Automated Portable Test (APT) System: Overview and prospects

A. C. BITTNER

Naval Biodynamics Laboratory, New Orleans, Louisiana

M. G. SMITH, R. S. KENNEDY, and C. F. STALEY

Essex Corporation, Orlando, Florida

and

M. M. HARBESON

Naval Biodynamics Laboratory, New Orleans, Louisiana

The Automated Portable Test (APT) System is a notebook-sized, computer-based, human-performance and subjective-status assessment system. It is now being used in a wide range of environmental studies (e.g., simulator aftereffects, flight tests, drug effects, and hypoxia). Three questionnaires and 15 performance tests have been implemented, and the adaptation of 30 more tests is underway or is planned. The APT System is easily transportable, is inexpensive, and has the breadth of expansion options required for field and laboratory applications. The APT System is a powerful and expandable tool for human assessment in remote and unusual environments.

The Automated Portable Test (APT) System has been developed as a tool for the assessment of human performance and subjective status. At present, it is being used in investigations of the effects of flight-simulator exposure on pilots and hypoxia effects on soldiers and in a variety of university studies. In the simulator study, for example, pre- and posttests are being administered in an effort to assess the magnitude and duration of simulator aftereffects on subjective and performance responses. Development of the APT System is based upon the concepts and empirical findings of the Performance Evaluation Tests for Environmental Research (PETER) Program (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984; Bittner, Carter, Kennedy, Krause, & Harbeson, 1984; Harbeson, Bittner, Kennedy, Carter, & Krause, 1983; Kennedy & Bittner, 1977). System development has been spurred by the general promise of microcomputers for human assessment, and the recent development of low-cost notebook-sized systems (Kennedy, Bittner, Harbeson, & Jones, 1982).

Portions of this report were prepared by personnel of the Naval Biodynamics Laboratory in support of Naval Training Equipment Center (NTEC) Task 3775-2P4 (Simulator Aftereffects) and by personnel of the Essex Corporation under NTEC Contract N61339-81-C-0105 and NASA Contract NAS 9-16982. Opinions or conclusions contained in this report are those of the authors and do not necessarily reflect the views or endorsement of the supporting government agencies. Trade names of materials or products of commercial or nongovernment organizations are cited where essential for precision in describing research procedures or evaluation of results. Their use does not constitute official endorsements or approval of the use of such commercial hardware or software. Requests for reprints should be sent to Robert S. Kennedy, Essex Corporation, 1040 Woodcock Road., Suite 227, Orlando, FL 32803.

The portability of notebook microcomputer systems has particular promise for environmental research, which frequently must be accomplished under difficult operational conditions. For example, the simulator-aftereffects study requires assessment at remote training sites. Time is extremely limited for operational reasons, and these locations often lack space and accessibility. To be practical, the computers used to perform this testing have to be inexpensive, portable, rugged, and user friendly, contain independent power sources, and provide adequate storage for data. The APT System has been developed to provide a human assessment capability suitable for use in remote operational environments.

The purpose of this report is to provide a descriptive overview and a brief prospectus of the APT System.

APT SYSTEM OVERVIEW

The APT System is comprised of three subsystems: (1) hardware; (2) test programs; and (3) system control.

Hardware

The hardware subsystem has been developed around a notebook-sized 8-bit personal computer: the NEC PC 8201A®. Integral to the microcomputer is a 32K internal read-only memory (ROM) containing, in addition to TELCOM and TEXT EDITOR, a version of Microsoft® BASIC. Table 1 abstracts the technical features of the microcomputer, which are more fully described in NEC Home Electronics (USA) (1983). Within a small and lightweight package, the system has: substantial onboard random access memory (RAM) ca-

Table 1
APT Computer Technical Specifications

Features	Specifications
Size	30 × 22 × 6 cm (11 × 8.25 × 2.5 in.)
Weight	1.7 kg (3.8 lb)
CPU	80C85 (CMOS version of 8085) with 2.4-MHz clock
ROM	32K (standard); 128K (optional)
RAM	16K (standard); 96K (optional)
Keyboard	67 standard keys: 5 function (5 more using shift key), 4 cursor directional, and 58 additional
Display	19 × 5.0 cm (7.5 × 2.0 in.) with reverse video option. May be configured as either a 240 × 64 element matrix or a 40 character × 8 line display.
Interfaces	1 Parallel (Centronics compatible) and 3 serial (RS-232C and 6- and 8-pin berg jacks)
Power Supply Options	4 AA nonrechargeable batteries, or rechargeable nickel-cadmium pack, or ac adapter 50/60 Hz @ 120 V ac, or external battery systems

capacity expandable to 96K; an external battery option (8 A h) providing for more than 100 h of continuous operation; and a built-in display. The NEC PC 8201A® provides the basis for an easily transportable and flexible human assessment system.

