The User Interface in Computer-based Selection and Assessment: Applied and Theoretical Problematics of an Evolving Technology

John F. Booth*

Although its central role in human-computer interaction is recognized by industry, the user interface (UI) has received scant attention in the computer-based testing (CBT) literature. Industrial and European Community ergonomic directives have been established regarding UI design issues, however, which may significantly impact upon CBT. Accordingly, the UI's CBT role is discussed regarding central CBT/UI constructs and dynamics. On the basis of one CBT UI design issue (here: paging) it is shown that many ergonomic regulatory definitional traps exist for the unwary CBT practitioner and UI designer. Newer UI developments in windowing technology and computer-based UI ergonomic regulation conformance analysis are described. Finally, open issues regarding CBT UI problematics are discussed, showing that, rather than continuing to lead a shadowed existence, the UI is deserving of considerably increased research effort regarding its CBT impact.

Key words: GUI, user psychology, user mental models, ISO standards, UI design.

Introduction

Although the user interface (UI) has long been recognized in industry as being central to effective human computer interaction (see Dzida 1988; Shneiderman 1987), the UI has led a somewhat shadowed existence in computer-based psychometric testing (CBT). For example, and to focus on the majority of CBT literature, where test authors commonly discuss factor structures of their assessment instruments, and adaptive testing specialists elaborate upon branching strategy and probabilistic modeling, UI aspects such as the rationale for a particular screen design and/or a selected user-input mode are rarely given. Seldom are UI design/usability aspects included in a discussion of experimental methodology and procedures. As such, it would seem that the quality and efficacy of implemented UI designs are either tacitly assumed or considered secondary to other research goals. Such an attitude is unjustified for at least two reasons:

First, an ever-growing body of research is pointing out that the UI plays a *central* role in human factors issues. If the CBT practitioner fails to take such findings into consideration (or glosses over UI problematics), he or she runs the risk of having ineffective/incorrect UI design negatively impacting upon testing validity.

Second, developments are occurring in the area of European national norming respective to user interface design which may, ultimately, have a serious and long-term impact upon CBT: industrial ISO standards and European Community directives are setting directions regarding UI design which can not only potentially help but also hinder the diagnostician in effecting accurate and valid CBT assessment. As will be shown, the proper interpretation and consideration of such directives respective to their use in UI design is essential if one is interested in meeting the 'spirit of the law' of such normative directives – while meeting one's goals of psychometric reliability and validity.

Discussional structure

Given this background, the following discussion is broken down into six major parts:

- First, and for the reader having little familiarity with UI issues, some UI definitions and key UI-related constructs will be given. Additionally, key stages of the diagnostic process will be discussed respective to their general relevance regarding UI design.
- Second, UI-related constructs will be discussed regarding UI dynamics – or, as it were, the UI 'in action': it will be seen that the

*Address for correspondence: John Booth, John Booth Consulting, Schleifweg 62, D-69126 Heidelberg, Germany, Tel: +49 6221 300 605, Fax: +49 6221 300 605, E-mail: john.booth@rhein-neckar.de

© Blackwell Publishers Ltd 1998, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

UI can involve a considerably complex myriad of perceptual and cognitive dynamics.

- Third, as they impact upon practically all European organizations utilizing computer technology, recent European and national norming directives respective to UI dialog design are outlined. Their general implications regarding CBT form a discussional platform for pragmatic and theoretical issues which are discussed in a later section.
- Fourth, and using the preceding sections as argumentative background, two GUI scenarios will be given which focus upon one major aspect of UI design in testing operations: paging. The scenarios will be discussed respective to certain ISO/EC directives regarding user dialog design. As will be shown, although well-intended, UI design efforts can result in unexpected violations of UI design directives and/or may inadvertently negatively impact upon test validity.
- Fifth, promising newer technologies will be described which center upon two UI research issues: (1) window navigation and structures; and (2) normative standards conformance models (oriented upon previously discussed EN normative directives). On the basis of ongoing research in these areas it is evident that the UI is becoming more and more a central target of intensive, ongoing international research and development which, inevitably, will impact upon future CBT theory and practice.
- Last, some examples of CBT UI applied and theoretical problematics will be discussed in relation to central UI issues which will have been discussed up to then.

UI Characteristics and Related Constructs

Characteristics

Regarding CBT operations, the UI can be defined as the input medium of an assessment system that has particular perceptual, cognitive and psychological valence for a CBT user. (It should be noted that the term 'user' is being used eclectically here, and can denote a test participant, assessment center candidate, job applicant, etc.) A well-designed UI facilitates accurate and valid psychometric assessment (Booth 1988a, 1990, 1991b). Objective outcomes indicative of UI design efficacy can be found in such parameters as reduced orientation time (in learning about the system), rapid user input and reduced input errors, subjective system-knowledge retention and user satisfaction with the system (Shneiderman 1987). For the purposes of this paper, primarily one major UI type will be discussed: the graphic user Interface (GUI), which can be characterized by its use of a pointing device (e.g., mouse, light-pen), drop-down/pop-up menus, windows and icons – and a wide range of objects enabling user control such as buttons and checkboxes (Hayes and Baran 1989; see also Booth 1991b; Rauterberg 1996a, 1996b).

User Mental Models

Invariably intertwined in the UI literature is the concept of the user's mental model for which there are various definitional orientations: Dutke (1987), for example, considers the construct to mean the way the user breaks use of technical devices and systems down into two knowledge areas: tool characteristics and required task outcome. (The tool vs. task concept has been central to much of the literature on UI issues – see for example, Booth 1991a, 1991b; Dutke 1994; Gentner and Grudin 1996; Hamborg 1996.)

Van der Veer and Felt (1988) orient upon varying user styles of information processing such as verbal vs. image preferences respective to syntactic/semantic aspects of a task. Jorna (1988) orients upon user operating system preference (such as MS-DOS vs. graphically-oriented direct manipulation systems). Dutke (1994) points out that mental models provide the basis for behavioral planning and management. Mental models provide a cognitive sampling, as it were, of the real world and individual mental models exist which can vary as to their focus: some are more oriented upon understanding such samplings, and others are more task oriented (*ibid*., p. 2). Ford and Kraiger (1995, p. 7) offer synonyms for the mental model concept such as 'knowledge structures, cognitive maps, or task schemata.' As they state (*ibid*.), mental models represent not only knowledge domain objects and properties, but also their functional/causal interrelationships.

Device Topology

In their work on mechanistic mental models De Kleer and Brown (1983) coined the term *device topology*. In essence, device topology is, in its dynamics, a special type of mental model: it focuses on the *surface characteristics* of a UI (or any device for that matter) which can trigger, as it were, perceptual and thought processes on the part of the user leading, in turn, to higher-order assumptions about a technical system and its use. Dutke (1994, p. 38) defines the term in detail[1]: '[By device topology is meant] the representation of the physical organization of a system – i.e., its external appearance with all its physically discernible components.' As Dutke states, the construct involves two major aspects: first, it involves the use of higher cognitive processes such as 'inference, induction and analogy'; Seel

© Blackwell Publishers Ltd 1998

(1986 p. 384) as cited by Dutke (1987 p. 38). Secondly, use of such cognitive processes leads, in turn, to the user's acquiring of an ever-evolving, newer understanding and conceptualization of the technical system he or she is using.

Platform Topology

The term *platform topology* is being coined here and represents an extension of the above-mentioned device topology concept. The term's intent is to focus upon different taxonomies, as it were, of technological paradigms which represent particular sources of influence upon user mental models regarding the perceptual, psychological and functional meaning of such technologies to the UI user:

Industrial OS standards. These involve established Operating System designs and environments by HW/SW industry leaders (e.g., MS-DOS, Windows 3.x, Win95, NT, Macintosh, etc.).

Industry UI standards. This source of influence involves *de facto* design standards of UI functionality such as button menus/panels, list boxes, etc.

National norms and directives. Here we have regulational efforts on an international (e.g., EN/ISO) and national (e.g., for Germany: DIN) level.

Diagnostic Process Stages

Respective to CBT operations, the diagnostic process can be broken down into three main stages regarding UI design considerations. As will be discussed in detail later, each stage places special responsibility upon the UI designer if reliability and validity of testing and user-friendliness of a system are to be achieved:

(1) The pre-test situation. Although in a narrow sense a pre-test situation occurs with every CBT instrument, what is meant here is the general situation involved when a user is being first introduced to a CBT system. Here, users must become acclimated to the general handling requirements of a CBT system (e.g., keyboard characteristics, mouse use, GUI layout). Additionally, various physical parameters of the user can (and should!) be accounted for, such as possible color-blindness, short-sightedness, handedness, etc.

(2) The test situation. For *each* CBT-presented test this diagnostic process stage consists of two parts: test instructions and the test itself. Test instructions involve two general informational areas: (i) general information (testing goal, sample questions and solution rationale and pacing strategies ('You should work as quickly, but as accurately as possible'); and (ii) test-specific 'handling' information (involving the user becoming familiar with help information, test-unique paging requirements – as in special multimedia environments – etc.).

(3) Post-test situation. This phase of CBT involves the termination of an individual test, or an entire series of tests during one testing session. Appropriate GUI design for selection of further tests/information (by either user or automatic system configuration, test proctor, etc.), and feedback concerning scores and their interpretation (which has been termed elsewhere as ergonomic data reduction; Booth 1986; see also Booth 1983, 1984) can play a significant role here. Additionally, user reactions to testing can be assessed at this stage.

UI Dynamics and operational perspectives

It has been argued elsewhere that the UI should not be thought of as being a static, unidimensional entity but, rather, as representing a multifaceted construct concerning human-computer interaction (Booth 1991b). Figure 1 helps to illustrate this concept.

