In situ repeated measures of affect and cognitive performance facilitated by use of a hand-held computer

PETER TOTTERDELL and SIMON FOLKARD

MRC/ESRC Social and Applied Psychology Unit, University of Sheffield, Sheffield, England

The advent of hand-held computers means that it is now feasible to administer standard measures of affect and cognitive performance in field studies without the direct intervention of the investigator. The functional requirements for such an instrument are set out, and a set of programs developed for an electronic Organiser to meet those requirements are described. The programs include alarm scheduling, subjective rating scales, a sleep diary, and choice reaction time and memory search tasks. Compliance rates and usability of the instrument are assessed and judged satisfactory on the basis of examples from three field studies of shiftworkers.

Within the fields of social and applied psychology, there appears to be an emerging interest in research methods that involve the study of patterns of change within individuals over time, in which many variables are measured on many occasions for each individual (Larsen, 1987). This is partly motivated by the recognition that psychologists know comparatively little of the cognitive processes by which individuals change over time—and in particular, of how changes in mental states and performance are affected by the individual's experiences at home and at work. This method of research, which involves collecting time series data for each individual, has been termed a *process approach* (Larsen, 1989).

This method of research is not only desirable but also necessary if the cognitive process under study is periodic. In such a case, there is simply no satisfactory alternative to repeated measures of the indicator variable(s) within and across a number of cycles. The properties of the underlying rhythm usually include its periodicity, phase, and amplitude. Research methods for the investigation of periodic variations in psychological performance, particularly variations that are of a diurnal nature, have been viewed as belonging to the domain of chronopsychology (see, e.g., Folkard, 1977; Naitoh, 1982).

As we will illustrate in the next two sections, researchers who collect time series data from each individual appear at present to have divided themselves between those who are primarily interested in the endogenous and exogenous control of affect, and those who are primarily interested in the endogenous and exogenous control of cognitive performance. This separation of interests

may be exacerbated by the limitations of current instruments for the combined measurement of affect and performance in situ.

INSTRUMENTS FOR STUDYING AFFECT

Process studies of affect include investigations of daily stress and mood (e.g., Bolger, DeLongis, Kessler, & Schilling, 1989; DeLongis, Folkman, & Lazarus, 1988); mood entrainment (e.g., Larsen & Kasimatis, 1990); and experience sampling (e.g., Csikszentmihalyi & Larson, 1987). In such studies, self-report paper-based diaries have normally been used for the collection of data.

Typically, the participants are asked to carry their diaries as they go about their normal routines and to record their subjective responses in the diaries either at regular fixed intervals or, in the case of experience sampling, according to a random schedule. Technology has been used in experience sampling studies in the form of radiopagers, programmable watches, and programmable calculators, but usually only to signal the participant to complete the diary and not to record responses (Csikszentmihalyi & Larson, 1987).

Barr-Taylor, Fried, and Kenardy (1990), however, have used a programmable hand-held computer to record as well as signal the responses. Their use of the hand-held computer is very similar to our own, except that we have extended the use of the instrument to include objective measures of cognitive performance as well as diary and subjective affect measures.

INSTRUMENTS FOR STUDYING COGNITIVE PERFORMANCE

Most of the chronopsychological studies of cognitive performance have been investigations of circadian and time-of-day effects (Folkard, 1990; Monk, 1990). Alongside the relatively numerous laboratory-based investiga-

We wish to thank Lawrence Smith and Angus Craig, who were principal investigators in Studies 2 and 3, respectively, and Dave Minors and Jim Waterhouse, with whom we collaborated on the Isolation Unit Study. Correspondence should be addressed to P. Totterdell, MRC/ESRC Social and Applied Psychology Unit, P.O. Box 604, The University, Mushroom Lane, Sheffield, S10 1FP, England.

tions of circadian performance (Folkard & Monk, 1985), there have been a number of field studies in which on-job measures of performance have been used, and a few field studies in which standard laboratory measures of cognitive performance have been applied (Monk & Folkard, 1985).

In the latter studies, researchers have normally relied on paper-and-pencil tests either administered infrequently by the investigator (e.g., Wojtczak-Jaroszowa, Makowska, Rzepecki, Banaszkiewicz, & Romejko, 1978) or by the participants themselves, who use stopwatches to time their own responses (e.g., Folkard & Haines, 1977; Monk & Embrey, 1981). The reliability of data obtained in the latter fashion is questionable.

Performance tests run on personal computers (see, e.g., Rosa & Colligan, 1988) are likely to be more reliable, but they are difficult to administer in field studies of this type. The size and cost of personal computers mean that participants normally have to be tested on the same computer in series (i.e., not at the same time of day), often with an operator present. This is likely to present problems when testing is required at all times of the day and night, as in studies of shiftwork, and it may act to restrict the number of times of day at which testing can take place (see, e.g., Dalbokova & Kolev, 1991). It is also disruptive to the worker's routine and virtually rules out obtaining measures during the worker's leisure time.

The advantages of being able to use the same standard measures of cognitive performance in both laboratory and field settings in terms of validity and comparability are currently counteracted by the inherent disadvantages of the available measuring instruments. The hand-held computer has the potential of overcoming these disadvantages.

FUNCTIONAL REQUIREMENTS FOR COMBINED INSTRUMENT

An instrument that can be used in field studies for the collection of repeated measures of both affect and cognitive performance should fulfill the following requirements.

1. It must be portable. The participant has to be able to carry the instrument at all times.
2. It should be inexpensive. Ideally, each participant should have an instrument for the duration of the study.
3. It must signal the participant to respond without the intervention of the investigator.
4. It must record the time at which the participant responds. This is particularly important in time-of-day studies.
5. It must accurately record elapsed time. This is essential for cognitive performance measures such as reaction time (RT).
6. It should have sufficient display size to allow presentation of tasks.
7. It should have sufficient memory for program and data storage.

8. It should allow downloading of data to a personal computer for subsequent analysis. This eliminates the need for a separate data coding stage.

We have developed a suite of programs for a programmable hand-held computer that appear to satisfy the requirements above. The usability of the instrument will clearly also be a key factor, but we believe that given proper instruction in its use, a good user interface, a safe user environment, and reliable performance, such an instrument can in fact motivate participant compliance.

Over the past 2 years, our programs have been used successfully in a number of field and laboratory studies. In this account, we will concentrate on the instrument itself—hardware, software, and usability—rather than on the substantive results of those studies. However, we have included a few results that should help illustrate the utility of the instrument.

METHOD

Hardware

The hand-held computer that we chose was a PSION Organiser II, Model LZ, which is advertised as an electronic personal organizer and costs approximately £150 sterling (it is available from PSION, Alexander House, 85 Frampton St., London, England). Its weight is 250 g without batteries, and its dimensions with the protective case closed are 142 × 78 × 29 mm. The LZ model has 64K ROM and 32K RAM, but this can be expanded. The LZ model has a 4 line × 20 character LCD screen. The keyboard is alphabetic and not QWERTY. The Organiser has its own programming language, OPL, and programs can be developed either on the Organiser itself or by using the Organiser Developer software on a personal computer and then downloading to the Organiser. Data can be transferred from the Organiser to a PC-compatible or a Macintosh, using a Comms Link (approximate cost, £60).

Our preference has been to use a 64K datapak (approximate cost, £59) for the programs and a 64K datapak for the data files. The advantages are the separation of programs and data, the safety of programs and data in the event of power failure, and the ability to use the Mk III Datapak Copier. The Mk III Copier (approximate cost, £995) enables programs to be simultaneously copied from a master datapak to up to eight datapaks, and also enables simultaneous transfer of files from up to eight datapaks to a personal computer. A datapak formatter is required to erase the contents of a datapak; this separate device costs approximately £45.

The Organiser has a built-in clock and calendar, which can be used for setting alarms and time-stamping events. Elapsed time can be calculated by reading the Organiser's frame counter, which adds an increment every 50 msec. This sets the upper limit on accuracy for a single RT measure.

Alarms

Although the Organiser provides built-in alarm functions, these were not felt to be sufficiently flexible for our purposes, so we developed our own programs that enable the investigator to schedule alarms that signal the participant to respond at scheduled intervals. Because of our particular interest in time-of-day effects, the alarms programs set fixed intervals between alarms. The interval can, however, be set to change within or between days. With minimal additional programming, it should also be possible to randomize the interval, for example, for an experience sampling study.

Before commencing a study, the investigator runs a program that assists him/her in creating a master alarms file; this file can then be copied onto the data datapaks. Each record on the alarms file specifies the date and time of the first alarm, the interval in minutes between alarms, the number of alarms required, and the number of consecutive days for which this should apply. When the alarm sounds, it beeps, pauses for 10 sec, and then beeps every 2 sec for up to a minute unless the user presses a key. Unfortunately, there is only one level of loudness on the Organiser, and it is not very loud.

The users can turn the alarms off for the duration of their sleep by answering in the affirmative a question that asks, at the start of each set of tests, whether or not they are going to sleep in the next interval. On waking, the user completes a sleep diary on the Organiser, which has the effect of automatically resetting the alarms for the day. The users can also override the alarms by selecting a menu option, which allows them to turn off either the next or all alarms. If they choose to turn off all the alarms, they can reinstate the alarms later by selecting another menu option.

User Environment

The user programs have been designed to run within a program shell. This protects the user from the other functions of the Organiser and improves the security of the data. An escape route is provided in the shell to allow the investigator to stop the program. Only through this escape route or by removing and replacing the battery can someone access either the alarms schedule or the recorded data.

When the users turn the Organiser on, they are first prompted for an ID number. They then see a menu of options, which might include a sleep diary, a nap diary, an alarms option, and a test option. If they choose the test option, they will then be taken through a set of selected tasks. Exit points are provided between tasks in case the user does not have time to continue. However, once the user has selected an option or entered a response, there is no means of backtracking.

The programs have been developed as a set of modules, one for each task. Among the advantages of the modular design are that tasks can undergo a number of iterations in their development without affecting other tasks, and that new tasks can easily be added. By editing the program files for the menu options and the test option, the investigator can easily choose which tasks to present and in which order to present them. The investigator can also decide whether or not to include in the test file procedure names that have the effect of giving users feedback on their performance.

Data Files

Modularity is also maintained in the data files, with each task module creating and updating its own file. Whenever a task is executed, a new record is created in the file with new values in each of the record fields. The first field in every record is a time stamp of the current date, hour, minute, and second. The file name indicates the task and the user ID number.

The modularity of the data files makes it possible to select subsets of the data for analysis. However, it has the disadvantage that the time stamps have to be matched when the files are merged for analysis. But given that the Organiser has a 16 fields per record maximum, one large data file was not viable.

Task Programs

A range of task programs have been developed for the Organiser.

Rating scales. These are 20-point subjective rating scales. A number of scales for alertness, mood, and task load (NASA Task Load Index: Hart & Staveland, 1988) have been programmed. However, it is easy for investigators to construct their own scales simply by supplying the label for a scale item as an argument to one of two program procedures. The two procedures construct either monopolar or bipolar scales.

Figure 1. Presentation of a bipolar and monopolar rating scale.

To complete a scale, users move the cursor, shown in Figure 1 as an asterisk, up onto the horizontal line, then move the cursor to the desired position along the line, and finally press a specific key to indicate that they are satisfied with the cursor's position. The date and time of completion as well as the position of the cursor are recorded.

Sleep diary. The user would normally be required to complete this on waking from a main sleep in each 24-h period. The diary prompts users for answers to a number of questions about the sleep that they have just finished. The following data are recorded:

Went to bed at (hours:minutes)
 Started trying to sleep at ...
 Number of minutes to fall asleep
 Number of times woke up
 Number of minutes sleep lost
 Awoke at ...
 Got up at ...
 Quality of sleep (scale 1-20)
 Start work at ...

Serial choice reaction time. RT has been shown to be sensitive to many factors, including fatigue and sleep loss (Wilkinson, 1964) and time of day (Gilliooly, Smolensky, Albright, Bartholomew, & Thorne, 1990). RT can be assessed in a number of ways; for example, both simple unprepared and choice RT can be measured. The serial choice RT task that we chose to implement has been shown to be sensitive to a range of stresses (Broadbent, 1963).

In our version of this commonly used task, an asterisk symbol appears on the display in one of four positions. These correspond to the relative positioning of the M, T, W, and R keys on the Organiser's keyboard (Figure 2). The user is required to press the corresponding key as quickly as possible. The next stimulus is presented immediately.

Users are instructed to use the second and first fingers of the left hand for the M and T keys, respectively, and the first and second fingers of the right hand for the W and R keys, respectively. There is a fixation point in the middle of the display, which the user is asked to gaze at between stimuli. Trials on which the user fails to respond within a set amount of time, usually set at 1 sec, are excluded from the RT and accuracy measures but are not replaced. These trials are termed *gaps*, and the percentage of gaps is recorded.

There are two task lengths: the full length, with five blocks of 80 stimuli each, and the restricted length, with just two blocks

Figure 2. Presentation of 4 choice RT task.

of 80 stimuli each. Each set of eight trials within a block of 80 has 2 stimuli in each of the four positions; the order is random, apart from this constraint. Data are recorded for each block. This causes a small (approximately 1-sec) interrupt between blocks. The Organiser's 50-msec limit on the resolution of timings means that a single RT measure may not be reliable, and hence the RTs are summed over blocks. The following data are recorded:

- Date and time of completion
- Current block
- Number of stimuli in block (default 80)
- Threshold setting (default 1 sec)
- Mean RT for left hand in seconds
- Mean RT for left hand correct responses
- Mean RT for left hand incorrect responses
- Percent trials left hand incorrect
- Percent trials left hand above threshold
- Mean RT for right hand in seconds
- Mean RT for right hand correct responses
- Mean RT for right hand incorrect responses
- Percent trials right hand incorrect
- Percent trials right hand above threshold

There is another programmed version of the task, which incorporates both two and four choices but which has a fixed interval between stimuli. An advantage, however, of using the task without an interval is that the number of user responses per unit time is maximized.

Memory search task. This task is based on Sternberg's (1969) paradigm and is very similar to the STRES (AGARD, 1989) implementation of the task. The task is useful for differentiating between factors that affect response processes and those that affect central memory scanning processes.

The user is required to memorize a set of letters that appear simultaneously across the display. The number of letters in the memory set can be changed by the investigator to alter the memory load. The user presses a key to indicate that the set has been memorized. A fixed number of *probe* letters then appears, one of them at a time, in the middle of the display. The user presses the T key if the letter was in the memory set and the F key if it was not.

The memory set letters are randomly selected but without repetition, without vowels, and without the response key letters T and F. Negative probes have the additional constraint that they are not members of the memory set. Positive probes are equally likely to match any of the memory set letters. Each probe has an equal chance of being or not being in the memory set. If the user fails to respond to a probe within 5 sec, the results are ignored and the next probe is displayed, but the probe is not replaced. The following data are recorded:

- Date and time of completion
- Size of memory set (1, 3, 5, or 7)
- Number of memory sets (default 1)
- Number of probes per set (default 20 or 40)
- Time taken to memorize set in seconds
- Number of true positives
- Mean RT for true positives in seconds
- Number of false negatives
- Mean RT for false negatives in seconds
- Number of true negatives
- Mean RT for true negatives in seconds
- Number of false positives
- Mean RT for false positives in seconds
- Number of probes not responded to within 5 sec

Search and memory (SAM). This task, like its predecessor MAST, was originally devised as a paper-and-pencil task for time-of-day studies (see, e.g., Folkard, Knauth, Monk, & Rutenfranz,

1976). Circadian performance rhythms on this task have been shown to vary with memory load (Folkard, Wever, & Wildgruber, 1983).

The user is required to memorize a set of letters. The size of the set, with 1, 3, or 5 target letters, can be changed by the investigator to alter the memory load. When the user presses a key to indicate that the list is memorized, two lines of 20 letters appear on the display. Users scan the lines from left to right and press a specific key each time they see one of the memorized letters. When they have finished the search, they must press another key. Another target set then appears for them to memorize.

The user receives two blocks of 4 trials, with a new target and stimulus set on each trial. The target set is selected at random but without vowels and without repetition. The targets are positioned at random in the stimulus set, and there is an equal probability of 0, 1, 2, 3, or 4 targets' appearing. The following data are recorded at the end of each block:

- Date and time of completion
- Size of target set (1, 3, or 5)
- Number of current block
- Number of trials (default 4)
- Total number of targets in block
- Mean time to memorize target set in seconds
- Mean time per trial in seconds
- Percent trials on which number of responses matched number of targets in stimulus set
- Percent targets detected
- Percent targets overdetected

The main drawback with the task is that one cannot assess whether or not users have detected a valid target when they respond. The paper-and-pencil version of the task requires the participant to mark each target, but keyboard skills are too unreliable to require the same in the computer version. However, the accuracy measure per trial will only be compromised if the number of omission errors equals the number of commission errors. In a study (Folkard, Arendt, & Clark, 1990) that included the paper-and-pencil version of SAM, 10 subjects searched 12 blocks of 360 letters and missed 341 of the 1,080 targets but made only 4 errors of commission.

A number of other cognitive performance tasks have been programmed for the Organiser, but either they have not yet been used in a field study or their sensitivity is unknown. A version of Baddeley's (1968) logical reasoning task similar to the STRES version (AGARD, 1989), and a version of Michon's (1966) interval production task similar to the CTS version (Shingledecker, 1984) are available but have not been used to date. A visual spatialization task, *vector*, is also available. This task briefly displays two aircraft as static arrows and requires the user to judge whether the situation is hazardous or not. The task was originally devised for ship pilots (Craig & Condon, 1984) and therefore has two dimensions only; the heights of the aircraft are not included. The sensitivity of the task is not yet known.

RESULTS

Compliance

We will illustrate participant compliance with the Organiser with data from the first study in which we used the Organiser. Study 1 was a study of university site security workers who monitored the university's premises from a central control room. There were 12 male workers, grouped in four teams of 3, providing 24-h cover on a weekly backward rotating shift system with 8-h shifts. Of the 12 workers, 1 was on holiday during most of the study, 2 declined to take part for personal reasons (a second job,

and distrust of the purpose of the study), and 1 experienced an Organiser display malfunction.

The remaining 8 participants were asked to carry the Organiser for the 28 days of their shift cycle; all started on the same calendar date. They completed tests every 2 h on the even hour during waking, which included both work and leisure time. The tests, which took approximately 6 min to complete, included alertness, mood, and disruption scales, and a five-target SAM and a four-choice RT task. The participants were given the option of discontinuing the test at the end of the rating scales, which they were instructed to use if they were very busy. The participants were also asked to complete a sleep diary each day.

A problem in the software enabled 1 participant to break out of the user environment and accidentally reset the clock. However, the participant continued with the study and so his compliance with the tests is included in the following summary.

The compliance figures for the study are shown in Table 1. Assuming a sleep average of 8 h, a maximum of eight 2-hourly tests per day per participant for the 28 days could be expected. This gives a compliance rate of 50.2% with the 2-hourly tests, 43.5% with the SAM tests (representing a dropoff of 6.7% due to lack of time), and 76.4% with the daily sleep diary.

We can further assume that in each working day there will be a maximum of four on-shift readings and four off-shift readings per participant, and that there will be an average of 3.5 days' holiday per participant each fortnight. This means that in the first half of the study, the on-shift compliance was 68.6%, dropping off to 54.8% in the second half of the study. The off-shift compliance in the first half of the study was 56.3%, dropping to 30.4% in the second half.

We can compare some of these figures with those from a study, part of which is reported in Folkard et al. (1990), in which paper-and-pencil measures were used to assess the sleep, mood, and performance of 20 police officers over a 28-day shift cycle. Compliance with the daily sleep diary used in that study was 78.4%, which is similar to the figure from the Organiser study. Compliance with 2-hourly ratings of alertness was 66.5%, but the participants were only asked to record alertness on 12 days, so that the best comparison figures from the Organiser study are those from the first half of the study, which are similar. Compliance with the paper-and-pencil SAM 1 and SAM 5

Table 1
Test Compliance of 8 Site Security Workers
During 28-Day Study

	<i>M</i>	<i>SD</i>	Range
Number of 2-hourly responses	112.4	61.1	33-227
Number of SAM tests completed	97.5	61.8	28-212
Number of sleep diary entries	21.4	7.4	5-28
Days 1-14 on-shift responses	28.8	9.1	
Days 1-14 off-shift responses	39.4	21.6	
Days 15-28 on-shift responses	23.0	15.9	
Days 15-28 off-shift responses	21.3	21.3	

Table 2
Usability Evaluation of Organiser

	Disagree	Neither	Agree
Study 1: 8 Site Security Workers (6 completed questionnaire)			
Q1. Enjoyed using Organiser	0	3	3
Q2. Prefer to pen & paper	2	0	4
Q3. Easy to carry around	2	0	4
Q4. Organiser easy to use	0	1	5
Q5. Tests didn't interfere with work	4	1	1
Study 2: 23 Power Plant Workers (18 completed questionnaire)			
Q1. Enjoyed using Organiser	1	5	12
Q2. Prefer to pen & paper	1	2	15
Q3. Easy to carry around	3	2	13
Q4. Organiser easy to use	1	2	15
Q5. Tests didn't interfere with work	5	8	5
Study 3: 17 Air Traffic Controllers (16 completed questionnaire)			
Q1. Enjoyed using Organiser	1	5	10
Q2. Prefer to pen & paper	1	0	15
Q3. Easy to carry around	2	3	11
Q4. Organiser easy to use	0	2	14
Q5. Tests didn't interfere with work	2	2	12

tasks was 63.5%, but the participants were only asked to complete tests once a day on 12 days, in contrast with the 8 times a day for 28 days asked of participants in the Organiser study. The paper-and-pencil compliance figures may also be inflated, because there is no possible check on when the ratings and tasks were actually completed.

Of the total of 899 2-hourly tests completed in Study 1, 19% commenced before an alarm was due, 63% commenced within 10 min of an alarm, and 18% commenced over 10 min after an alarm. This suggests that participants did make use of the programmed alarms.

Usability

At the end of Study 1, and at the end of two other field studies of shiftworkers (Study 2 and Study 3), the participants were asked to complete a simple questionnaire to assess their feelings about the usability of the Organiser. The results are summarized in Table 2. With the "disagree," "neither," and "agree" responses to questions scored as 0, 1, and 2, respectively, the mean level of agreement for each question can be represented graphically as in Figure 3.

In Study 1, the compliance rates of the 3 subjects who reported that they enjoyed using the Organiser were higher than those of the 3 subjects who were indifferent, suggesting that subjective reactions to the Organiser may translate to behavior.

Study 2 was a study of 23 power plant workers. The workers were on one of two shift systems, depending on whether they were engineers or operators/craftworkers. The participants kept the Organiser for a shift cycle, which was 5 weeks on both systems, but they were asked to complete 2-hourly tests on only 8 specific days of the cycle. The tasks in the 2-hourly test were an alertness scale, task load index scales, a four-choice RT task, and SAM 5. The tasks took approximately 6 min to complete.

Figure 3. Usability evaluation of Organiser.

Study 3 was a study of 17 air traffic controllers who worked a rapid forward rotating shift system with 8-h shifts. They were on duty for an average of 90 min between breaks. The participants kept the Organiser for 20 days and completed tests at the start of a shift, during their breaks, and at the end of a shift. The tasks during the breaks were workload ratings, response to workload ratings, mood ratings, vector, and memory search (with memory set sizes of 1 and 5). The tasks took approximately 5 min to complete.

Reliability and Validity of Tasks

In many studies, it would be possible to compare the results of the first half of the study with those of the second half in order to assess the reliability of the measures. However, this was not possible for our set of studies because of the confounding factor of the shift cycle. The results were likely to depend on the position of the shift-worker within the shift cycle; and the two halves of each study were not symmetrical with respect to this factor. For example, Figure 4 shows the average daily on-shift and rest-day SAM 5 time-per-trial results of 1 participant from Study 1. It shows quite clearly that RT was slower when the participant was working night shifts.

The mean SAM 5 RT per trial of this participant was 15.52 sec, which is the equivalent of 388 msec per item searched and thus compares with the paper-and-pencil SAM 5 mean of 393 msec per item derived from the 20 subjects used in the Folkard, Arendt, and Clark (1990) study.

In a recent laboratory study (Folkard, Totterdell, Minors, & Waterhouse, 1991), Organisers were used to collect data from 3 participants every 2 h while they were awake over a 26-day period. They spent the last 15 days in a temporal isolation unit, where they were on a 30-h

sleep-wake schedule. Figure 5 shows how the practice effect for the four-choice serial RT results of one participant persisted throughout the course of the study. This suggests that practice effects in performance are unlikely to be eliminated and must therefore be taken into account during analysis.

The mean four-choice RT of this participant was 380 msec, excluding gaps (RTs > 1 sec). Woodworth and Schlosberg (1954) quote a figure of 364 msec for three-choice and 434 msec for four-choice, RT tasks, which probably includes gaps.

Sensitivity of Tasks

Figure 6 shows how the on-shift SAM 5 time per trial of the same participant from Study 1 varied during the

Figure 4. Mean on-shift and mean rest day SAM 5 time-per-trial measures over 28-day shift cycle of security worker.

Figure 5. Practice effect for four-choice serial reaction time ($n = 1$).

Figure 6. Mean SAM 5 time-per-trial results of security worker on different shifts over 28-day shift cycle.

Figure 7. Power spectrum of four-choice serial reaction time ($n = 1$).

morning, afternoon, and night shifts. A two-way analysis of variance (ANOVA) with shift and hours into shift as the two factors gave a significant main effect of shift [$F(2,75) = 26.34, p < .001$]. The night shift was associated with the slowest SAM 5 RTs. Casting the data into 12 2-hourly bins and using a repeated measures multivariate ANOVA gave a significant time-of-day effect [$F(11,72) = 2.8, p = .005$] and explained 30% of the variance. This could be due to an endogenous circadian rhythm, but it could also be due to exogenous effects such as time since waking and time on shift.

Figure 7 shows the power spectrum of four-choice serial RT for the same participant from the laboratory study. This is based on the time series of 88 2-hourly mea-

sures obtained during the last 10 days of the study. Cosine curves varying from 20 to 32 h in 0.2-h increments were fitted by a least squares method in order to assess the "power" of the rhythm at different periods. The figure shows quite clearly that there are 25-h, 30-h, and possibly 21-h performance rhythms on this task. Similar but less powerful rhythms have been obtained for SAM 5 RT and memory search (1 and 5) RT (Folkard et al., 1991).

Performance of Organiser

There have been very few hardware or software problems with the Organiser. Apart from the Organiser display malfunction in Study 1, there were also 2 nondiagnosable "trap" errors in Study 2, which were rectified by resetting the Organiser.

The number of sampling points that can be recorded on the Organiser depends both on the capacity of the datapak and on the number and combination of tasks used. But as a rough guideline, one may note that the most compliant participant in Study 1 completed 227 2-hourly tests and 28 sleep diary entries. Each 2-hourly test recorded a total of 32 fields of information and each sleep diary entry 10 fields, giving a grand total of 7,544 data points for this participant. This occupied 61K of the 64K datapak. A warning to contact the investigator would have been displayed had the datapak reached within 2K of its maximum. Datapaks of 128K are available for longer studies. A warning is also displayed if the battery is low, but in this case, one long-life battery was sufficient.

Another aspect of the performance of the Organiser that is clearly critical to our use of the instrument is the reliability of the timer and associated software procedures. To check the timer's reliability, the choice RT module was altered so that it did not empty the key buffer before starting the timer. By keeping a recognized key continually depressed to stock the key buffer, it was possible to check whether the module recorded an RT of 0 sec for each stimulus. Of 8,000 responses, 7,170 were recorded as 0 sec and 830 were recorded as 0.05 sec. If we therefore assume an average overestimate of 0.05 sec on 10.38% of the trials, this implies that the average overestimate of mean RT on a block of 40 serial choice stimuli will be +0.0052 sec.

DISCUSSION

The hand-held computer has proved itself to be an invaluable and practical instrument for repeated measures of cognitive performance and self-reported affect in our field studies. It allows the investigator to cue the participant to respond without the investigator being present; collect data at all times of day and night in all locations without direct intervention or excessive disruption; objectively record when a task is completed; obtain objective measures of response time; and automatically code data and hence speed analysis and eliminate human coding errors.

Participants in the field studies have responded favorably to using the Organiser. Only 2 out of 40 participants

reported that they had not enjoyed using the Organiser, and only 1 participant reported difficulty in using the instrument.

The compliance rate of 50.2% on the 2-hourly tests in Study 1 was not spectacular. However, given that the study lasted for 28 days and that the tasks took approximately 6 min to complete, the participants were being expected to give up over 22 h of their time to the study with no incentive, financial or otherwise. It is interesting that compliance was higher during work hours and that leisure time readings dropped off considerably in the second half of the study. In subsequent studies, we have been more judicious with regard to the length and number of tests, particularly in the number of leisure time measures that we expect. This has also been prudent, in view of the sheer amount of data generated by this type of study.

The major limitation of using the Organiser as we have described is that the investigator has no control or knowledge of the conditions under which a test is completed. Ambient conditions such as light and noise are likely to vary considerably. Interruptions during tests and social inhibitory factors may also vary. Hand and finger positioning and the eye-to-display distance are also likely to change, despite instructions to the contrary at the start of a study. The act of faith on the part of the investigator is that these variations will not change or swamp the phenomena under investigation.

Another problem, which is common to all measuring instruments, is that the Organiser may interfere with the participant's daily life and hence alter the constructs being measured. The extent of this problem is difficult to assess, however, because there is no acceptable yardstick that is not also subject to the same criticism. The best one can hope for is to look for discrepancies between results obtained with the use of different instruments.

Although we have suggested that the hand-held computer will prove most useful in field studies in which there is a need to measure multiple variables of affect and/or cognitive performance on many occasions for each participant, clearly there are many other situations in which it would be beneficial too. For example, Barr-Taylor et al. (1990) have related that they are already finding such instruments useful for therapeutic interventions. Similarly, we have found that our instrument can be cost effective in cross-sectional laboratory studies simply because it is possible to use many Organisers in parallel.

These applications, plus the potential for simultaneous measures of cognition and affect and for viable comparisons of laboratory and field results, make the hand-held computer an exciting practical addition to the psychologist's toolbox.

PROGRAM AVAILABILITY

Those who wish to obtain the programs should send a 64K datapak to the first author. The returned datapak will include two source files for changing the choice and order of tasks (which allow other programs to be linked in) and the executable programs for the environment and tasks.

REFERENCES

- ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT (AGARD, 1989). *Human performance assessment methods* (AGARDograph No. 308). Springfield: National Technical Information Service.
- BADDELEY, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, **10**, 341-342.
- BARR-TAYLOR, C., FRIED, L., & KENARDY, J. (1990). The use of a real time computer diary for data acquisition and processing. *Behaviour Research & Therapy*, **28**, 93-97.
- BOLGER, N., DELONGIS, A., KESSLER, R. C., & SCHILLING, E. A. (1989). Effects of daily stress on negative mood. *Journal of Personality & Social Psychology*, **57**, 808-818.
- BROADBENT, D. E. (1963). Differences and interactions between stresses. *Quarterly Journal of Experimental Psychology*, **15**, 205-211.
- CRAIG, A., & CONDON, R. (1984). Operational efficiency and time of day. *Human Factors*, **26**, 197-205.
- CSIKSZENTMIHALYI, M., & LARSON, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous & Mental Disease*, **175**, 526-536.
- DALBOKOVA, D., & KOLEV, P. (1991, July). *Changes in performance and heart rate variability under 12 hour shiftwork*. Paper presented at meeting on "Shiftwork and Job Demands" associated with the 11th International Ergonomics Association Congress, Paris.
- DELONGIS, A., FOLKMAN, S., & LAZARUS, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality & Social Psychology*, **54**, 486-495.
- FOLKARD, S. (1977). Chronopsychology: Implications for education. *Chronobiologia*, **4**, 111. (Abstract)
- FOLKARD, S. (1990). Circadian performance rhythms: Some practical and theoretical implications. *Philosophical Transactions of the Royal Society of London: Series B*, **327**, 543-553.
- FOLKARD, S., ARENDT, J., & CLARK, M. (1990). Sleep and mood on a "weekly" rotating (7-7-7) shift system: Some preliminary results. In G. Costa, G. Cesana, K. Kogi, & A. Wedderburn (Eds.), *Shiftwork: Health, sleep and performance* (Proceedings of the IX International Symposium on Night and Shift Work, Verona, Italy, 1989). Frankfurt: Lang.
- FOLKARD, S., & HAINES, S. M. (1977). Adjustment to night work in full and part time night nurses. *Journal of Physiology*, **267**, 23-24P.
- FOLKARD, S., KNAUTH, P., MONK, T. H., & RUTENFRANZ, J. (1976). The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system. *Ergonomics*, **19**, 479-488.
- FOLKARD, S., & MONK, T. H. (1985). Circadian performance rhythms. In S. Folkard & T. Monk (Eds.), *Hours of work, temporal factors in work-scheduling* (pp. 37-52). Chichester, U.K.: Wiley.
- FOLKARD, S., TOTTERDELL, P., MINORS, D., & WATERHOUSE, J. (1991, September). *Dissecting circadian performance rhythms: Implications for shiftwork*. Paper presented at 10th International Symposium on Night and Shift Work, Sheffield, England.
- FOLKARD, S., WEVER, R. A., & WILDGRUBER, C. M. (1983). Multi-oscillatory control of circadian rhythms in human performance. *Nature*, **305**, 223-226.
- GILLOOLY, P. B., SMOLENSKY, M. H., ALBRIGHT, D. L., BARTHOLOMEW, H., & THORNE, D. R. (1990). Circadian variation in human performance evaluated by the Walter Reed performance assessment battery. *Chronobiology International*, **7**, 143-153.
- HART, S. G., & STAVELAND, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshaki (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier.
- LARSEN, R. J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. *Journal of Personality & Social Psychology*, **52**, 1195-1204.
- LARSEN, R. J. (1989). A process approach to personality psychology: Utilizing time as a facet of data. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 177-193). New York: Springer-Verlag.
- LARSEN, R. J., & KASIMATIS, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality & Social Psychology*, **58**, 164-171.

- MICHON, J. A. (1966). Tapping regularity as a measure of perceptual motor load. *Ergonomics*, **9**, 401-412.
- MONK, T. H. (1990). Shiftworker performance. In A. J. Scott (Ed.), *Occupational medicine: Shiftwork* (pp. 183-198). Philadelphia: Hanley & Belfus.
- MONK, T. H., & EMBREY, D. E. (1981). A field study of circadian rhythms in actual and interpolated task performance. In A. Reinberg, N. Vieux, & P. Andlauer (Eds.), *Advances in the biosciences: Vol. 30. Night and shift work: Biological and social aspects* (pp. 473-480). London: Pergamon.
- MONK, T. H., & FOLKARD, S. (1985). Shiftwork and performance. In S. Folkard & T. Monk (Eds.), *Hours of work, temporal factors in work-scheduling* (pp. 227-238). Chichester, U.K.: Wiley.
- NAITOH, P. (1982). Chronobiologic approach for optimizing human performance. In F. M. Brown & R. C. Graeber (Eds.), *Rhythmic aspects of behavior* (pp. 41-104). Hillsdale, NJ: Erlbaum.
- ROSA, R. R., & COLLIGAN, M. J. (1988). Long workdays versus rest-days: Assessing fatigue and alertness with a portable performance battery. *Human Factors*, **30**, 305-317.
- SHINGLEDECKER, C. A. (1984). *A task battery for applied human performance assessment research* (AFAMRL-TR-84-071). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.
- STERNBERG, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, **57**, 421-457.
- WILKINSON, R. T. (1964). Effects of up to 60 hours sleep deprivation on different types of work. *Ergonomics*, **7**, 175-186.
- WOJTCZAK-JAROSZOWA, J., MAKOWSKA, Z., RZEPECKI, H., BANASZKIEWICZ, A., & ROMEJKO, A. (1978). Changes in psychomotor and mental task performance following physical work in standard conditions, and in a shift working situation. *Ergonomics*, **21**, 801-810.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology* (rev. ed.). London: Methuen.

(Manuscript received July 29, 1991;
revision accepted for publication January 13, 1992.)

5th European Conference for Research on Learning and Instruction
Aix-en-Provence, France
August 31 - September 5, 1993

The 5th European Conference for Research on Learning and Instruction, organized by the European Association for Research on Learning and Instruction (E.A.R.L.I.), will be held in Aix-en-Provence, France, August 31 to September 5, 1993.

The conference will include invited addresses, symposia, paper sessions, poster sessions (workshops), and demonstrations. The official language will be English.

The main themes will be: Cultural and social aspects • Learning and teaching in specific subject areas, including foreign languages • Text comprehension and text production • Learning disabilities • Classroom interactions and learning processes • New media and information technologies • Evaluation and assessment • Adult learning and continuing education • Learning in natural settings • Knowledge acquisition, problem solving, and metacognition • Developmental processes • Teaching process and teacher thinking • Industrial and professional training • Social, motivational, and cognitive factors • Parental involvement in learning.

The address for further information is: Jean-Paul Roux, 5th EARLIE Conference Secretariat, U.F.R. de Psychologie et Sciences de l'Éducation, Université de Provence, 29 Avenue Robert Schuman, 13621 Aix-en-Provence Cedex, France (Telephone, +33-42 27 03 08; Fax, +33-42 20 59 05).