Augmenting the notebook microcomputer are the wide variety of auxiliary components shown in Figure 1. Among these, the (32K) RAM cartridges have proved particularly useful in applications to date. For field applications, the APT System and Testing Programs are maintained in internal RAM, and, after data collection, data are transferred to a RAM cartridge for mailing or carrying from remote sites to a centralized data-base location. For laboratory applications, it is anticipated that researchers may find it useful to extend the capabilities of the microcomputer with an external display (CRT), floppy disks, and computer interfaces. Overall, the NEC PC 8201A® has the expansion options required for a wide range of field and laboratory applications.

Figure 1. Equipment options for the APT System.

Test Programs

APT Test Programs are developed following an iterative three-stage process: identification; mechanization; and evaluation. These stages are described in the following sections.

Identification

The identification stage has involved the collaborative selection of tasks by test-development personnel. Initially, selection was from a set of 30 performance measures found to be most suitable statistically for repeated-measures applications in an analysis of 140 measures from the PETER Program (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984; Bittner, Carter, Kennedy, Krause, & Harbeson, 1984). Two examples of implemented tests drawn from this metaanalysis are: Pattern Comparison (Klein & Armitage, 1979), in which a subject compares two clusters of dots and determines whether or not they are identical; and the Grammatical Reasoning Test (Baddeley, 1968), in which the validity of a logical statement must be established. Recently, other tasks have been selected on the basis of identified requirements to supplement previously studied performance measures (cf. Carter, Kennedy, & Bittner, 1980). The Dynamic Landolt-C Test, in which a subject's dynamic visual acuity is adaptively assessed, is an example of a more recent development. Also, subjective-status questionnaires (e.g., Mood Adjective Check List) and related instruments have been mechanized to provide a more comprehensive assessment tool. In these recent developments, care has been taken to avoid test characteristics (e.g., proportional scoring) that are unsuited to repeated-measures applications (Bittner, 1981). The goal during the identification stage is to avoid mechanizing unstable or otherwise unsuitable tests.

Mechanization

Tests are programmed during the mechanization stage. Following the development of test requirements, assessment programs are initially written in BASIC. BASIC is used to facilitate modification and documentation, as well as transfer to other computer systems. Care is taken during mechanization to take advantage of the breadth of programming tactics for facilitating program execution [e.g., NEC Home Electronics (USA), 1983, Sec. 7, p. 7]. In many cases, logic simplification algorithms are also applied to facilitate program simplification and, consequently, to expedite execution (Mendelson, 1970). This approach frequently results in program speeds that have been evaluated as "effectively optimal" (this matter is discussed in the next section). However, when identified as required during evaluation, assembler routines are developed. An assembler routine, for example, has been developed that achieves 4-msec resolution for timing tests with the built-in clock (an alternate routine that promises 0.1 msec is under study). The goal during task mechanization is to develop efficient test programs with high construct validity.

Evaluation

Assessment programs are examined for efficiency and construct validity during the evaluation stage. Evaluation is initially conducted with interim criteria, to promote early appraisal of the quality of the mechanization. Subsequently, evaluation is conducted with ultimate criteria, which are experimentally more difficult to obtain. Hence, the evaluation stage is conceptually two-tiered, with feedback to mechanization conducted largely on the interim tier. Ultimate and interim criteria are described below.

Ultimate estimates. Following the approach developed during the PETER Program, efficiency is ultimately defined by the "reliability-efficiency of a differentially-stabilized test" (cf. Bittner, Carter, Kennedy, Krause, & Harbeson, 1984; Kennedy & Bittner, 1977). Reliability-efficiency, in this context, is operationally defined by the cross-day test-retest correlation, normalized to a fixed test time (viz., 3 min). Typically, this normalization is accomplished by application of the Spearman-Brown equation. Construct validity is ultimately assessed, similarly to reliability, by the corrected-for-attenuation correlation between the original and computerized versions of a test. These ultimate efficiency and validity criteria have been fully applied only to computerized tests outside the APT System (i.e., Smith Krause, Bittner, Kennedy, & Harbeson, 1983). However, investigations of these criteria are currently underway, and preliminary results indicate outstanding characteristics for tests in the APT System.

Interim estimates. An interim estimate of construct validity has been developed to expedite computer test development. To determine an estimate, judgments of the computer-original cross correlations (r_{co}) are made by psychological test experts who are knowledgeable about both versions of the tests. This approach was suggested originally by the apparent success of judgments regarding the tasks reported by Krause (Krause & Woldstad, 1983; Smith et al., 1983). Recent reports of the success of expert estimations of test validity provide support for this approach (Schmidt, Hunter, Croll, & McKenzie, 1983). Estimates of computer-original validity have approached their theoretical, reliability set, limits for several tests on the APT System.

Quick estimations of reliability-efficiency are made only after construct validity has been judged adequate to justify the exercise. Although in principle this could also be accomplished by direct estimation, an experimentally based approach has been developed that is more productive. This approach provides both an estimate of the current relative efficiency and an estimate of the "asymptotic computer relative-efficiency."

Computer test relative-efficiency (r_c) may be estimated by a variant of the Spearman-Brown equation. Specifically,

$$r_c = (k r_o) / [1 + (k - 1) r_o], \quad (1)$$

where r_o is the relative-efficiency of the original test (derived from previous work; e.g., Bittner, Carter, Kennedy, Harbeson, & Krause, 1984), and k indexes the ratio of

computer to original item completions per fixed period of time (T). In practice for performance tests, this ratio (k) is experimentally determined by

$$k = \bar{n}_c / \bar{n}_o, \quad (2)$$

where $\bar{n}_c$ and $\bar{n}_o$ are the average numbers of items (corrected for guessing) completed by a group of well-trained subjects in the fixed period. This latter index (k) has proved particularly useful in determining whether a computer test is operationally ready ($k \geq 1$) to replace its original counterpart.

Asymptotic computer test relative-efficiency (r_c') may be computed from Equations 1 and 2, given an estimate of the items that may be completed ($\bar{n}_c'$) with an infinitely fast computer. This number may be estimated by the equation

$$\bar{n}_c' = \bar{n}_c \bar{t} / (\bar{t} - \bar{t}_c), \quad (3)$$

where $\bar{t} = T/\bar{n}_c$ is the average time per computer item with the present system and $\bar{t}_c$ is the average time of the computer to accept a response and generate a new item. Interestingly, many current APT tasks have evolved to the point that their current asymptotic ratios ($\bar{n}_c / \bar{n}_c'$) are approaching unity (e.g., 1.25 for Pattern Comparison). These tasks have been termed "effectively optimal" because little improvement in reliability is expected [aside from possible input and output (I/O) format changes]. The asymptotic metrics ($\bar{n} / \bar{n}_c'$ and r_c') provide guidance regarding the limits of a computer test that is in development.

System Control

Figure 2 provides an overview of the system control subsystem. The control functions are divided between the experimenter and the computer.

Experimenter Functions

The experimenter is required to perform three system functions—two preexperimental and one postexperimental: system input, testing control input, and data transfer to an external medium (off-lining). Prior to the experiment, the system control and testing programs are bootstrapped into the computer, if they are not already present. For field applications, this typically involves inserting a (System) RAM cartridge and selecting the transfer program (BACKUP). Also, testing control parameters are entered into the control file (ORDER) as an ordered list. The test sequence is established by inputting the names of the tests to be used, in their order of use. Associated with each test (e.g., TEST.BA) are three values specifying the length of practice (if any), the maximum time allowed for subject response, and the total duration of the test (e.g., TEST.BA, 0, 10, 90). Once ORDER has been specified, a subject run may be started by selecting the initiating program (BEGIN). Postexperimentally, data are off-lined from the computer to an external device by selection of the transfer program (XFRDAT). The two preex-

Figure 2. Overview of the APT System flow from setup through data collection.

perimental functions are required only once per study, and postexperimental data transfer is done only after blocks of subjects have been run.

The experimenter functions have been designed such that they may be accomplished by paraprofessionals. Typically, 1 h of training is sufficient for functional mastery of the APT System.

Computer Functions

When a run begins, the computer prompts the subject to enter his or her unique (9-digit) coded identification number (ID). The system also requires the subject to enter salient data (e.g., hand dominance) the first time the subject is tested. After these preliminaries have been completed, the testing sequence is initiated and is continued

until the (ORDER) testing specifications have been met. Assessment scores, together with subject ID, are stored in a date- and time-coded data file (DATA) to be off-lined later. An APT System goal is to separate computer functions from experimenter functions to the greatest extent possible.

DISCUSSION

Overview

The APT System has been developed to provide a human assessment capability suitable for use in remote operational settings. As presented in the system overview, the hardware, test program, and system control subsystems meet the requirements for such a system. The notebook-sized NEC PC 8201A® provides the basis for an easily transportable and flexible assessment system with expansion options required for a wide range of field and laboratory applications. Additionally, the development of test programs is being conducted by a process to assure efficiency and construct validity. This process is based both on evaluation tools developed for computer tests and on lessons learned during the PETER Program (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984; Bittner, Carter, Kennedy, Krause, & Harbeson, 1984; Smith et al., 1983). Lastly, the experimental control subsystem has been simplified for use by paraprofessionals with minimal training. The APT System provides an assessment tool suitable for use in many remote settings.

Prospects

The APT System has substantial prospects for future growth and development. Attesting to this are recent and ongoing studies that have indicated that it has considerable promise for use in a broad range of unusual environments. For example, both an explosive decompression study and flight testing have found that the system is suitable for high-altitude, chamber, or airborne applications. These applications, coupled with the evaluation of the design for NASA, have indicated that the system could easily be adapted for orbiting shuttle use by applying spray coatings to the interior and to the external case. In addition, the NEC PC 8201A® has demonstrated robust capabilities to operate in at least the range of 0° to 32° C, to survive drop tests, and to withstand multiple airport x-ray exposures. The reliability of the system has been demonstrated during extensive field studies (> 10³ operational hours without failure). In addition to this robustness is the range of extension options shown in Figure 1. Overall, the APT System has substantial potential for application in unusual environments.

Conclusion

The APT System is a powerful and expandable tool for human assessment in remote and unusual environments.

REFERENCES

- BADDELEY, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, **10**, 341-342.
- BITTNER, A. C., JR. (1981). Use of proportion-of-baseline measures in stress research. In G. Salvendy & M. J. Smith (Eds.), *Machine pacing and occupational stress* (pp. 177-183). London: Taylor & Francis.
- BITTNER, A. C., JR., CARTER, R. C., KENNEDY, R. S., HARBESON, M. M., & KRAUSE, M. (1984). Performance evaluation tests for environmental research (PETER): Good, bad, and ugly. In *Proceedings of the 28th Annual Meeting of the Human Factors Society* (pp. 11-15). Santa Monica, CA: Human Factors Society.
- BITTNER, A. C., JR., CARTER, R. C., KENNEDY, R. S., KRAUSE, M., & HARBESON, M. M. (1984). *Performance evaluation tests for environmental research (PETER): Evaluation of 112 measures* (Research Rep. No. NBDL-84R006). New Orleans, LA: Naval Biodynamics Laboratory.
- CARTER, R. C., KENNEDY, R. S. & BITTNER, A. C., JR. (1980). Selection of performance evaluation tests for environmental research. In *Proceedings of the 24th Annual Meeting of the Human Factors Society* (pp. 320-324). Santa Monica, CA: Human Factors Society. (NTIS No. AD111296)
- HARBESON, M. M., BITTNER, A. C., JR., KENNEDY, R. S., CARTER, R. C., & KRAUSE, M. (1983). Performance Evaluation Tests for Environmental Research (PETER): Bibliography. *Perceptual and Motor Skills*, **57**, 283-293.
- KENNEDY, R. S., & BITTNER, A. C., JR. (1977). The development of a Navy performance evaluation test for environmental research (PETER). In L. T. Pope & D. Meister (Eds.), *Productivity enhancement: Personnel performance assessment in Navy systems*. San Diego, CA: Navy Personnel Research & Development Center. (NTIS No. AD A111180)
- KENNEDY, R. S., BITTNER, A. C., JR., HARBESON, M. M., & JONES, M. B. (1982). Television-computer games: A "new look" in performance testing. *Aviation, Space, and Environmental Medicine*, **53**, 49-53.
- KLEIN, R., & ARMITAGE, R. (1979). Rhythms in human performance: 1½-hour oscillations in cognitive style. *Science*, **204**, 1326-1328.
- KRAUSE, M., & WOLSTAD, J. C. (1983, June). *Massed practice: Does it change the statistical properties of performance tests?* (Research Rep. No. NBDL-83R005). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A139338)
- MENDELSON, E. (1970). *Boolean algebra and switching circuits*. New York: McGraw-Hill.
- NEC HOME ELECTRONICS (USA) (1983). *NEC PC-8201A® User's guide*. Tokyo: Nippon Electric Co.
- SCHMIDT, F. L., HUNTER, J. E., CROLL, P. R., & MCKENZIE, R. C. (1983). Estimation of employment test validations by expert judgment. *Journal of Applied Psychology*, **68**, 570-601.
- SMITH, M. G., KRAUSE, M., BITTNER, A. C., JR., KENNEDY, R. S., & HARBESON, M. M. (1983). Performance testing with microprocessors: Mechanization is not implementation. In *Proceedings of the 27th Annual Meeting of the Human Factors Society* (pp. 674-678). Santa Monica, CA: Human Factors Society.