Figure 1 illustrates a hypothetical moment in CBT where the diagnostic process focus is upon the test stage (see left portion of Figure 1).

Whereas UI design and implementation issues can be considered respective to their momentary diagnostic process focus (see above discussion of diagnostic process stages), one can also conceptualize UI issues regarding their impact upon the user during the entire testing/assessment process. In this regard the CBT UI can involve a fairly complex set of perceptual and cognitive dynamics, as described below.

UI Operation levels

As has been described in detail elsewhere (Booth 1991b), a CBT UI can be viewed as involving various operation levels or perspectives:

Tool level. Here, a user must come to grips with the UI functionality of a CBT system. In initial phases of testing (see previous discussion concerning diagnostic process stages) this involves the learning of basic UI operations such as inputting of an answer to a test question (e.g., marking a checkbox, entering text in a field), paging from one test item to the next, accessing help information, using a pointing device (e.g., mouse) etc.

Test level. This perspective deals with the user's having to work with a UI regarding an assess-

© Blackwell Publishers Ltd 1998

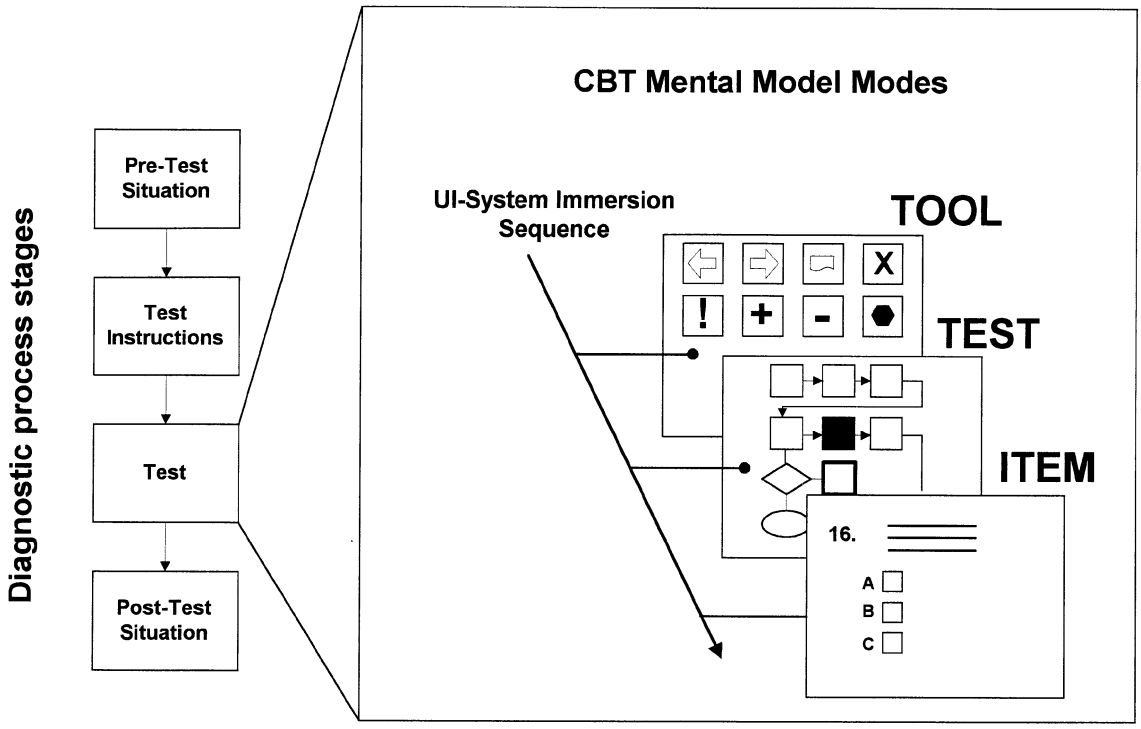

Figure 1: Schematization of UI dynamics: CBT mental model modes. (Adapted from: Booth, 1991b).

ment principle or domain such as a single test, or a subtest of a test battery. Additionally, and respective to information accessing, hypertext use and user/system event-driven navigation (as in multimedia applications) is found at this level with test and subtest branching being typical features of this level.

Item level. This involves the CBT presentation of a particular test question, and/or the assessment moment requiring specific input from the user.

UI Semantics and Syntactics

This aspect involves the valence of a UI element or perspective to the user in the CBT diagnostic process. In initial stages (as in, for example, the pre-test situation) the syntactics (i.e., the 'how' of a UI) predominate as in, for example, item paging, checkbox marking. As one proceeds in testing, semantics (or, the 'what and why' of the CBT system) gain attentional dominance (as in, for example, *strategically* navigating through a test system, accessing hyperlinked data, etc. in order to achieve a particular testing goal (see Shneiderman 1987; Ulich 1987; Jorna 1988; van der Veer and Felt 1988; Norman 1991).

UI Task Demands

In a true sense, of course, the first task of a CBT user is to learn how to use the tool functionality of a UI (mentioned above). Once this is mastered, however (i.e., a basic familiarity with the UI layout and its elements is achieved), the actual task meant here concerns use of the UI in *strategically* best meeting test situation demands. This involves using a CBT system in coping with *direct* demands (such as correctly answering the problem-solving class of a question complex (e.g., adding numbers, selecting discrepant figures out of a group of figures), and *indirect* demands (such as skipping over difficult questions at the beginning of a test session, efficiently navigating back and forth among items in order to make comparisons) in general, 'working quickly but efficiently' which have been described elsewhere as 'item response discipline demands' (Booth 1991b). It should be noted here that the user (provided the test and/or test situation is new) isn't usually capable of optimally integrating all UI elements (here: functionality) and attenuating efficiently to all UI levels (i.e., TOOL, TEST and ITEM) at the beginning of a test session. It is only when the user evolves from being a 'novice' to becoming an 'expert' in working with a CBT system that such a UI level integration, or strategically-best UI management on the part of the user is possible. (The *novice vs. expert* issue has received heavy attention throughout the UI literature. For various perspectives of this topic see Norman 1991; Heineken and Kuckhoff 1992; Rauterberg 1992, 1993; Gentner and Grudin 1996; Hamborg 1996; Kandogan and Shneiderman 1997).

© Blackwell Publishers Ltd 1998

CBT Mental model modes

The attentional focus of a user during CBT can vary considerably regarding the above-mentioned perspectives. For example, in answering a test item, the user must consider a test question (ITEM level) within the context of a series of questions of a subtest or test (TEST level) and then correctly enter an answer and then page to the next test question (TOOL level). Depending upon the type of testing involved, this attentional oscillation (i.e., regarding the tool, test and item perspectives) can be relatively simple (as in the solution of a test item involving Likert-type multiple-choice input to, say, a vocabulary question), or become fairly complex – as in the case of multimedia applications and complex problem-solving scenario assessment situations. In brief, however, all of these UI levels or perspectives can play a role in the user's understanding of and use of a technical system (here: CBT UI). Thus, and regarding CBT operations, as the user's attention to the UI can focus upon one or the other operational/functionality perspective – within a general assessment situation – it is being proposed here that, regarding CBT operations, instead of speaking of mental models, the term mental model *modes* may be more appropriate.

Thus, and to summarize UI dynamics in the CBT situation as discussed in this section, the UI requires on the part of its user mastery of the Tool, Test and Item levels which can have different valence for the user depending upon the diagnostic process stage or testing situation demand involved. For example, the Tool level could have more syntactic valence at the start of a test session (as in, for example, the pre-test situation where one must learn how to page forward and backward to test items), and then attain more semantic valence later on in the same test session (as in the case of a user's using the functionality of the Tool level to strategically navigate to special items, call up special help information etc.). Additionally, although it is being postulated here that the UI-System immersion sequence (i.e., the perceptual and cognitive 'focus' of a user regarding use of a UI and its various levels) normally takes the Tool-to-Test-to-Item path sequence (see Figure 1), once testing is 'under way' attentional oscillation regarding these levels can occur at practically any testing moment, with the user's UI mental model switching from one UI mode to the other.

UI Design Directives

CBT is never conducted in a social vacuum. Beyond issues of scientific methodology and psychometric testing and assessment, use of computer technology in training and assessment is ever-increasingly becoming subject to industrial constraints regarding GUI design and SW/HW system implementation (Booth 1977a, 1977b). In this regard, two major directives will be briefly discussed which have become mandatory guidelines for European Common Market members respective to user-friendliness issues of SW. The directives differ as to their focus and applicability. Both, however, have impact upon design and implementation of UIs.

Directive 90/270/ECC

On 31 December 1996 the European guidelines for computer-monitor use (EU 90/270/ECC) became law. As described by Zwingmann (1996), the directive contains new directions for work safety which, among other aspects, set minimum safety standards for all employees – whether in private industry, government or general public/civil service. Three key aspects of the directives place particular responsibility upon employers:

- The directives are required to be *systematically* implemented (which implies the necessity for well-conceived work environment assessment and control measures regarding work safety and ergonomics respective to computer use).
- The regulations are to be implemented and integrated into *all* industrial organizational structures and processes (to include governmental, industrial, military, general public and private institutions).
- Employers are required to become informed about the state-of-the-art of developments in science and technology regarding office equipment and technology environments and are obligated to pass on this information to worker-representatives (e.g., union officials).

In addition to ergonomic requirements respective to HW, general office equipment and work environments, the regulations also establish guidelines regarding man–machine interface aspects of SW (see SANUS 1996). Of particular salience here are the requirements that:

- SW must be appropriate for the task at hand.
- SW must be user-friendly and fitted to the knowledge and experience of the user.
- The SW/computer system must keep the user informed as to the process at hand.
- System information must be presented in a format and pace which is fitted to the user.

EN 9241 Part 10

This directive (Ergonomic requirements for office work with visual display terminals, Dialogue principles, DIN 1996b) replaces the former

© Blackwell Publishers Ltd 1998

directive DIN 66234-8 (DIN 1988) which, essentially, covered the same major ergonomic issues.[2] The newer directive, however, has added two new general principles to dialog requirements. The newly-added principles involve *individualisability* of SW systems and *learning suitability* (which, in their thrust, are very similar to provisions of the above-mentioned 90/270/ECC directive.

Although a detailed discussion of European laws and normative directives cannot be attempted here, suffice it to say that both directives pose implementation problems respective to their being applied to CBT. For example, Directive 90/270/ECC requires that SW be tailored to user knowledge and experience. EN 9241-10 specifies that, where possible, functions should be automated, redundant operations should be avoided and the user of a computer system should have a good overview of where he or she is 'at' in the used system. As will be seen, the achieving of such goals in CBT can be very problematic.

Illustrative UI Design Approaches to one CBT operational aspect: Paging

Given the above background considerations, it is now time to see how such aspects can become involved in GUI design considerations. Accordingly, two instances, or scenarios as it were, will be offered which involve standard CBT approaches to GUI design regarding item paging. The examples demonstrate that the GUI designer, if he or she is not careful, may unwittingly come into conflict with certain recent ISO/EC directives regarding user dialog design requirements. Further, it will also become evident that very often good GUI design requires 'trade-offs' regarding pragmatic application and psychometric problematics and, further, that conflicts between good UI design goals and such user dialog directives may, at times, be unavoidable.

Occupational aptitude testing and counseling: Project DELTA

Figure 2 represents a typical test item presented in a standardized GUI of the DELTA CBT system of the *Bundesanstalt für Arbeit* (Nürnberg, Germany; BfA 1995). The test item involves mechanical comprehension. The user must decide in which direction the weight (see 'y' in the figure), and the device arm-extension (see 'x' in the figure) will move when the handle is depressed (see the largest arrow, left-middle in the figure).

Regarding item solution, and as one can see,

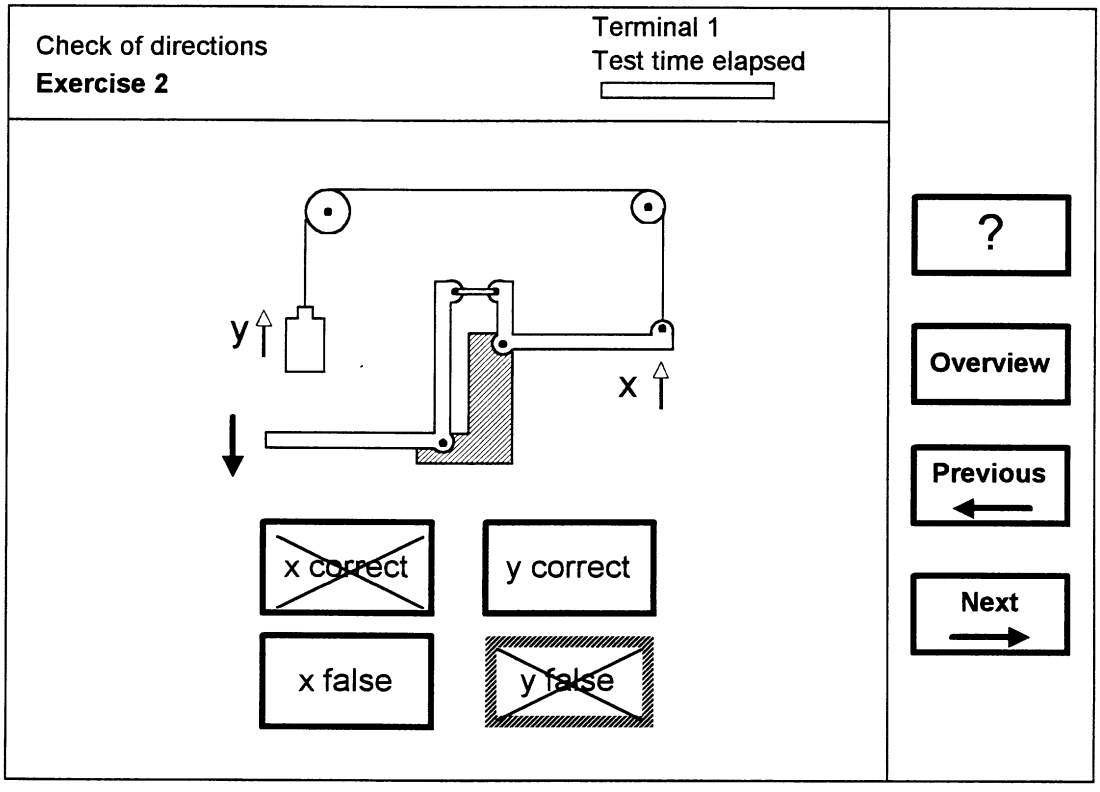

Figure 2: Test item of the DELTA CBT system. BfA (1995).

© Blackwell Publishers Ltd 1998

when the handle is depressed, the device arm-extension (x) pivots upward, which results in the weight being *lowered* (y). Accordingly, x, as a direction, is correct and y (as a direction – see the arrow adjacent to the y) is *incorrect*. As the figure shows, 'x correct' and 'y false' have been selected as being the answers to the test item.

The right-side of the GUI contains a button panel consisting of various functions:

(1) With the [?] function[3], a user can call the test proctor's attention via networked communication.
(2) The [Overview] function allows the user to see what items have been answered and which are still open.
(3) The [Previous] and [Next] buttons enable paging to lower-numbered or higher-numbered questions.

Figure 3 represents an instance where a user has activated the above-mentioned [Overview] button in order to see where he or she 'is' regarding answering of the test items.

As illustrated in the figure, the user is presently at item 30 (see inversed square, lower-right), and has still to answer items 11 and 19. If, for example, the user wishes to proceed to item 11 from the current position, he or she only needs to mouse-click on square 11. This results in the Overview dialog being removed and automatic branching to that question. If the user decides, however, to remain at the current position (i.e., item 30), he or she merely has to click on square 30, and that item will be redisplayed.

The 'Overview' function of the DELTA GUI (illustrated in Figure 3) represents an attempt to meet what has been termed elsewhere as Intertechnological Validity Congruency (IVC; Booth 1991b). In essence, the GUI design of Figure 3 represents an effort of DELTA project staff to effect a 'status transparency' similar to that of pencil-paper technology. That is, in paper-pencil testing users (via answer sheets), have, automatically, an efficient listing of items which are unanswered. The present GUI is intended on emulating this effect.[4] As such, it has a dual function: it informs about test reply status *and* functions as a navigation device.

Although well-intended regarding IVC (see above), the 'Overview' function of the DELTA GUI *could* be considered as violating provisions of the EN ISO 9241-10 directive regarding dialog design (DIN 1996a, 1996b) – depending upon definition of the user's informational and navigation needs during testing:

(1) First, and regarding a very common test behavior, a user might well be interested in *solely* branching to and reviewing *unanswered* items of a test such as is often done when sufficient testing time remains and a user wishes to review his or her answers. The only way to do this via

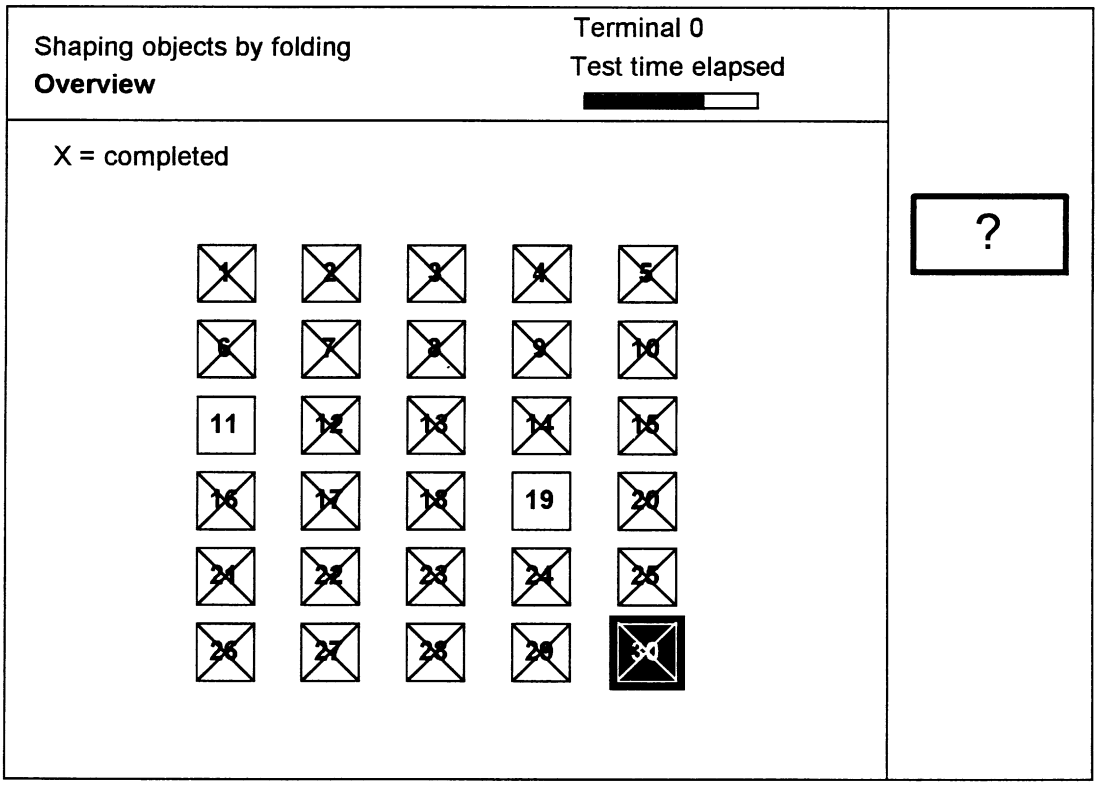

Figure 3: Example of the overview function of the DELTA CBT system. BfA (1995).

© Blackwell Publishers Ltd 1998

the DELTA GUI is for one to use the [Overview] function: one activates the function's dialog and then selects the item to be branched to (in this case, an open question). This, however, requires that the user 'pass through' the Overview dialog each time an open question is to be selected and branched to. Section 3.2 of the 9241-10 directive (*task suitability*) requires, however, that dialog systems should eliminate unnecessary work (application example 9)[5].

(2) Second, and interpretationally more difficult to evaluate as to *possible* 9241-10 non-conformance, section 3.2 (*task suitability*, application example 1) requires that a dialogue should present only that information needed for completion of a task. Further, application example 5 (same section) specifies that the format of dialog output should have a level of precision that is appropriate for a specific task. This particular design (i.e., the Overview function) *possibly* violates this design principle in that the present dialog design is too complex regarding the discernment of unanswered questions: as can be seen in Figure 3, unanswered items (being in the minority) are more difficult to discern at the final stage of testing. Accordingly, if the user of the DELTA system is primarily interested in detecting unanswered questions, then, perhaps, only unanswered question checkboxes should be displayed.

In contrast to the above, the following UI paging design for a different test system solves some problems as have been mentioned above regarding the DELTA system. (But, as will be discussed, potentially creates new ones!)

Intelligence testing: FRT

Figure 4 shows an illustrative test question of a parallel form of the computerized figure reasoning test (FRT 1997). The original FRT (Daniels 1962) and the newer parallel form to it, is a non-verbal test which measures general intelligence.

In appearance, FRT items (there are 45 in each test) are similar to the non-verbal Raven series (Raven, 1958). An FRT test item consists of a matrix of eight figures (a ninth cell of the 3×3 matrix is empty). The user must select out of six alternatives (A–F) the figure element which completes the logic of the matrix. In this case, alternative 'E' has been checked (see radio-button 'E' in the 'Select an answer' alternative group of Figure 4). The solution rationale of this item (which is, actually, the fifth FRT practice item; for expositional and test security purposes, the item has been renumbered here) is simple: item patterns here differ as to their rotation and figure-element shading. For example, and to use the upper-left figure as the starting point, as one proceeds left-to-right, or top-to-bottom, each figure in succession is rotated one quarter-turn clockwise. Additionally, only one figure in a row or column has a gray-shaded figure element.

In addition to buttons allowing paging to previous and subsequent items ([<], [>]) two buttons of the FRT GUI allow rapid paging to unanswered questions: ([< ?], [? >]. The procedure is simple: let us assume that we are, as was the case with the DELTA test (see above discussion), interested in branching to items 19 and 11, and that we find ourselves at a higher-numbered item (here: 30). All we have to do to branch to those items is to activate the [< ?] button – which branches the test automatically to the next open item 'to the left' (in this case: item 19). After answering that item, we merely have to activate the [< ?] button again to proceed to item 11. In brief: this GUI design allows the same branching power of the 'Overview' function of the DELTA system respective to adjacent unanswered items – with one, instead of two, mouse-clicks. Further, it obviates the need to activate any dialog pass-through for such navigation.

In addition to allowing rapid paging to unanswered questions, the FRT GUI also allows rapid paging to items which have been 'dog-eared' by the user via the 'Marker' field radio-button of the FRT GUI (see Figure 4): if a user considers an item to be critical or particularly valent (especially hard, containing a problem-solving pattern of special interest, etc.) the user can mark the radio-button via mouse click. As Figure 4 shows, the user has thus marked item 30. To procede solely to marked items, the FRT user needs only to press the toggle button below the two quick-branching buttons in order to have their function 'toggled' from unanswered item branching to marked item branching. Once that button is pressed (or, 'toggled') the labeling of the quick-branching buttons changes from [< ?], [? >] to [< !], [! >]. The quick-paging function for marked items is intended as an IVC feature for FRT testing via PC: if a user tends to mark various item numbers on his or her answer sheet during paper-pencil testing, he or she should be allowed to do the same in CBT. (As a side-note: one could substitute the toggle button with another pair of buttons which would be placed below the open-question branching buttons. As shown here, no GUI space is saved via the use of the toggle button concept. This toggle function, however, is an ongoing GUI design prototype. In certain situations such a toggle-button function would be helpful in reducing 'surface cluttering' of the GUI. For example, one FRT GUI toggle prototype is tiered for several branching levels: one can start with left/right branching, and then toggle respectively to open left/open right, marked left/marked right, first/last questions. Thus, such

© Blackwell Publishers Ltd 1998

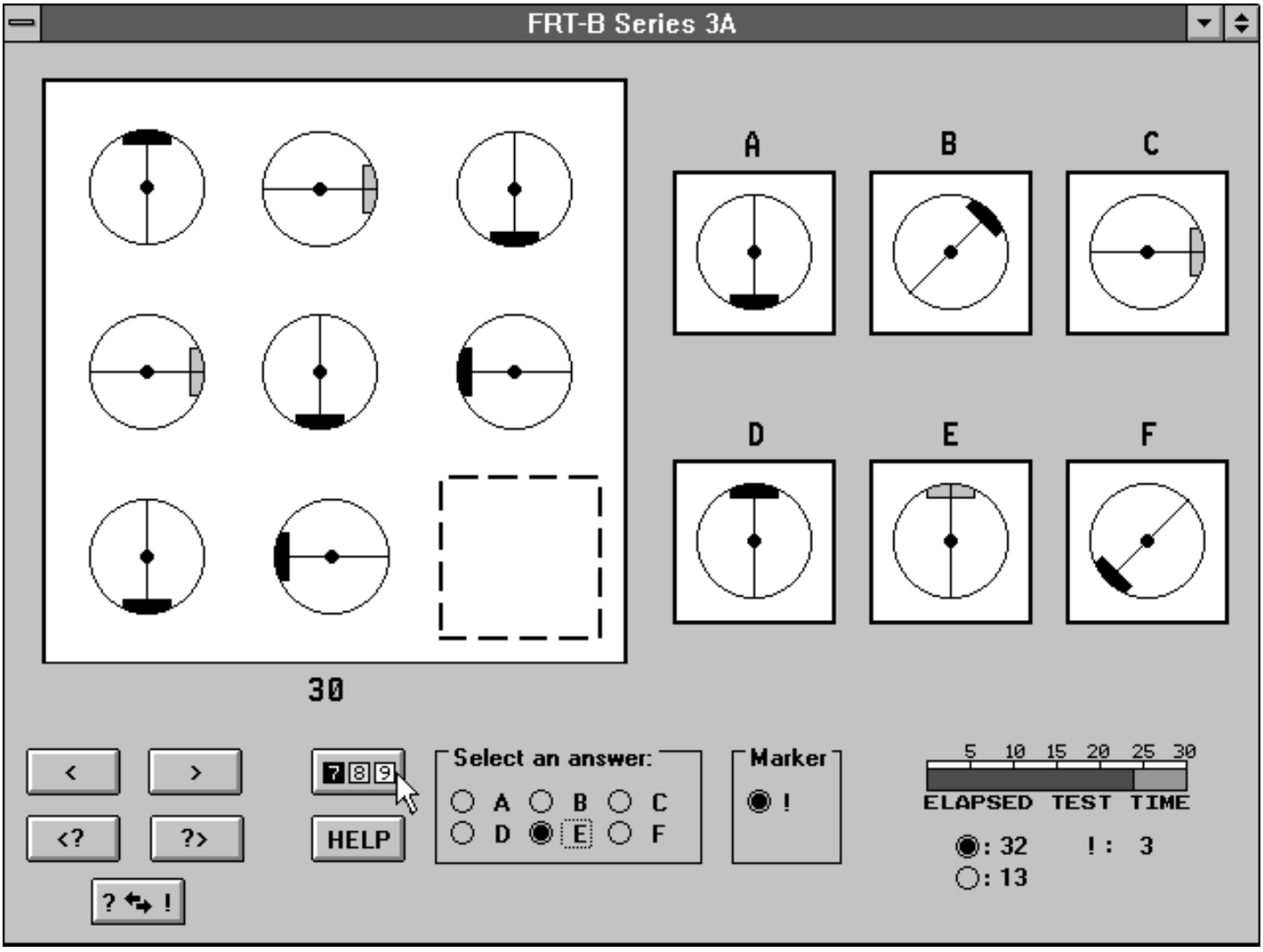

Figure 4: Standard GUI of the FRT. © 1998, John Booth Consulting.

a button function could eliminate having to place additional pairs of branching buttons for each desired branching level.)

As does the DELTA system, the FRT also allows a user to gain an overview of the entire test respective to item status: the user merely has to activate the Item Status function of the FRT GUI, which is the button to the right of the [>] button (see Figure 4: the mouse cursor is pointing to the button); when that button is activated, the display in Figure 5 is the result.

As can be seen in Figure 5, the user is at item 30 of the FRT (see large arrow next to item number 30). Additionally, 'below' item 30 there are three unanswered questions: 11, 16 and 19. No items beyond 30 have been answered. Three items have been marked by the user (11, 30, 32). In order to navigate to any listed item of the display, the user merely needs to click on an item number; when clicked, a confirmation dialog appears (see lower-right of the figure) which enables the user to re-select if necessary, or cancel the operation). If branching is desired, the user presses [OK] and the dialog is removed and the selected item is branched to. Activation of [Cancel] blends out the confirmation dialog and the user can make a new selection. Activation of the button-field for the current item (in Figure 5: item 30) returns the user to the current item.

Non-designed GUI use factors

Regarding quick-paging to open questions (as has been described here regarding the FRT GUI), it should be noted that open-question branching can be used to effect, so to speak, a GUI 'streamlining'. For example, the open question function (i.e., [< ?], [? >]), can be used during an entire test from the start if one would so choose: all items are unanswered at that point. Thus, use of only that set of buttons for item paging

© Blackwell Publishers Ltd 1998

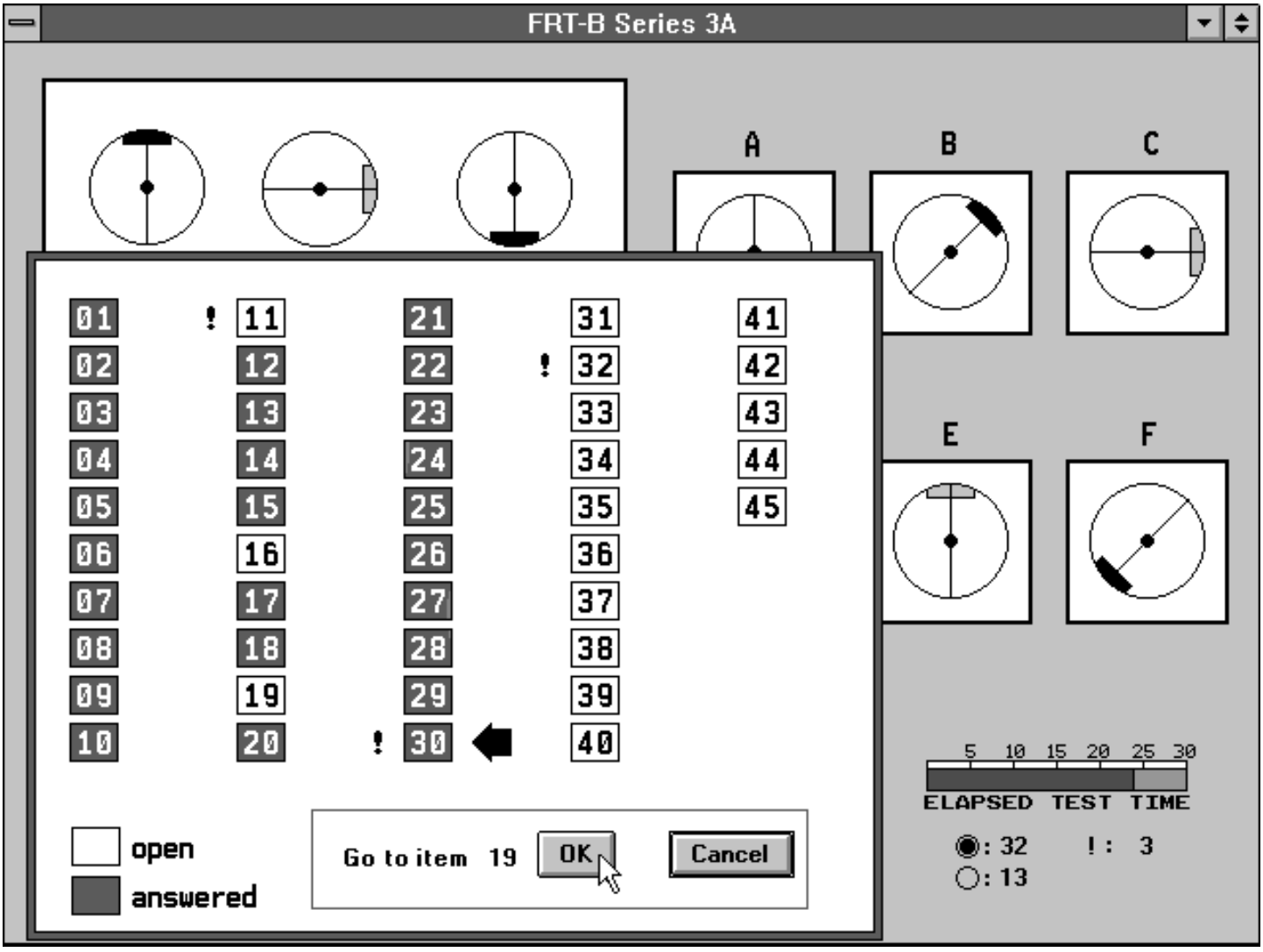

Figure 5: Test Status display of the FRT © 1998, John Booth Consulting.

(rather than the regular [<] [>] buttons) can simplify, in a sense, item navigation: only one set of navigation buttons would, thus, have to be used.

The open question function can also be used to 'mark' positions in a test. For example, a user can remove an answer to a higher-numbered item and then back-page item-for-item to a previous item in order to compare its pattern with the harder question (higher-numbered questions in this case, being more difficult). In this procedure, the answer for the lower-numbered item would also have to be removed. Then, by 'toggling' back and forth between the two items (using alternatively the [? >] and [< ?] buttons, a comparison between the two items could be more efficiently effected than would be possible via item-for-item paging or item accessing via the Item Status function. Being satisfied with the comparison, the user could then re-enter an answer for the lower-numbered item and then (using the open-question function) quickly re-navigate to the higher-number question which, then, would be re-answered. Such GUI 'streamlining' as mentioned here (which has been detected in post-test interviews and reading of UI use protocol analyses; see Booth 1991b), suggests that the device topology of a UI (in this instance, a previous FRT GUI using such open-question quick-branching) can differentially trigger user perception and implementation of a GUI beyond what may have been design-wise intended and, thus, may enable some users to utilize a GUI in a particularly efficient (and advantageous) manner. (As such strategies were not originally intended in FRT GUI design, research or development, they are being termed here as non-designed GUI use factors; although unintended, they may play a role in CBT UI use-impact regarding reliability, validity and user satisfaction.)

© Blackwell Publishers Ltd 1998

Interpretation problematics

Respective to national norms and directives, it will be recalled that aspects of the DELTA GUI have been characterized here as being *potentially* in violation of the 9241-10 directive. For example, and to focus again upon its Overview function, if the DELTA system were to be predominantly used by power users (defined here as persons possessing considerable PC experience and/or having high-ability regarding strategic use of a GUI in a particular test situation, see previous discussion of this topic respective to UI dynamics and task demands) the present DELTA GUI could be, in this regard, characterized as being 'underpowered': as previously explained, it doesn't allow for rapid branching to unanswered or marked questions. On the other hand, under certain circumstances the FRT GUI could represent functional 'overkill': its four-tiered branching strategy (i.e., [<], [<?], [<!] and Item Status) could very well be overly complicated for users having little or no PC experience. In essence, both the DELTA and FRT GUIs can be considered good or deficient designs – depending upon their user audience and intended test/assessment goal. The point being made here is this: no UI is suitable for all user populations in all CBT situations. This fact is recognized by the previously mentioned European directives. In its introduction, for example, Directive 9241-10 states that in the specification, development or evaluation of any dialogue system, the directive can be applied only as a general guideline. The actual implementation of the directive respective to dialogue principles depends largely on the intended user, tasks involved, environment and used dialogue technique (see DIN 1996b, p. 3). In essence, and regarding such UI normative specifications as the 9241-10 directive, the UI designer is being challenged to fulfill the 'spirit of the law' of such directives while meeting particular CBT psychometric and assessment goals.

Evolving UI-Related Platform Topology

Up to now, most argumentation and examples respective to UI problematics have centered upon fairly conventional uses of PC technology. Newer technologies, however, are being developed which may also impact upon CBT UI design and implementation. In this regard, two examples of evolving platform topology (PT) will be briefly discussed. The first deals with the desktop metaphor (see Dutke 1994; Gentner and Grudin 1996; Norman 1991; Rauterberg 1992, 1993; Shapiro 1995) which has permeated the software industry and, in the extensive use of 'windowed' technology has become an industrial OS standard. The second PT influence concerns efforts to use PC technology in standardizing evaluation of user interfaces respective to 9241-10 directive conformance.

Window management

Seymour (1996) has pointed out certain advantages of working with larger PC monitors respective to heavily graphically-oriented SW so that one can use multiple windows more efficiently than is the present case with smaller screens. As anyone with enough PC experience knows, the logistics of keeping track of what is on the screen (because of window overlapping and non-structure) can be difficult. Additionally, users of the World-Wide Web (WWW) and other hypertexted systems often find that hypertext hierarchies and linkage are not without their own problematics: many individuals get quickly lost in using their WWW browser or in using multimedia-based hyperlinked information. How can one develop more optimal GUI design regarding *both* of these factors? That is, how can one provide for better user management of large computer monitor screens (and concomitant multiple window use), along with improving structure and depth cues for windowed information? The 'elastic windows' concept focuses on both of these issues.

Figures 6 and 7 schematically illustrate an approach being taken by Professor Ben Shneiderman's human factors team at the University of Maryland Human-Computer Interaction Laboratory (HCIL) which has been termed by its authors as 'Elastic Windows' (EW; Kandogan and Shneiderman 1997). In short, the design is intended on aiding the GUI user in navigating through large amounts of information in multiple windows while, at the same time, affording a high degree of informational structure transparency.

Figure 6, for example, represents a computer monitor screen which contains newly-opened pages of information. The main window of the display (see '1' in Figure 6) concerns project information about four current projects. Two projects have been selected (see '2' in the figure). The names are, actually, hyperlinks to further information, and their selection (indicated by inversing) has resulted in two additional windows being displayed (see '3' in the figure). Each of the two newer windows (which are child windows to the 'Project Information' parent window) has its own window label bar ('Alpha', 'Bravo') which has a different color shading than the parent window and aids the user in differentiating window grouping levels. The tileing of the two child windows indicates that they are contained in the same hierarchy level. In brief, Figure 6 represents a two-level windowed

© Blackwell Publishers Ltd 1998

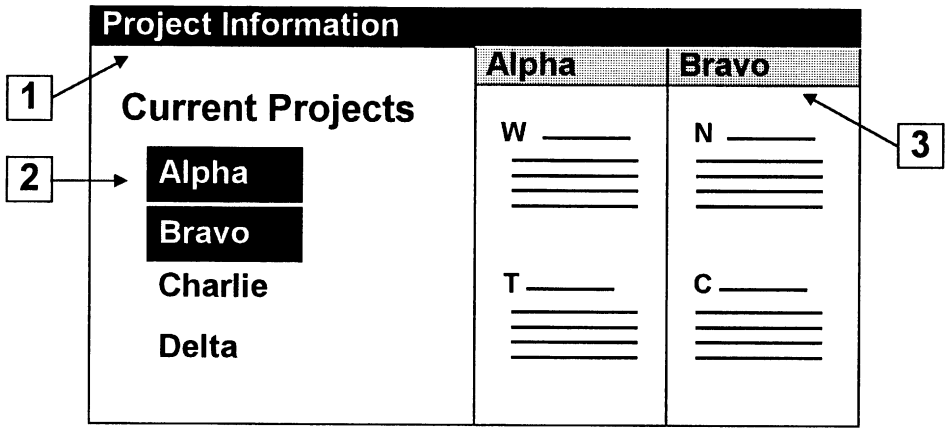

Figure 6: A schematized example of the Elastic Windows GUI: 2-level hierarchy.

hierarchy: windowed information listing current projects represents the first level; the windowing of both Alpha and Bravo project information represents the second level.

Figure 7 represents a different informational hierarchy: here, the main window concerns personnel information of a 'Project Alpha' of which a subtopic (here: 'Research'; see '1' in Figure 7) has been selected. This has resulted in a child window ('Research') being opened within its parent. A further selection of a topic in that window ('Section 3'; see '2' in the figure) has, in turn, resulted in a further window being opened ('Section 3'; see '3' in the figure). In brief, there are three hierarchy levels here: 'Section 3' which is within 'Research' which, in turn, is within 'Project Alpha' – representing a form of hierarchical *nesting* of windowed information.

HCIL staff report that the EW procedure results in more user-friendly use of large monitors and that users can flexibly organize information, restructure it and browse through it – while retaining a sense of location in the information structure (Kandogan and Shneiderman 1997).

ISO 9241 Conformance Evaluation

The author can only sympathize with GUI designers when they try to come to grips with systematically evaluating their dialog designs in accordance with the ISO/EN 9241 directives. Initial steps are being taken in providing structured and user-friendly solutions to this problem: for example, Professor Reinhard Oppermann and his staff of the German National Research Center for Information Technology (Sankt Augustin) have developed an evaluation procedure based upon the ISO 9241 directive which is designed to allow an analysis of any SW system as to its conformity with that standard. The system is designed to afford a practical but comprehensive evaluation of user interfaces. The ISO 9241 Evaluator is an expert-based evaluation system that uses pre-structured guidelines which involve about 300 test items.

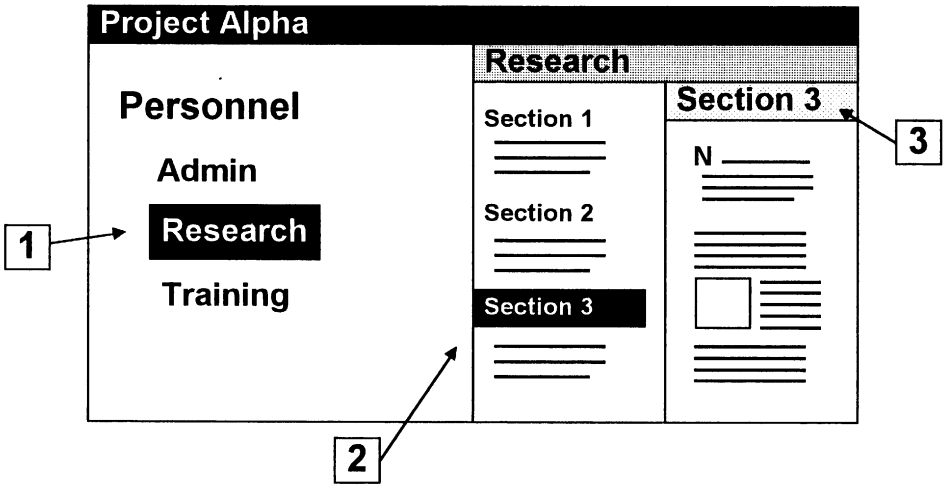

Figure 7: A schematized example of the Elastic Windows GUI: Hierarchical nesting.

© Blackwell Publishers Ltd 1998

The items are oriented upon assessing the conformance of a software system to basic requirements of part 10 of the standard. In addition, test items pertaining to provisions of parts 12–17 of the ISO 9241 may be included in the evaluation as needed. (EVADIS 1997).

The intended user of this system is the human factors expert who orients 9241 Evaluator assessment primarily on software ergonomic principles rather than the specific task of the targeted system. In practice, the system is used as a standardized checklisting regarding conformance of an SW system's major components (e.g., input/output dialogs, help information and error handling, functionality of the system regarding its utility, system reaction time, technical interfaces, etc.) in regard to criterion specifications of the 9241 directive (e.g., error tolerance, learnability, task suitability etc.). In effecting sufficient evaluation coverage of an SW system respective to directive conformance, an SW system can be evaluated in regard to the checklisting of the SW system in various application situations. Accordingly, a text-processing system could be first evaluated as to operations in document formatting, then evaluated according to printing and file transfer operations etc. Thus, eventual weaknesses of an SW system can be detected and differentiated regarding any non-conformance that is context specific or generally system imminent. As the authors report (EVADIS 1997), the system results in ISO 9241 conformance analyses that are relatively fast, resource economical, provide system transparency and can be used on a wide range of GUI/system aspects. Psychometric evaluation of the 9241 Evaluator as to its reliability and validity has not yet been conducted but is expected within this year.[6]

Applied and Theoretical CBT UI Problematics: Some open issues

By now it should be clear that the UI is a key element of the CBT diagnostic paradigm and, regarding implementation and design issues, can involve a considerably complex spectrum of psychological and cognitive aspects. Although no comprehensive review will be attempted here, a discussion of some applied and theoretical problematics will be offered respective to central UI issues which have been discussed so far and which, on the basis of the author's experience in CBT research and development, are considered to be especially problematic and deserving of future research effort.

Platform topology issues

Window management. Regarding industrial OS standards, the current direction of HW and SW technology developments seems to be pointing to an increased use of larger PC monitors and richer graphical environments. In this regard, the orientation of the Elastic Windows (EW) concept is intriguing and represents a possible strategic alternative in affording improved user management of such graphically-based windowed environments. But: do we need this much 'push' in technology for CBT? For example, tests such as the FRT and those of the DELTA system don't require such a direction. To be sure, other test designs (such as those using multimedia techniques) could conceivably profit from the EW approach. But before we become much enamoured of such newer technology, shouldn't we find out if there are limits to the number of windows that CBT users can comfortably work with? Are there testing paradigms (such as, perhaps, adaptive testing) where such technology could be potentially paradigmatically damaging or psychometrically inappropriate?

OS platform mixing. In a study of the impact of user interface design upon the learnability of text processors, Altmann (1987) found that learning transfer was better for users of a DOS-based text processor when they had to subsequently learn a graphically-oriented text processor system (here: Macintosh) than was the case for Macintosh-experienced users having to learn how to use a text-oriented, DOS-based text processor. In this regard, one can well imagine a situation involving testing under mixed operating systems, where testing might start with an MS-DOS oriented UI, to be followed by tests running under a MS-Windows (3.x) and then, say, a Win95 or UNIX platform. What would be the impact upon assessment reliability and validity – and user attitudes toward CBT – in using such mixed UI platform topologies in the same testing battery?

Monitor surface user input locus. As has been argued elsewhere (Booth 1991b), in a testing battery involving use of various PC-based tests, inter-test consistency regarding GUI design should be a prioritized goal, for consistency in UI design can significantly aid in the establishment of a uniform and standardized testing environment. In this regard, however, a major limitation of current PC technology which hampers such UI standardization lies in limited PC monitor 'real estate' (see Seymour 1996): i.e., very often a CBT UI designer is faced with the problem of having to figure out how to optimally place test content (e.g., graphics, text, etc.) onto the monitor screen while also providing for user-friendly and ergonomically optimized UI design regarding button controls, input fields and system status capability (such as

© Blackwell Publishers Ltd 1998

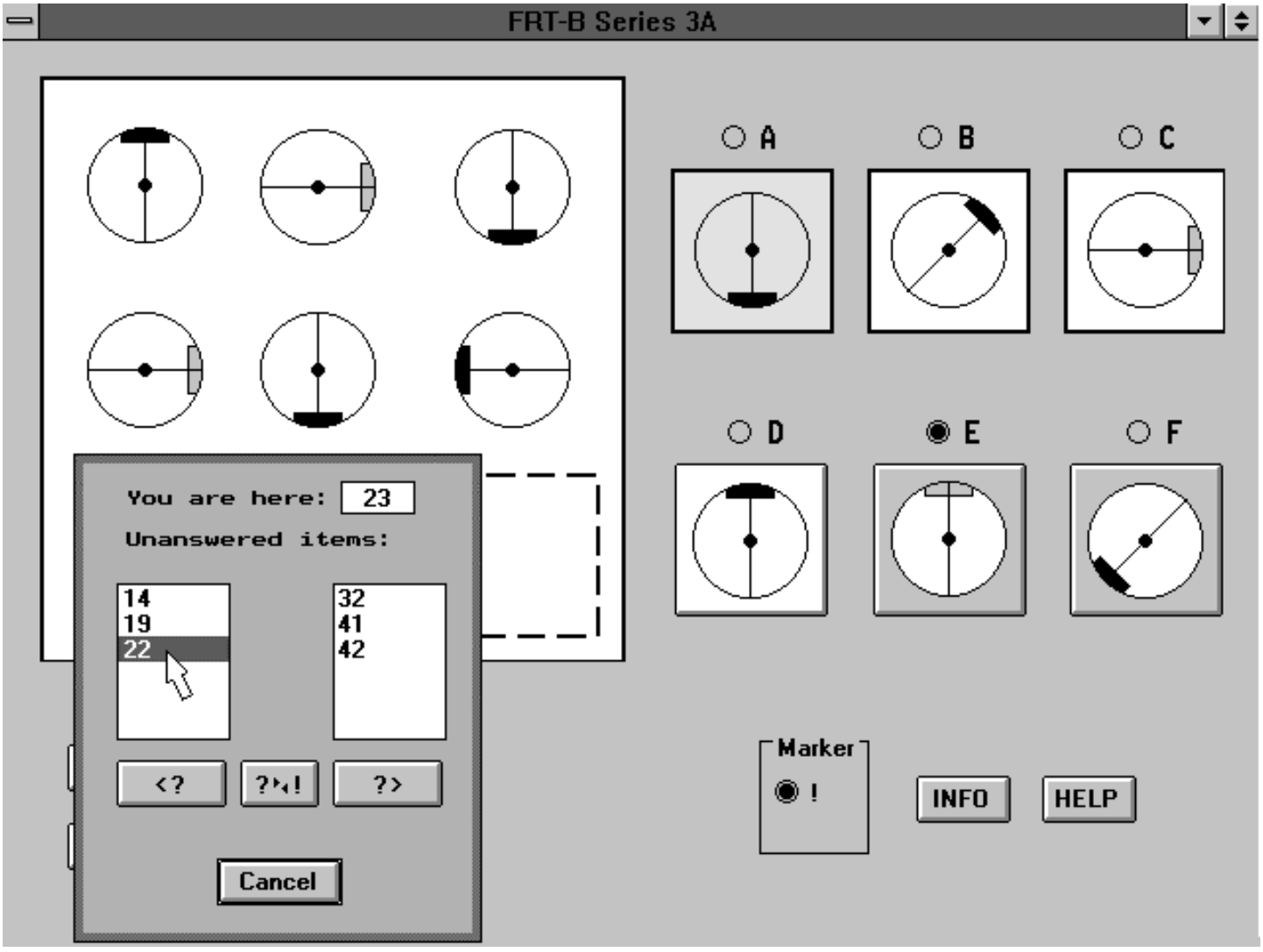

Figure 8: Prototype GUI of the FRT. © 1998, John Booth Consulting.

displaying time remaining in a test, the number of items already answered, etc.). In this regard, a UI design alternative which makes more efficient use of monitor real estate can be found in incorporating user input functionality into the test content screen area. Figure 8 helps illustrate this principle.

As can be seen in the figure (see Figure 8, middle-right, D–F), the item alternatives are imaged to button surfaces. Instead of having to mark one's answers to a test question via a checkbox group (as in the GUI illustrated in Figure 4), the user of the GUI illustrated in Figure 8 merely has to activate the button containing the image of choice (i.e., buttons D–F). One advantage of such an input design approach lies in the freeing of monitor space. For example, use of such button input for alternative choices obviates the need to use a checkbox group for such input. Additionally, as the alternatives have to be depicted anyway, their placement upon buttons is not effected via any significant loss of screen real estate (although buttons require a slightly larger surface area to accommodate for their 3D imaging and sufficient peripheral margins between the involved figure and button edges). Given that tests can vary considerably regarding the number and type of their response alternatives, however, such an approach could result in a user's having to, so to speak, readjust from GUI to GUI regarding the location of such input response buttons (or input locations: the response alternatives could also be made into 'hot areas' sensitive to mouse-cursor double-clicking – which is the case for alternatives A–C in the figure) in a test battery using tests of varied structure and appearance. (It should be noted that Figure 8 represents a design prototype. A mixing of 'hot-areas' and imaged buttons would not be a suitable design. Also, the first alternative (i.e., 'A') has a different colored surface (actually, yellow) which was used to

© Blackwell Publishers Ltd 1998

determine if accentuating buttons with that color would better effect user input regarding their locating of mouse-sensitive hot-areas. Further, the dialog at the lower-left of the figure is an alternative test-item status display to that discussed in Figure 5: here, only items falling into a searched category are displayed in left or right listings. This design is intended on meeting EN 9241-10 requirements that displayed information should be kept at an ergonomic minimum.

In essence, we are dealing here with a major CBT design choice principle: either one is interested in confining user input and control of a testing system to a consistently used area of the computer monitor (as in the case of the FRT and DELTA UI designs), or one is willing to 'mix' monitor surface use in this regard. The advantages or disadvantages of either approach, and the potential impact of mixing the two approaches in, say, a test battery, would have to be ascertained via empirical investigation.

The diagnostic process

Although, as previously discussed, the diagnostic process can be broken down into several phases regarding GUI aspects, attention here will be focused upon the pre-test situation.

It has already been seen that the pre-test situation should include a careful informing of the user regarding system requirements. Additionally, sufficient user-practice regarding use of the keyboard, mouse, touch-sensitive monitors, etc. is often necessary at this diagnostic process stage. In the case of few tests, such a pre-test situation (with its emphasis upon user PC-system training and 'acclimatization') can take longer than testing itself. In this regard, (and regarding the eventual acceptance/rejection of such technology in the ubiquitous cost-benefit atmosphere of industrial settings), the development of efficient, encompassing, psychometrically justifiable and user-friendly acclimatization procedures which minimize system familiarization time are, and will remain, a major challenge to the CBT system designer. Other problematics, however, can also be found concerning this stage of the diagnostic process:

Differential user knowledge of GUI design conventions. Although general advances in PC technology and applications of same are, in this ever-increasingly internet-oriented world, becoming more evident to the general public, one should not be tempted into taking user knowledge of general GUI design conventions for granted: GUI design aspects, which one could think would be self-evident, can become stumbling blocks regarding user understanding

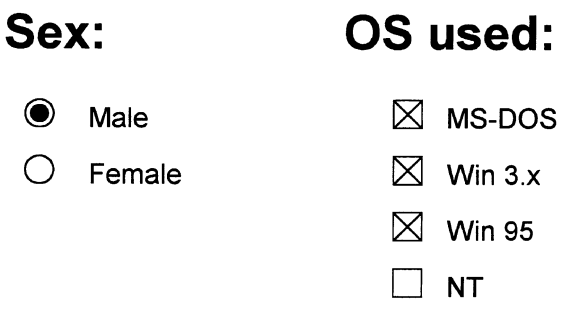

Figure 9: Radio buttons and checkboxes.

of a computer system's use. For example Ashworth (1996) reports a study where Macintosh users were found to be lacking in their general understanding of two industry-standard GUI graphic conventions: radio buttons and X-boxes (or, checkboxes). Examples of both are given in Figure 9.

By industry convention, radio buttons concern parameters or dialog situations involving mutually-exclusive events. For example, the radio button options of Figure 9 involve selection of gender. Once one radio-button option is selected, it becomes marked and any other previously-selected option in a radio-button group becomes de-selected. Checkboxes, on the other hand, involve selections which involve options any of which can be selected. Thus, and as illustrated in Figure 9, one could be presently using none or any of the operating systems listed in the figure. As Ashworth reports, even though the Macintosh system is well-known as a graphically-based system, surprisingly not all subjects of her study were aware of the selectional difference between radio-buttons and checkboxes. Thus, and as Ashworth points out, GUI designers should not assume that users are familiar with such industry-standard GUI conventions regardless of their previous computer-use experience. This would suggest the need of reviewing such GUI conventions with CBT users in the pre-test situation.

Is there a need for GUI-use coaching? It will be recalled that the FRT GUI allows very rapid branching throughout the test and, if used 'strategically', can even result in very efficient navigation for item comparisons. But: for some subjects this could result in shorter test-taking time, better test overview, etc. Could this reflect a type of 'GUI bias' for sub-populations of CBT test users? Should we inform users about such 'hidden GUI agendas', in order to afford them a fair chance regarding knowing or not knowing about such GUI use – or would this information make the testing/assessment situation overly complicated for some CBT users?

© Blackwell Publishers Ltd 1998

UI Dynamics

By now it should be clear that human-computer issues respective to UI dynamics can be considerably complex. This is particularly so for the field of CBT. Although no attempt can be made here to comprehensively deal with such issues, some aspects concerning UI dynamics as applicable to CBT will be raised below which reflect applied and basic research and practice topics deserving of empirical scrutiny.

Can UI characteristics unintentionally bias user perceptions of CBT goals? We know that 'official' test instructions should include information for a CBT user as to the appropriate test-taking strategy regarding speed and power tests. Research suggests, however, that the nature of the CBT situation (and perhaps GUI designs themselves) can send 'cues' to the user – which may result in lowered assessment scores and poorer test performance. Neubauer and Urban, for example (1991) found in a study involving paper-pencil vs. CBT testing involving the Advanced Progressive Matrices (Raven, 1958), that subjects using computers got significantly lower scores (total scores and means); they were also quicker in their work and, apparently, more careless than the paper-pencil subjects. Neubauer and Urban suggest this outcome may have been due to the 'challenging' nature of PC technology, in that the PC-using subjects felt the need to work more rapidly under CBT conditions. Questions here, of course, concern how we can control for such a possibility – and if so, when. Before testing? During? After?

What impact does a GUI have upon its user's mental model in CBT problem-solving situations? Ulich (1987) conducted a series of studies investigating the impact of dialog flexibility upon computer-based office work. The goal of the studies was to effect an adaptive software system that could be fit to user experience and preference for various dialog forms. In one experiment (involving the simulated processing of customer orders), Ulich had two groups of participants (15 female office workers each) using different menu systems: one group had a fixed-menu system to work with. The other group could modify and combine menu commands and menus into customized menus or input masks. On completion of the experiment, Ulich found that there were no significant differences between the groups regarding time to complete the task, errors made or outcome of performance (i.e., meeting the particular task goals). In interviewing the individuals after the experiment, however, and asking them for suggestions regarding any improvements they could think of regarding future experiments, he found that there *was* a difference between the groups regarding their mental models of their task: the fixed-menu group had over five times as many suggestions regarding improvements in the menu system as did the flexible-menu group. The flexible-menu group, however, had almost twice as many suggestions regarding ways to improve the work process and problem-solving strategies. In essence, the more flexible dialog situation led to more innovative job performance. One could say that the flexible-menu individuals, once freed from the syntactic aspects of the task domain, were able to develop more semantic levels of behaviorally-relevant cognition.

Are our assumptions concerning interface design impact upon ease of user learning correct for CBT purposes? One common assumption of GUI design is that simplification of interface control (i.e., making it easier to use) results in improved user-learning of a technical system. This has long been a basic tenet of direct manipulation advocates (see for example, Shneiderman 1987). A study by Golightly (1997) indicates that such an assumption may not always be correct. Figure 10 helps to explain this.[7]

Figure 10 illustrates two UIs representing different interface manipulation paradigms regarding the solving of the '8-puzzle' (see Golightly 1997). The 8-puzzle consists of the large 3×3 matrixed tiles on the left of Figure 10. (The bar on the far-left is a time thermometer showing the amount of time remaining for a trial.) The task goal for the user is to arrange the numbered tiles so that the pattern depicted in the smaller 3×3 matrix on the right is duplicated. Direct manipulation (DM) is effected by allowing the user to press, via mouse-control, a matrix element/tile adjacent to the empty cell. For example, in order to move the number 5 to the empty cell, one would merely have to mouse-click on the 5-tile. This would result in that tile being shifted to the right – thus, filling the previously vacant cell. The cell position where the 5-tile was originally present would then be vacant. In DM trials, the middle set of buttons is absent. The goal of the game is to complete as many pattern goals as possible within a set time limit; after a pattern has been duplicated, all tiles are 'shuffled' and the user starts matching again. Indirect manipulation (IM) is effected via the middle two rows of buttons: in the IM situation, the user is required to activate the button containing the number of the matrix tile to be moved. Thus, to move tile 5, the user would have to mouse-click on button 5 of the 2-columned group of buttons.

Golightly reports that, regarding overall success, the two groups are equal in

© Blackwell Publishers Ltd 1998

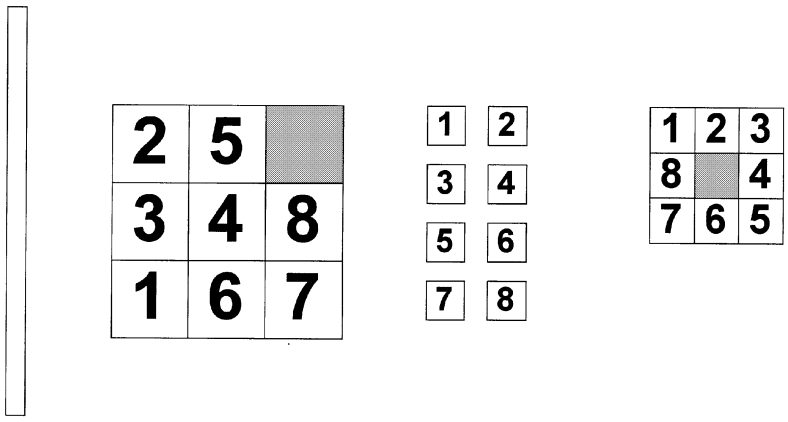

Figure 10: The 8-puzzle GUI. Adapted from Golightly (1997).

performance. Regarding problem-solving of the task, however, the two groups differ. As Golightly states, the two interfaces result in different problem-solving processes: the DM users are quicker in moving tiles but require more moves to achieve their goals. The IM users are forced to perform more poorly (regarding tile movement speed) but, having to invest a higher degree of cognitive effort (via the necessity of their having to 'work through' the button columns in order to effect tile movement) employ different problem-solving strategies than DM users. Golightly claims that the IM situation requires a 'richer problem-solving style' involving tile movement. In this respect, movement times for the DM group have a more constant rate, thus, suggesting that DM users employ a more 'trial-and-error' problem-solving style (*ibid.*). Golightly reports that his findings show that, under certain circumstances, use of a less direct interface can result in a user's developing a 'richer understanding of the [task] domain.' As he further writes, the results of his study would not support any claim that the IM paradigm is superior to that of DM. Indeed, and as he states, the DM paradigm offers the most holistic UI model for both novices and experts regarding a wide spectrum of applications. In some instances, however, the use of IM as an educational-support strategy in order to strengthen (Golightly 1997) 'enhanced domain knowledge' in such areas as situations involving safety critical systems, and use of educational software may be a viable learning enhancement strategy.

Cognitive/perceptual styles and UI use. It will be recalled that Ulich (1987) found that allowing users to 'design' their own user interfaces resulted in their attaining a deeper, semantic understanding of the problem-solving task at hand than was the case with users who were required to use a fixed-menu system. Ulich also found that problem-solving styles can result in differential outcomes regarding UI use. In another experiment of the series, Ulich devised a simulation game (PRIMP-1; Programmable Robot for the Investigation of Mental Processes) which enabled users to direct a robot in a labyrinth in which wine bottles were stored in three chambers. The bottles had three states: full, half-full and empty. The task of the game was to sort the bottles in shelves. Standard commands of the program were GO, TURN, DEPOSIT, TAKE, OPEN (see Ulich p. 90). The program allowed the construction of macros (consisting of groupings or aggregations of such simple commands). The goal was to program a robot so that the bottle-sorting could be accomplished in the shortest time with the fewest commands. Subjects also had to optimize the programs of other subjects according to the same criteria: few commands, minimum runtime. The results of the study showed that the most efficient programs were those which had been 'crafted' by their owners, i.e., programs of others required a (Ullich) 'retransformation processing' which involved a subject's having to try to understand the mental model of the other programmer's thinking regarding 'functionality, strategy and concept-building.' Additionally, subjects who were *'position-oriented,'* employed simpler program commands and tended not to build macros.

Alonso and Norman (1996) have studied the impact of UI design aspects upon the differential ability of users navigating through a database hierarchy. In their research they have focused upon users differing in Spatial Visualization Ability (SVA). They define individuals being high in SVA as persons who can, for example, effectively solve paper-folding tasks, and who are able to generate, retain and transform mental representations of information holistically

© Blackwell Publishers Ltd 1998

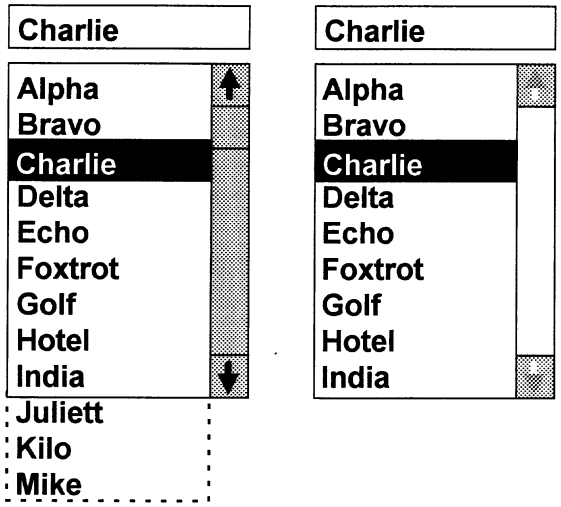

Figure 11: Two dialog lists with differing amounts of information.

(Alonso and Norman (1996) in reference to Lohman (1989). High SVA individuals, for example, have no problem with navigating through large and complex and hyperlinked databases such as those found on the Internet (i.e., the Web: WWW). Subjects low in SVA, however, are much less adept at working with such navigation and often get 'lost' in their attempts at traversing such data.

CBT user Personality and UI topology interactions. Horn (in preparation) in discussing personality assessment issues and paradigms in CBT has written that research has shown that testee impulsiveness can negatively impact upon intellectual assessment as measured with standardized intelligence instruments (Gibson, 1975, as referenced by Horn). In this regard, Figure 11 helps to depict a hypothetical CBT situation where use of a standard industry GUI dialog element in CBT could lead to biased testing results.

Mention has already made of industry-standard GUI designs which have become topological for CBT purposes. Figure 11 depicts two dialog lists with differing amounts of information. The selection list on the left contains nine visible entries; three further entries are 'hidden'. In order to view those entries, the user would have to activate the scroll bar on the right of the list. By moving the slider downward, entries 10–12 (Juliet, