

COMPUTERIZED TESTING IN THE FEDERAL ARMED FORCES

WOLFGANG WILDGRUBE
GERMAN MINISTRY OF DEFENSE

The Federal Armed Forces (FAF) consists of about 480,000 soldiers (240,000 of these are draftees); the FAF administration comprises 170,000 civilians; and in the FAF Psychological Service there is a civilian staff of 1,300 psychologists. Figure 1 presents an overview of the organization of the FAF Psychological Services. The center of activities is in personnel psychology, with more than 80% of the psychologists in the area of aptitude diagnosis. Figure 2 shows the psychological aptitude testing procedures for selection and classification for both the FAF and the FAF administration. Aptitude diagnoses are carried out for various purposes for large samples, such as for draftees (about 300,000 diagnoses per year); for volunteers (about 30,000 per year); for advancement from sergeant to an officer career; and for selection of pilots, pyrotechnists, civil servants, and personnel for linguistic services. Aptitude and intelligence tests are administered by paper and pencil to groups of about 50 persons. Special apparatus tests or other special procedures and psychological interviews follow as necessary, dependent on the selection process or on the individual result. With these procedures, the Psychological Service thus attempts to make the best possible personnel decision.

Problems

The large number of testing procedures and the wide areas of testing create numerous problems. Mass testing (about 350,000 testees per year) requires a large quantity of material and manpower. The test application, scoring, and decision-making consist of many routine activities that require a great expenditure of personnel.

For each selection procedure all testees of a group process standard testing batteries: All testees undergo the same test battery during a limited period of time. For a certain number of testees the test is too difficult; for others, too easy. Thus, motivation decreases and fatigue increases. Special knowledge, attitudes, or personal spheres of interest or inclinations are not taken into consideration. Moreover, very rarely are special procedures possible, so that in the limited time allotted only some aptitude dimensions are carried out in an undifferentiated manner.

At present the mass data, collected by paper and pencil, do not permit follow-up analyses. Statistical evaluations of the testee data are impossible, and

Figure 1
Organization of the Psychological Services in the Federal Armed Forces of Germany

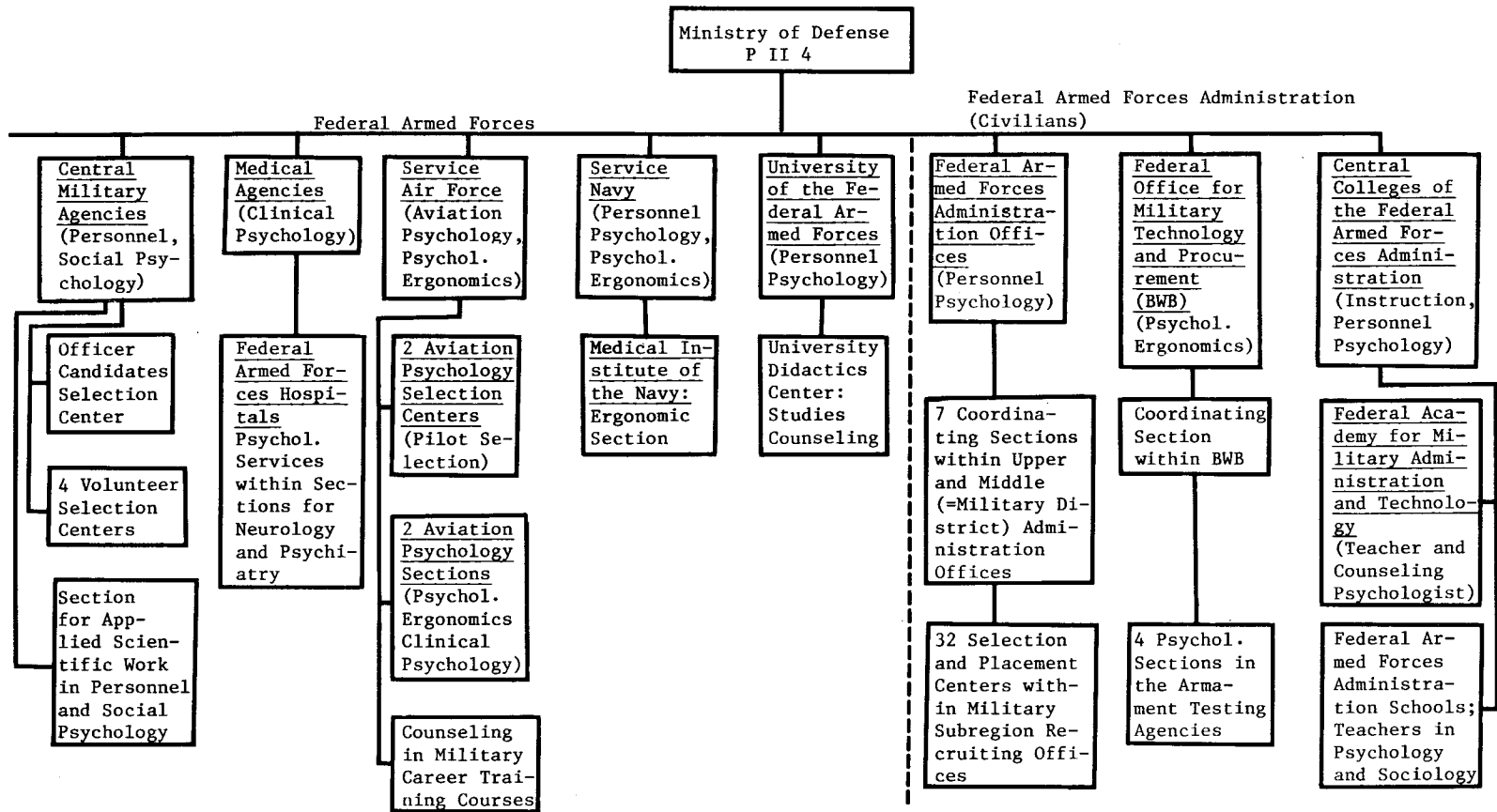
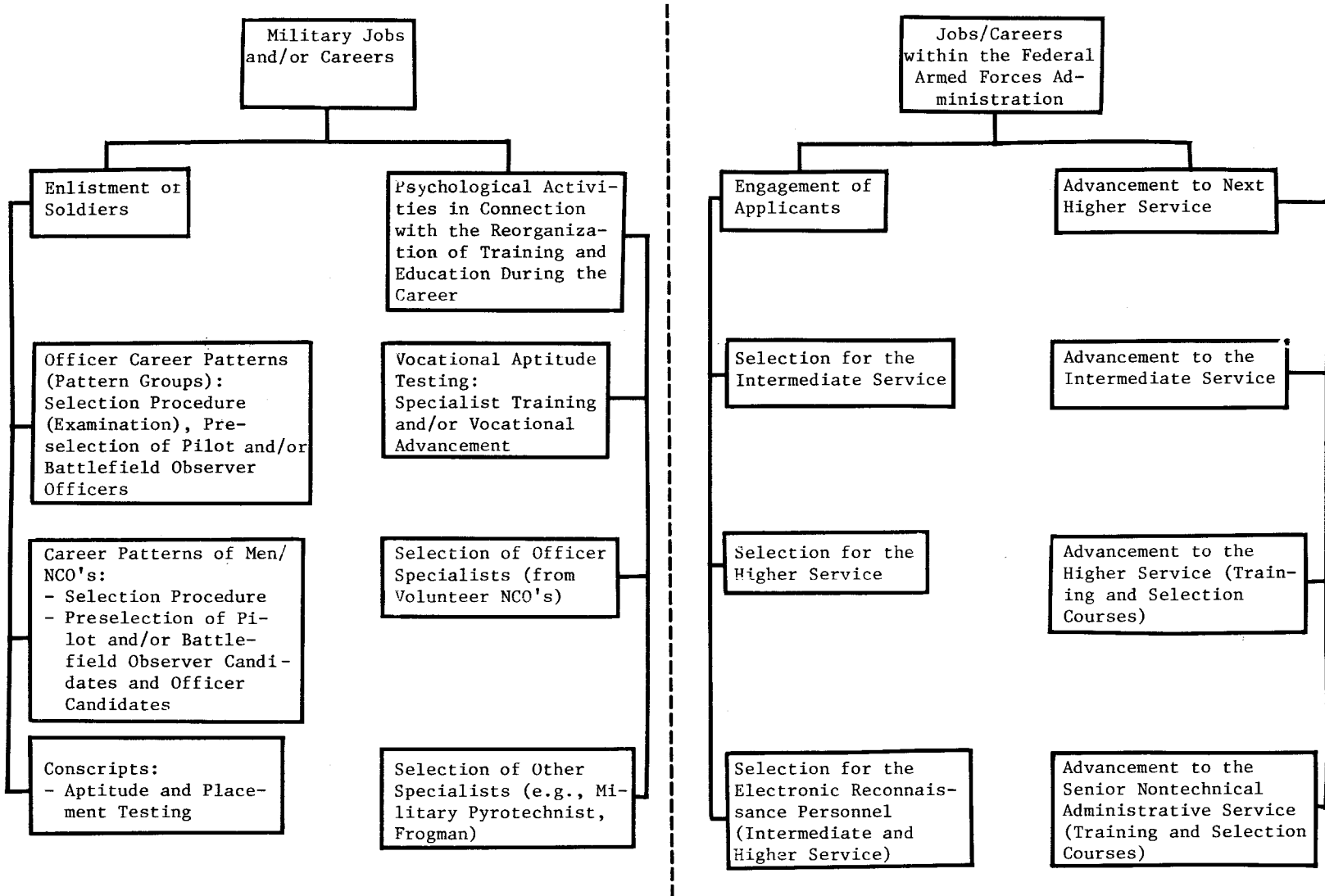


Figure 2
Psychological Aptitude Testing in the Federal Armed Forces of Germany



changes within the tests--for singular items or for the whole norm values--are not analyzed. Technical, organizational, and legal problems (as for instance, the security of tests) are connected with the mass testing and the different areas of aptitude diagnoses. It is necessary that the tests and the selection procedures be modified shortly. Above all, not only do the tests--that is, the psychological selection procedures within the scope of decisions or careers--become obsolete very soon but the patterns for their solutions (the items and the corresponding correct answers) become known after a very short time. It is not possible to perform a permanent modification in addition to the tests for career selection with the limited capacities available for such updates.

Requirements for the Diagnostic Process

Cognizance of these problems of aptitude diagnoses as well as the daily practice in the FAF provides a basis for the following requirements for future diagnostic work:

1. Improvement of the diagnoses is necessary; greater importance should be given to the differential diagnoses. A useful method should be found for solving the "bandwidth fidelity dilemma" so that, in spite of the use of mass testing, differential decisions are possible ("the right person in the right place"). This problem will become urgent for the FAF from about 1985 onward, when there will not be enough draftees available because of the rapid decline of the birthrate in the late 1960s.
2. Paper-and-pencil tests alone will not suffice in the future; with the improvement of diagnoses and the consideration of further aspects, skills, and experiences, it will be necessary to include new testing procedures and to test other psychological dimensions. Additionally, interests, motivations, and personality aspects should be tested, and perception and motor tests should be carried out to make more perfect diagnoses.
3. In addition to the test result--the score or ability parameter--other data should be included in the diagnostic process. Therefore, re-search programs concerning the testing process are necessary, including item solution time or time needed for solving a subtest, so that testing protocols (e.g., for counseling) can be produced.
4. Finally, mass testing makes it necessary to develop economy in the entire testing process and aptitude diagnosis. Scores and other computations should be carried out during the session, and results should be directly available at the end of testing. With these procedures and the proposed applications of items and subtests, it will be possible to save time and, moreover, to improve the diagnostic process.

Potential Solutions

Computerized testing will provide solutions to these problems in the following three areas:

Item production. Parts of item production can be performed by computer-assisted test construction (CATC). In a separate project, software was produced and implemented for item production and for individualization of tests, modifying the tests by computer. The first computer tests are in the empirical phase, and extensive results are expected in 1980.

Test data. For computation and interpretation of test data (selection and decision; "the diagnostic process") multi-faceted aids are possible: Simulations are being used in the FAF for computerized decision-making, and possibly the test results will be used to call up draftees.

Use of tests, the presentation of items, and scoring procedures. In addition to the presentation and computation of items using the classical concepts, there is a special case of test application--Computerized Adaptive Testing (CAT). Considerable savings and improvements of the aptitude diagnoses in the FAF are expected, especially from the adaptive methods and the new techniques of CAT.

Components of Computerized Testing

For the planning stage and implementation of computerized testing in the FAF a catalog was produced, containing the most important components of computerized testing and therewith also of CAT. These components, some of which will be empirically investigated by the FAF, include the following (the minimum requirements are preceded by an *):

Hardware

The requirement is defined to set up a test station, for example, for 50 testees carrying out diagnostic procedures of draftees. Many technical details (e.g., conception: connection to a large-size computer or stand-alone terminal station or a microprocessor for each testee; CPU and periphery, special screen and keyboards, and other facilities) are clarified and compared. Different products will be rated with regard to the requirements in the FAF.

1. *Requirement for flexibility of technology (e.g., extensions, innovations) and modular concept of hardware;
2. *Concept of the test station (connection to a large-size computer or stand-alone computer or microprocessor for each testee); system of minicomputers with foreground (input/output operations) and background (e.g., computations, estimations); multi-tasking, multi-processing;
3. *Central Processing Unit/Core Memory: construction, capacity/size, response/access/cycle time (e.g., station with 50 terminals testing draftees); byte or word, bit per word; accuracy/precision; floating point arithmetic (hardware or software, binary or decimal; number of bits for parameter estimations); *real-time execution, system-response time (processing an input immediately, without delay time);

4. *A printer for each testing station (production of testing protocols, plots); console display; possibility of storage, capacity of disks (magnetic disks or floppy disks); other storage on periphery; access mode and time; *archives/output of the raw data, compatibility (making copies to magnetic tapes, computations on an IBM large-size computer); processing the data on line/off line (among others, for the personnel division, using the test data in the data bases); connection to other computers in the FAF; definition of interfaces;
5. Equipment for a testing station for each testee (number of places connected to one processor); *special displays for presentation and processing the items; special keyboards (only digits and few buttons); display quality (sharp definition, contrast); graphic with 200,000 points, color equipment; use of video pictures; periphery, connection of further equipment/devices (tachistoscope, light pencil for figural tests or labyrinth items); usage of other apparatus or testing additional psychological dimensions with hardware or/and software (e.g., determination tool); controlling the testing process by acoustic stimulus, input of the answers using the terminal keyboard; employment of an A/D converter, making digitals using the physiological data or further testee data from other equipment;
6. *Infrastructure (e.g., power, power consumption, air conditioning); *mobility, possibility for transportation when testing draftees at different locations.

Test Applications/Concepts

The type of aptitude diagnoses to be taken over by a computer needs to be specified, for example, which psychological dimensions should be tested, which contents and methods should be used during the pilot projects (among others, the item-response time for ability estimation), and which further tasks (e.g., next item presentation and scoring) are possible with the test station, for example, computerized decision-making or counseling aspects.

1. *Flexibility for using different tests or methods; flexibility for time limits, sequence of subtests, power/speed tests, types of items, item material; flexibility for different data, changing the input of the test station (e.g., insertion of personal data or item-solution times); recording further psychological dimensions (perception, motor skills, concentration, coordination, fatigue, curves of learning, tracking); recording of interests, motivations, personality aspects; *possibility for different testing processes, omnibus procedures versus criterion-referenced measurement;
2. *Application of tests using the classical concept, presentation of conventional items by display (such as the present procedure for draftees); *jumping to different items, similar to the paper-and-pencil application (selection of different items by the testee, jumping forward and backward, as in a test book); *usage of sequential strategies based on subtests (screen and main test; indication of "critical

items" for the next subtest or for use in the interview); *processing the tests in groups of testees but continued application of individual tests/subsets of items;

3. Testing the pyramidal approach with the self-scoring aspect (in sensu Hornke, 1978); *application of tests with variable branching strategies, using different methods, different algorithms for parameter estimations, different scoring procedures, different criteria for cut-off; solving different methodological problems using different estimation procedures (Bayesian, maximum likelihood, and so forth); inquiry of CPU/execution time using different methodological approaches (different probabilistic models, various software);
4. Input of additional criterion data (e.g., age, date of final graduation from school), interests, special knowledge; recording the biographic data (using a questionnaire or free responses); *immediate computation of test data during the test process so that results are finished at the end of the session (i.e., scoring, norm values); interpretation of the test data, computerized diagnoses (classification with discriminant or cluster analyses); decision-making, placement recommendation for the draftees, taking into consideration the different requirements, priorities, or various criterion data of the armed forces; computerized personnel management (in contact with the data bases for the military personnel in the FAF); additional use of the test station for counseling aspects (e.g., possibilities of career, study at the universities of the FAF);
5. Possibility for giving feedback, processing several subtests; noting time limit if tests with time limit are in use (rest time per subtest, time used per item); *recording the item-solution time and processing the time as an additional ability estimator or for counseling; producing testing protocols with the response patterns (method for solving the subtest);
6. Possibility of computerized test construction; computation of follow-up analyses, validity approaches, and so forth.

Software

The system and the assembler programs monitoring the microcomputer, the possibilities for updating, the compatibility to an IBM large-size computer for follow-up analyses, and the real-time execution for presentation and computation of items should all be considered.

1. *Requirement for a modular system of software, implementation of new methods and testing procedures within a short time;
2. *Conversational/dialog program for processing the test sessions (selection of items, presentation, and computation; processing the item-solution time; possibly giving feedback); supervision of the test station (e.g., input/output, computations, interruptions, error han-

dling); monitoring the test process, operating log (e.g., internal statistics for usage of the subtests, items, error for handling, CPU time); *introduction for handling the CRT and the keyboard, processing of examples, operating the keyboard by various types of items; *check of the input for formal correctness (e.g., only one digit permissible or only a digit less than 5);

3. Requirement for programming the minicomputer by the user (e.g., the psychologist); using the higher software languages, such as FORTRAN or BASIC (interpreter or compiler); installation of a compiler for all stations or only for the development institution; usage of overlay techniques or virtual storage concepts, optimizing the core capacity; expense for programming, implementation of new tests, new methods, new software; support by utilities; improving the software and the assembler programs; updating the system of the minicomputer (e.g., presentation of the items, data management to archive the raw data, initial calculations); *storage of the data for follow-up analyses, transferring to a file of a large-size computer (calculations by SPSS or other software), development of software using a large-size computer via teleprocessing, simulating the minicomputer (e.g., conversational processing, compiler, assembler);

Organization and Usage

Checkpoints are the organization of the testing session during the entire selection process (with sport examination, medical check-up by physician, interview by psychologist, and so forth), operating the test station and the single screen/keyboard, handling for system trouble, and maintenance services.

1. *Requirement for simplicity of operations (nonspecialized operation of the testing station); *explanation for handling the CRT and keyboard for the testees (e.g., input, corrections, skipping forward and backward, giving assistance by a function HELP); monitoring the test process using the classical concept, i.e., for side-by-side terminals, parallel versions are presented;
2. Handling the test station for system trouble; restarting/restoring the system, rerunning the session, continuing with similar items (controlling the last transfer operation, the last processed item; successful processing of the last written operation; security of data (safe dump of the the raw data));
3. Breakdown time; maintenance services, agreement; spare parts; requirement for high readiness of operations;
4. Cost for purchase or lease, for maintenance and spare parts, and for operation (price-performance ratio);
5. Specific points of the firms (special features not described by the requirements above).

Planning and Procedure

Following the information and concept phase, in which information is collected and redefined for application of adaptive tests using a computer and for incorporation into the FAF, the first research programs are planned for examination and trial of the different contents, methods, and techniques; and the corresponding pilot projects are prepared. After checking computerized test applications in the FAF--their methods and techniques--the following parts and steps are designed:

1. Application of tests using the classical concept; presentation of conventional items by display; research program for the "psychology of computerized testing"; an experiment by Birke (1979) on the use of item-solution time as an additional ability estimator;
2. Testing the pyramidal approach with self-scoring (in sensu Hornke, 1977, 1978); and
3. Application of tests with variable branching strategies using different methods and approaches.

For these pilot projects item pools have been prepared and larger tests/subtests are presently in preparation. The extensive software should be produced in FORTRAN using the existing TSO connection to an IBM computer 370/168 and simulated corresponding test applications. Parallel to the planning of content and methods is the procuring of hardware, considering the components as previously designated in the catalog.

Since last year the FAF has had intensive contacts with the German firms Zak and Hogrefe, which--after many years of experience with the production of psychological-physiological tools--have offered microprocessor-based stand-alone computers for test application and for analyses of physiological data. Both firms are in the development phase, thus all offers have still not been realized (e.g., graphic equipment for 200,000 points, light pencil, use of video-tapes). Zak offers a modular system with 10 intelligent terminals, two floppy disk drives, and the central processor for one station; whereas Hogrefe offers a screen, a CPU, and a floppy disk for each testee.

The developments in the market are being observed and checked. Based on the requirements of the FAF, directions and concomitant requests are being formulated, and the German Ministry of Defense is providing a test station for the first pilot project for computerized test application.

Conclusion

The Psychological Services of the FAF is today at a starting-point of a new, rapid development of the testing process and aptitude diagnoses. At this time there is neither background experience nor a special approach to computerized testing in the FAF. Until now, problems of discussion and research designs have been oriented toward the practice in the FAF, derived from the everyday aptitude diagnoses requirements. I am certain, however, that in the coming

years the traditional concept of testing by using paper and pencil will be eliminated.

REFERENCES

- Birke, W. Non-parametric measurement of intelligence on the basis of item-response-time: Outline of a model. Unpublished manuscript, 1979.
- Birke, W., & Wildgrube, W. Die anwendbarkeit computerunterstützter adaptiver testmethoden im bereich der Bundeswehr (Wehrpsychologische Mitteilungen 1/78). Bonn: Bundesministerium der Verteidigung Psychologischer Dienst der Bundeswehr, 1978.
- Buchtala, W. Entwicklung eines testinstrumentes für individualisierte testung durch adaptive parameterschätzung. In W.H. Tack, Bericht über den 30. Kongress der Deutschen Gesellschaft für Psychologie in Regensburg (Band 2). Göttingen: Hogrefe, 1977.
- Fritscher, W., & Koch, E.W. Elektronische datenverarbeitung und personalklassifikation im Psychologie dienst der Bundeswehr (Wehrpsychologische Untersuchungen 3/75). Bonn: Bundesministerium der Verteidigung Psychologischer Dienst der Bundeswehr, 1975.
- Hornke, L.F. Antwortabhängige testverfahren: Ein neuartiger ansatz psychologischen testens. Diagnostica, 1977, 23, 1 - 14.
- Hornke, L.F. Vergleich zweier adaptiv-antwortabhängiger testsrategien. Diagnostica, 1978, 24, 103 - 112.
- Steege, F.W. Personnel psychology in the Federal Armed Forces of Germany. In J.W. Miller (Ed.), The 12th International Symposium on Applied Military Psychology (C-26-27). London: Office of Naval Research, London Branch Office, 1976.
- Weiss, D.J. (Ed.), Proceedings of the 1977 Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.
- Wildgrube, W. Computergestützte hilfen bei der eignungsdiagnostik--problemaufriss (Wehrpsychologische Mitteilungen 1/78). Bonn: Bundesministerium der verteidigung, Psychologischer Dienst der Bundeswehr, 1978.
- Wildgrube, W. Computergestütztes adaptives testen (CAT)--Neueste entwicklungen (Arbeitsberichte Nr P-3-78). Bonn: Bundesministerium der Verteidigung, Psychologischen Dienst der Bundeswehr, 1978.

years the traditional concept of testing by using paper and pencil will be eliminated.

REFERENCES

- Birke, W. Non-parametric measurement of intelligence on the basis of item-response-time: Outline of a model. Unpublished manuscript, 1979.
- Birke, W., & Wildgrube, W. Die anwendbarkeit computerunterstützter adaptiver testmethoden im bereich der Bundeswehr (Wehrpsychologische Mitteilungen 1/78). Bonn: Bundesministerium der Verteidigung Psychologischer Dienst der Bundeswehr, 1978.
- Buchtala, W. Entwicklung eines testinstrumentes für individualisierte testung durch adaptive parameterschätzung. In W.H. Tack, Bericht über den 30. Kongress der Deutschen Gesellschaft für Psychologie in Regensburg (Band 2). Göttingen: Hogrefe, 1977.
- Fritscher, W., & Koch, E.W. Elektronische datenverarbeitung und personalklassifikation im Psychologie dienst der Bundeswehr (Wehrpsychologische Untersuchungen 3/75). Bonn: Bundesministerium der Verteidigung Psychologischer Dienst der Bundeswehr, 1975.
- Hornke, L.F. Antwortabhängige testverfahren: Ein neuartiger ansatz psychologischen testens. Diagnostica, 1977, 23, 1 - 14.
- Hornke, L.F. Vergleich zweier adaptiv-antwortabhängiger testsstrategien. Diagnostica, 1978, 24, 103 - 112.
- Steege, F.W. Personnel psychology in the Federal Armed Forces of Germany. In J.W. Miller (Ed.), The 12th International Symposium on Applied Military Psychology (C-26-27). London: Office of Naval Research, London Branch Office, 1976.
- Weiss, D.J. (Ed.), Proceedings of the 1977 Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.
- Wildgrube, W. Computergestützte hilfen bei der eignungsdiagnostik--problemaufriss (Wehrpsychologische Mitteilungen 1/78). Bonn: Bundesministerium der verteidigung, Psychologischer Dienst der Bundeswehr, 1978.
- Wildgrube, W. Computergestütztes adaptives testen (CAT)--Neueste entwicklungen (Arbeitsberichte Nr P-3-78). Bonn: Bundesministerium der Verteidigung, Psychologischen Dienst der Bundeswehr, 1978.