system is independent of the technology available. Only in a later step—that of the technical design—does adaptation to the hardware take place, which simplifies the problem of adapting to constantly occurring changes in the equipment.

Although there are several methods that can support the development of information systems, very few meet the criteria of a project like TSS in which several disciplines must be made to work together. These criteria are:

1. The method should be quick and easy to learn; it should involve the use of as few specific symbols and interpretation rules as possible.
2. Sufficient material and literature on the method should be available in order to simplify a first acquaintance with it.
3. External information sources (i.e., data from other users of the same development method) should be available. This is of essential importance in case of unanticipated problems.

The ISAC design method (Lundeberg, Goldkuhl, & Nilsson, 1981) meets these requirements. ISAC was developed at the University of Stockholm and is short for Information System Work and Analysis of Change. In contrast to most other development methods, ISAC does not start from the information system to be constructed, but from the activities that need support; therefore, its use precludes the construction of information systems that either insufficiently fit the corresponding activities, or do not solve the problems for which these systems were primarily developed. ISAC effectively supports the design and construction processes. It gives equal attention to the processes and to the data involved in the processes; it also takes account of both the size and the profoundness of the system.

In addition, ISAC links up with existing methods that can be used for the development of software (methods for structured programming and testing).

It should be noted that the choice of a method offers no guarantees with respect to the quality of the information system that is eventually developed. Every method, including ISAC, is only a way of looking at the problems, analogous to a pair of glasses which can be worn or removed as desired, depending on the situation.

System Development

As stated above, ISAC starts with the investigation of the needs, problems, and ideas of the users, and it ends with the specifications for manual procedures and computer programs. The activities within this stage can be divided into various areas, as shown in Table 1.

The activities contained in each of these areas and the products of each area will be discussed below. It is clear, however, that the areas are not separated very strictly, because this is an iterative process. During a particular area new ideas and knowledge are acquired, which may render obsolete certain parts of areas covered previously, making a redefinition necessary.

Change Analysis

The aim of change analysis is to investigate the improvements which might be pursued in order to solve, or at least reduce, the problems and needs that have been observed. Social, human, and eco-

Table 1
 Scope of Activities and ISAC Areas

Scope of Activities	Areas
1. Condition setting and system definition	Change analysis
2. Problem orientation	Activity studies Information analysis
3. Data orientation	Data system design Equipment adaptation

conomic factors play a role in this analysis. In this phase the existing organization is analyzed, and attempts are made to locate the problem areas. These problem areas may occur not only in the area of information technology, but also in fields such as management and organization. The steps within the change analysis are (1) analysis of the current situation, and determination of the problems and needs; (2) study of change alternatives; and (3) choice of the change approach.

ISAC describes the results of the change analysis primarily through tables and diagrams. The tables include a problem table, a list of interest groups, a table of objectives, and a table of needed changes, in which the desired improvements are recorded. The existing situation and the proposed improved situation(s) are recorded by means of so-called A-graphs. These graphs are defined by means of the following method, which is common in systems engineering. First, the boundaries of the system are determined, and the system's input and output are investigated without looking at the functions of the system itself (black box approach). Then the content of the black box is determined, and, by means of A-graphs of a lower, more detailed level, the system itself is described. Also part of this ISAC procedure is to check whether the goals set for the system contribute sufficiently to the solution or reduction of the problems mentioned in the problem table.

Activity Studies

Activity studies begin by determining the information systems required, and by dividing them into subsystems. ISAC provides very few guidelines for dividing an information system into subsystems, although some were provided by Bemelmans (1984). Apart from similarity of functions and output demands, he also mentioned the independence of data and processes, and the practical feasibility ("makeability") within an acceptable time-span.

Next, the information systems identified in this way are analyzed. An important part of this analysis is the determination of the ambition levels of each system. An ambition level of a system or

subsystem depends on the system's requirements for optimum functioning and on the demands of the user. Higher demands of the user are mostly related to higher requirements of the system. The system requirements can be translated into requirement factors. In an ambition level table, the possibilities for the user are mentioned for each factor. These possibilities are grouped according to ambition levels of the system or a subsystem. Finally, the ambition level is chosen that is most acceptable and, obviously, most economical. It is important to maintain a consistent choice of ambition levels across different subsystems.

Information Analysis

Information analysis begins with the performance of precedence and component analyses. For the precedence analysis, the output desired is the starting point. The next step is the identification of the data comprising each type of output (i.e., report, expected information inquiry). This is followed by a consideration of which datasets are needed to construct the different types of output. Thus, the information precedence relations are analyzed. The precedence analysis is recorded by means of information precedence graphs, the so-called I-graphs.

In component analysis a detailed analysis is made of each data component (i.e., output) or intermediate dataset involved in the precedence analysis. In other words, an investigation takes place into which objects require data, and which attributes are used in describing the object. The findings are recorded in C-graphs.

The next step in the information analysis is the process analysis. In contrast to the precedence and component analyses, which are only concerned with data, this step focuses on the processes necessary to produce the final products, and the conditions under which the processes must be carried out. The last step is the analysis of performance requirements, or the property analysis, in which the ambition level chosen in the activity studies is tested in light of the newly obtained insights.

Data System Design

The aim of the data system design is the development of an information system data model which is independent of the equipment. This means that there are still manual parts and parts to be automated in the model. A data system is a system that collects data and maintains datasets. Its design entails consideration of aspects such as maintainability, flexibility, reliability, and the processing philosophy chosen. The data system design is identical with the technical system design only if it is restricted to the parts to be automated. Decisions are made with respect to processing mode (transaction mode or batch mode), storage method (sequential, direct, etc.), processing location (centralized, decentralized), and the moment of processing (delayed, immediately). The decision of whether to purchase a standard package is also made in this stage. If a database is used, the logical database design is now made. (ISAC, because it is a process-oriented approach, provides little help in designing an outline for a database.)

Equipment Adaptation

The last area distinguished in ISAC is equipment adaptation, which concerns the eventual tuning between hardware facilities and the system designed. The physical medium (tape, disk) for storage of datasets is now determined. Then the actual construction ensues, which consists of programming, elaborating procedures, testing, and implementing the system.

Development of TSS: A Practical Case

At the start of the development it was decided that two variants of the system were required. One variant was to be a general provision for CITO (the CITO-TSS); the other was to be a system to be used in schools (the School-TSS). The CITO-TSS is a facility for professional test constructors; the School-TSS is meant for teachers without special training in psychometrics or evaluation methodology.

Change Analysis

The first area of ISAC, the change analysis, was performed separately, first for the CITO-TSS and then for the School-TSS. As mentioned above, change analysis consists of the determination of problems and needs and the study of change alternatives. As an example of the results concerning the determination of problems and needs, the problem table (Table 2) and the table of objectives (Table 3) of the School-TSS are given.

Table 2
Problems in the School-TSS

-
-
1. Time-consuming processing of responses and writing of reports.
 2. Limited interpretation of test results.
 3. Time-consuming test construction.
 4. Insufficient quality control.
 5. No characterization of items in terms of learning content.
-

The study of the change alternatives implies a description of the existing situation and the proposed future situation. In the case of TSS it has not proved useful to record the existing situation, because the desired situation entails an entirely new approach which cannot be deduced from the existing situation and the changes required. There-

Table 3
Goals of the School-TSS

-
-
1. The development of facilities for storing items on the computer, including both items constructed by CITO and user-constructed items.
 2. The development of facilities for adapting items provided by item publishers to specific user demands.
 3. The development of facilities for computer-assisted test construction from item banks. Items are selected on the basis of learning content characteristics.
 4. The development of facilities to make tailored/adaptive testing possible.
 5. The development of procedures for scoring, analyzing, and reporting test results.
 6. The development of facilities for guaranteeing the quality of items and tests.
 7. The development of procedures for interpreting student test results for the purpose of making decisions.
-

fore, only the A-graphs for the future situation were drawn and described. In the next two sections the main results of this analysis are given for the CITO-TSS and the School-TSS.

The CITO-TSS. The overview graph of CITO-TSS is shown in Figure 1. The graph shows, inside the frame, the activities and sets of messages involved in the CITO-TSS. The sets of messages are represented by parallelograms; the activities are represented by circles. The lines are flows of messages. The parallelograms above the frame represent the inputs that enter the CITO-TSS, and those below it represent the outputs or products of the system. All activities and sets have been given a number, which will be used in all tables.

In addition to the graphs, text pages and property tables were made. The text pages give a more detailed description of the activities and sets in the A-graphs. Table 4 is an example of a text page. It gives part of the description of an item in the item bank. The property tables describe the quantitative

aspects of the set of messages such as the frequency of activities and the size of sets.

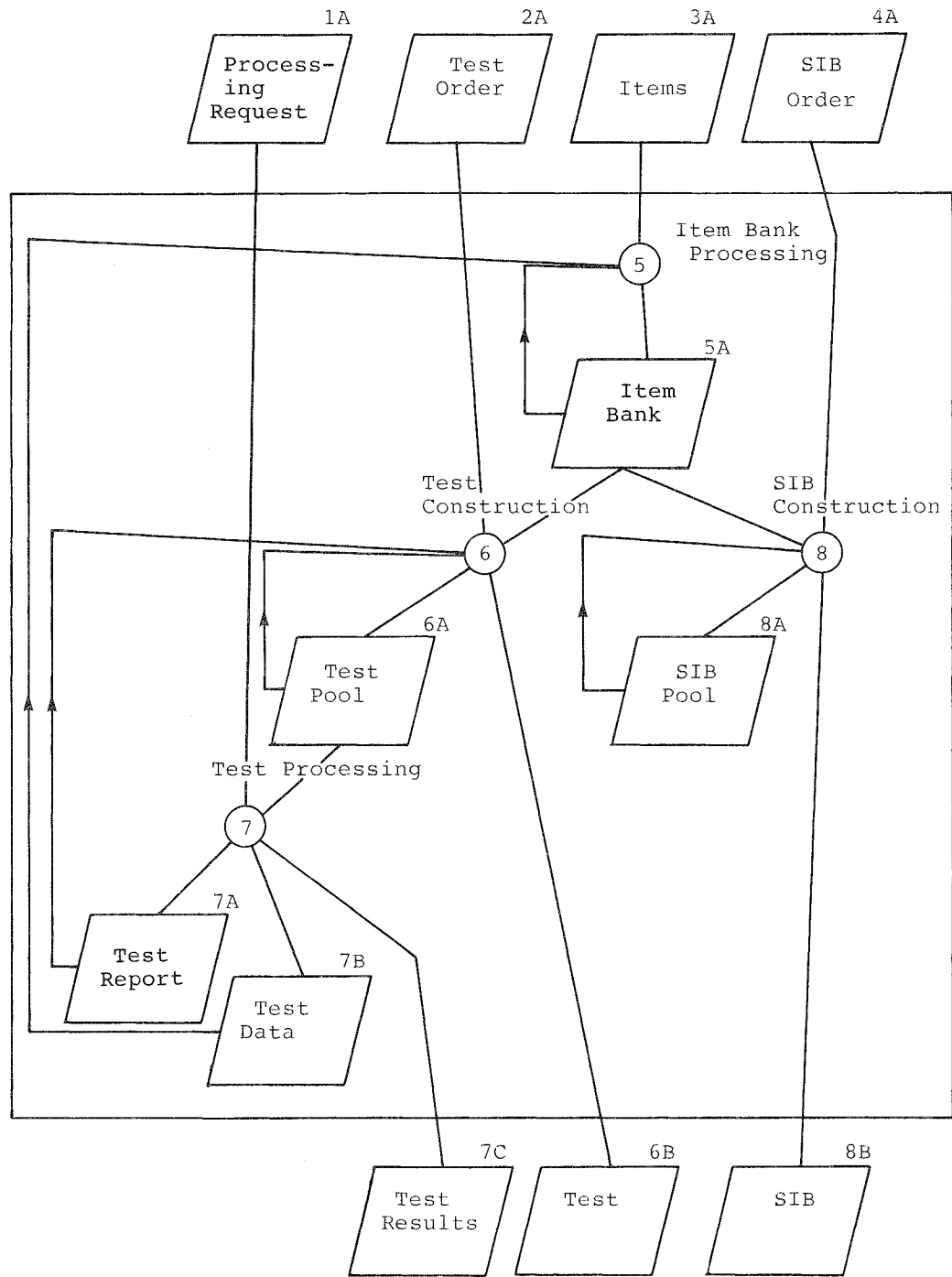
Item bank processing has items, item banks, and test data as input, and item banks as output. An item bank is a structured collection of items, structured both from a learning content and a psychometric point of view. The structure of the learning content forms the framework by which the items can be classified and retrieved. The categories of the structure should represent the content matter in such a way that it is possible to construct tests that possess both content and construct validity. For a number of school subjects the information-processing approach in cognitive psychology is useful for constructing such a learning content structure (e.g., the schema theory of Rumelhart and Ortony, 1977).

The psychometric structure of an item bank consists of the set of unidimensional scales that can be discerned in the bank. In a sense it is a partition of the set of items in homogeneous subsets. The meaning of the scales depends on the psychometric

Table 4
 Part of a Text Page of CITO-TSS

5A	Item bank	
5A1	Objectives	Text characteristics
5A2	Auxiliary material for items (if required)	Text/description characteristics
5A3	Items	
5A31	Item text	Question Answering alternatives Key/scoring instruction
5A32	Item characteristics	
5A321	General characteristics	Item type Additional material Admissible use Precluded items
5A322	Content characteristics	Content area Grade level Learning objective Auxiliary material
5A323	Psychometric characteristics	Number of responses Number of correct responses Rasch scale Item difficulty
5A324	User statistics	Date of last use in a test
5A4	Populations	School type Grade level Estimated parameters

Figure 1
Overview Graph of CITO-TSS



model employed. In the case of item banking it is necessary to use an item response theory test model in order to be able to test the validity of the psychometric structure and to be able to use parameter estimates independent of the sample used to calibrate the bank.

Item bank processing itself includes a range of activities such as the input of items, the classification of items, the determination of deficiencies in the item bank, the assessment of the quality of items, and the improvement of items. Most important, however, is finding the most suitable learning content structure and psychometric structure within a collection of items.

Next to item bank operations, test construction is the main activity in the CITO-TSS. The test order is input and contains the criteria which should be met by a test or a school item bank. Test order is determined by professional test constructors. This is both a specification in terms of learning content and a psychometric specification. The learning content specification leads to a collection of items which are admissible in the test. From this collection the items are selected which optimize the characteristics of the test. These characteristics are determined by the psychometric specifications. These specifications may include, among other things, reliability, information functions, Bayesian risk, and scale values. The result of the test construction is stored in the test pool. Apart from a description of the test (items, order of items), this pool also contains the estimates of the psychometric characteristics, and the instructions for administering the test and for processing the results. On request, a test can be drawn from the test pool and can be output in a form in which it can be administered.

The test processing consists of the processing of the item responses. The processing request contains these item responses and the specification of the type of output desired: for instance, the way in which item characteristics are presented, and the psychometric scale used for reporting the scores. The output consists of three different parts: (1) the scores for the students and schools involved (test results); (2) a report for the constructors of the test (test report); and (3) feedback to the item bank (test

data). By means of this feedback, the necessary psychometric characteristics can be attached to the items.

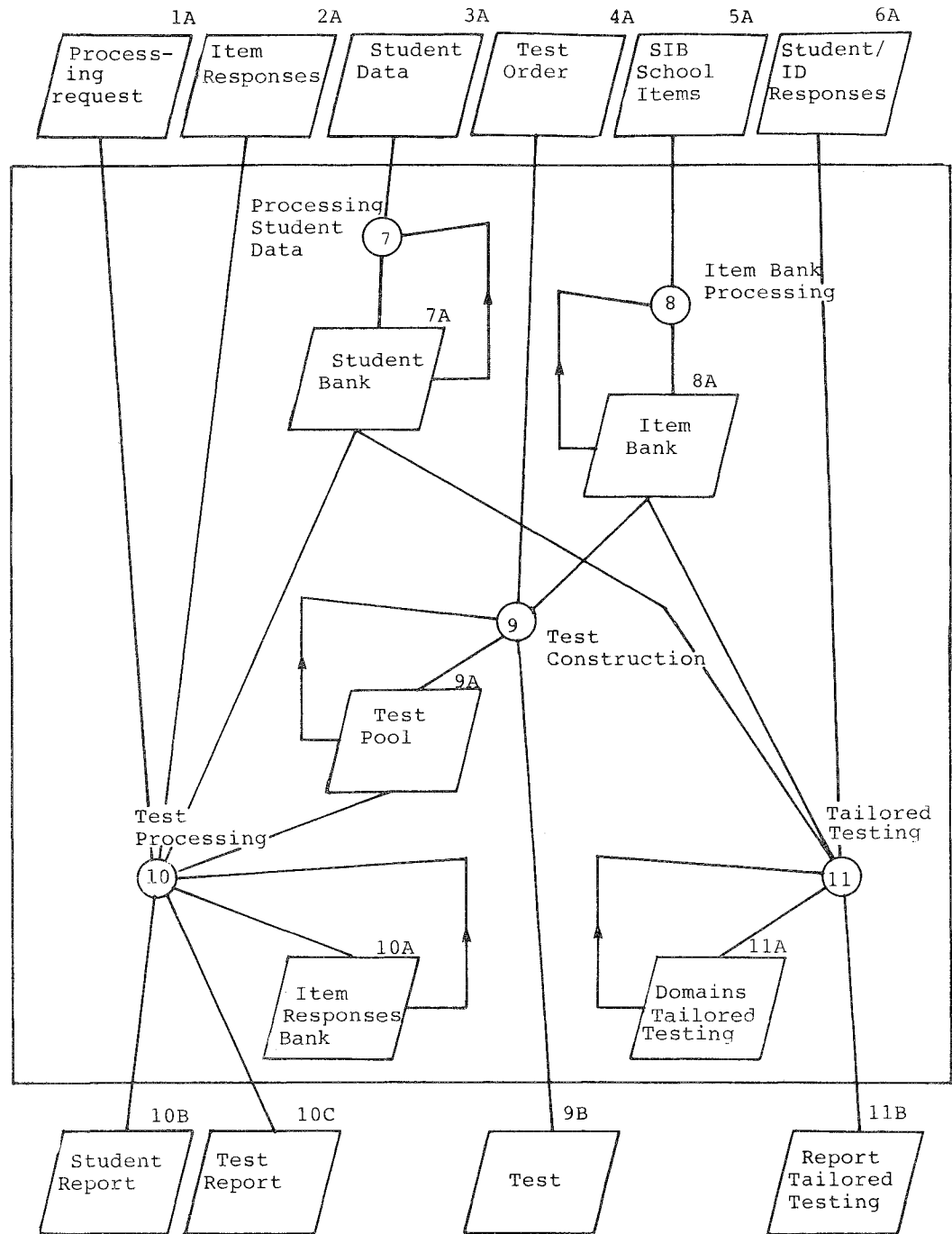
A School Item Bank (SIB) is an item bank that has been constructed by means of the CITO-TSS and is input for the School-TSS. This activity will be discussed when describing the School-TSS. Construction of a SIB is done in a manner similar to the construction of tests.

In addition to the A-graphs and their description, the study of the change alternatives also includes the property tables, which contain the quantitative aspects of the sets of messages and the activities. Item banks are the most important sets of messages of the system. In order to determine the expected number of banks and the size of the different banks, it is assumed that an item bank should contain all items that a user may need during the construction of one particular test. In this case, the resulting numbers are a maximum of 70 banks for the CITO-TSS, with 2,500 to 5,000 items each. The number of tests and the number of school item banks that will be constructed from these item banks depends to a large degree on the extent to which schools will be using the School-TSS. A conservative estimate is that 30 tests and 2 school item banks per year will be constructed from each item bank. These estimates are based on the current test production of CITO.

The School-TSS. Figure 2 shows the overview graph of the School-TSS. The CITO-TSS and the School-TSS are similar in appearance; the latter has been extended to include adaptive/tailored testing and processing of student data.

The main input for a School-TSS is school item banks. In addition to professionally constructed items, these item banks also contain information concerning content structure and psychometric structure. These structures are basically derived from the CITO item banks, but users of a School-TSS are able to enter their own items into the structure, and to alter existing items. These newly entered and/or altered items, however, cannot be included in the psychometric structure. As a consequence, the item banks of the School-TSS contain two kinds of items, referred to as CITO items and school items

Figure 2
 Overview Graph of School-TSS



respectively. Because the psychometric characteristics of school items are not known, they can only be used if no psychometrics are involved. A CITO item that is changed becomes a school item.

The opportunities for test construction in a School-TSS will be more limited than in the CITO-TSS. An important consideration in this regard is that the required knowledge of learning content and especially of psychometrics is not available in schools, and therefore misuse cannot be ruled out completely. Possible ways of assembling tests will be based on intuition, reliability, information functions, or scale values. A test can be stored in the test pool. For the purpose of administering the test, a master copy can be produced, along with directions for administering and scoring the test.

The ways in which item responses are processed will differ substantially between a School-TSS and a CITO-TSS. The emphasis in the School-TSS will be on providing results in a form that is useful for teachers in terms of making decisions. This means that different scalings will be possible, and that the results will always be accompanied by an interpretational framework. In processing the item responses, the psychometric information contained in the item bank can be used. Thus, the results can be compared with those of a norm population. Finally, it can be investigated whether the response pattern of a student deviates from that of other students in the same school or in the norm population (see Harnisch & Tatsuoka, 1983).

In the case of tailored/adaptive testing, the items, which are presented one by one to a student, are chosen in such a way that the maximum information about the student's proficiency is obtained. Such a test is, in general, much more efficient than a test that is the same for all students (Lord, 1980).

Before the adaptive test can be administered, it is necessary to determine the items from which it can be drawn and which define the meaning of the test. These so-called domains can be stored in the system. A domain will consist only of CITO items, because it is only for these items that psychometric characteristics are available.

The number of items in a SIB is the main figure from the property tables of the School-TSS. A small survey has shown that up to 700 items are necessary for a bank that covers one grade for one school subject, and that it should be possible to construct about 12 large tests a year. Teachers should be able to add a maximum of 200 school items.

The Activity Study of TSS

In the second area of ISAC, activity studies, determination of the information systems and subsystems takes place and the ambition level of each subsystem is determined. As an example, Figure 3 shows the subsystems within the TSS system.

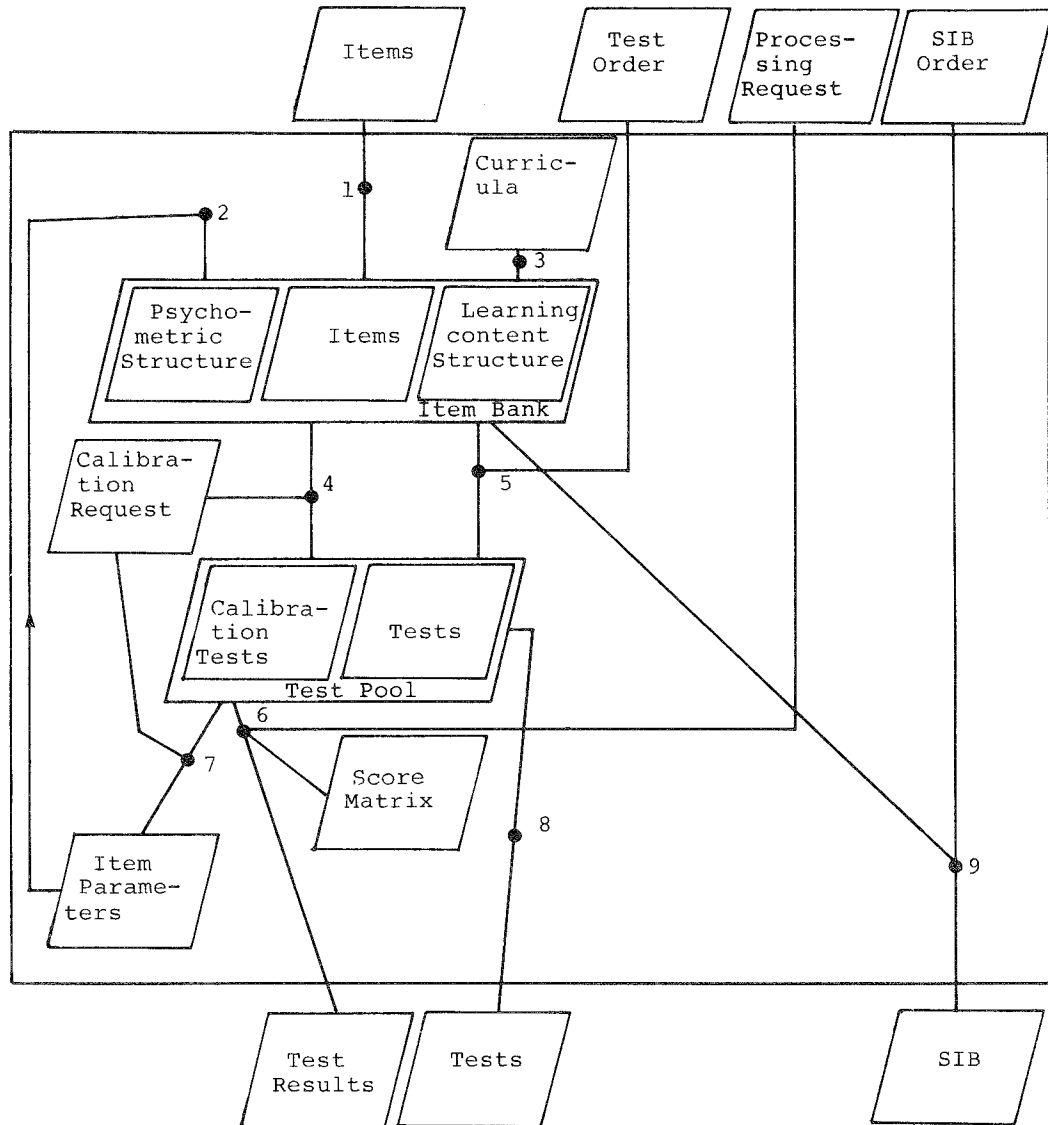
The main ambition levels for the CITO-TSS are related to the main aspects or factors that are to be implemented in the system. These factors are the learning content structure and the psychometric structure. The ambition levels themselves are given in Table 5. For the development of the CITO-TSS the highest level has been chosen. The reasons for this decision have primarily to do with the required validity of the test constructed from the item bank.

As mentioned before, an item banking system requires an item response model. No other model makes it possible to predict the characteristics of tests on the basis of item characteristics. The Rasch model was chosen because there is a clear interpretation of the model parameters and the model

Table 5
 Ambition Level Table for the CITO-TSS

Level of TSS	Content structure	Psychometric structure
Simple	Hierarchical topic structure	None
Elaborated	Hierarchical topic & concept structures	Classical test theory (p values)
Advanced	Hierarchical topic & concept structures	Mixed Rasch model

Figure 3
 Overview of Information Systems



- 1 Item construction and entering
- 2 Determination and entering psychometric structure
- 3 Entering and determination learning content structure
- 4 Construction of calibration design
- 5 Test construction
- 6 Test processing
- 7 Calibration processing
- 8 Test production
- 9 SIB construction

can easily be extended to incorporate polytomous scoring (Masters & Wright, 1984). A mixed model has been chosen because in an item banking context the main interest lies in estimating the distribution of student parameters, rather than in the parameters of the individual student (Goldstein, 1980).

Information Analysis

The development of the TSS is now at the stage of information analysis. The main components of information analysis are precedence and component analysis. Since the TSS system is a system that reflects a future procedure, the information need (i.e., the output description) is not yet known in every detail. Consequently, it is not certain whether the precedence and component analysis can be done in the way ISAC prescribes it. Perhaps for this step another methodology, which does not take the output description as a start, will be more suitable than ISAC.

Influence of ISAC on Development and Results

Problems During the Introduction of ISAC

Some problems occurred during the initial use of ISAC because it was the first time that a systematic approach was followed in the initial phases of a design. With every new step some experimentation was required before an adequate approach was found, for every product an adequate format had to be found, and it had to be determined whether it was useful to make the product at all, given the situation. The preliminary investigation for School-TSS, which was carried out about half a year later, was considerably shorter (18 months for CITO-TSS; 6 months for School-TSS), in part because the School-TSS is a less comprehensive system, but also due to greater experience with the approach. How much of the elapsed time for CITO-TSS can be attributed to inexperience with ISAC is difficult to determine in retrospect. Yet the experience with School-TSS suggests that it lies in the range of 25% of the total effort.

Advantages of Using ISAC

What were the advantages of following a systematic approach in the development of the system, and in particular, what influence did ISAC have on the final version of the system? Because a method of development is simultaneously a "way of looking" and a means of communication, it is obvious that the use of ISAC affected the eventual shape of the system. On the other hand, it is very difficult to predict what the system would have been like had another approach been followed.

However, there are a number of elements and moments in the development that clearly have been influenced. During the stage of change analysis, the boundaries of the system must be determined. ISAC forces the user, in drawing up the overview graph, to explicitly determine which data collections and activities belong to the system, and which do not. Thus, the decision was made to free the TSS from the responsibility for the reproduction and distribution of tests. It was also determined in this early stage that subscriptions of schools to recurring tests such as national exams would be dealt with within the system. By making these decisions at such an early stage in the development, all attention is concentrated on the functions that are truly relevant.

During the stage of the activity studies, it should become clear which activities and sub-activities are involved in the system under development. As a result, it was necessary to have a more specific idea of the processes of test construction and item construction. This necessitated analyzing the learning content (and entering and classifying items already available), as well as considering the integration of activities such as item construction, item classification, and control over the completeness of item banks. ISAC demands that the user maintain a consistent idea of the relations between the activities involved; otherwise, the necessary diagrams cannot be drawn. Here also, the discussion of the diagrams involves a conceptualization of the activity under discussion, and it is decided whether or not certain options will be implemented. Questions such as these must also be considered: When is an item bank complete? What is the correct slot

for an item in an item bank, or is there more than one correct slot? If so, how many? What structures can be identified in the subject matter and how can they be identified? Which instruments can be helpful in this respect?

Making the property tables, although they only contain preliminary estimates, also provides early insight into the practicality of the activities concerned; a more realistic view of item bank calibration has resulted. By estimating and calculating how many schools or students are required for calibrating an operational item bank, it has become clear that not all item banks that will be developed can be calibrated and recalibrated once every 3 years. The forming of ideas about collecting and processing test and item information obtained in constructing and administering tests in schools (by means of the School-TSS) was also influenced by the employment of ISAC and initiated at an early stage.

In the School-TSS, the employment of ISAC has influenced, among other things, the size of the system. In making the property tables, the use of the system in a school environment was a major concern. It became clear that only a limited number of teachers can use the same hardware configuration because there is only a limited amount of time available for the activities involved in test construction and test processing. The use made of the machine(s) in tailored/adaptive testing was also calculated and compared to the prospective availability of financial resources in schools. This meant, among other things, that the system will be constructed less elaborately than was assumed originally. The discussion also led to the conclusion that the facility for tailored/adaptive testing will be implemented as the last system component.

Finally, the employment of ISAC in the development of both the CITO-TSS and the School-TSS has enabled both systems to be optimally attuned to each other. As a result, the place for restructuring the item information for use in the School-TSS was not left to chance, but was consciously incorporated in the School-TSS itself.

The most important influence of a systematic approach such as ISAC is that the system to be

developed is clear from the beginning. In addition, all aspects of the system (such as collecting the necessary input data, distribution of output, maintenance, implementation, organizational adaptations and external cooperation required of schools and students, size of the operations, and research required) receive their rightful attention, and nothing is overlooked. Because information about these aspects is obtained at an early stage and can be incorporated in the implementation of the project, surprises are reduced to a minimum and the attunement of the different parts of the system is improved substantially.

Future Developments

The preceding sections have made it clear that a methodical approach to system development offers many advantages. Such an approach will likely be followed in the future; but it is another matter whether the method followed will be ISAC. Though it has been adjusted on certain points to suit CITO's needs (such adjustments are always necessary when following a standard approach), ISAC has so far proved very satisfactory. However, problems are anticipated with the current area, information analysis, as well as with future areas, especially data system design. By using advanced tools, such as fourth-generation languages and relational database systems, it is possible to produce the information required when the need arises. For the TSS system a relational database system has been chosen. A current investigation is considering whether ISAC is suitable for endorsing these new system development tools.

References

- Bemelmans, T. M. A. (1984). *Bestuurlijke informatiesystemen en automatisering* [Administrative information systems and data processing]. Leiden, The Netherlands: Stenfert Kroese.
- Goldstein, H. (1980). Dimensionality, bias, interdependence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology*, 33, 234-246.
- Harnisch, D. L., & Tatsuoka, K. K. (1983). A comparison of appropriateness indices based on item re-

- response theory. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 104–122). Vancouver BC: Educational Research Institute of British Columbia.
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Lundeberg, M., Goldkuhl, G., & Nilsson, A. (1981). *Information systems development: A systematic approach*. Englewood Cliffs NJ: Prentice-Hall.
- Masters, G. N., & Wright, B. D. (1984). The essential process in a family of measurement models. *Psychometrika*, 49, 529–544.
- Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson,

R. J. Spiro, & W. E. Montagu (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale NJ: Erlbaum.

Acknowledgments

The authors thank B. Weltens and N. van Zuijlen for translating and refining the text.

Author's Address

Send requests for reprints or further information to Michel Zwarts, crro, P.O. Box 1034, 8601 MG Arnhem, The Netherlands.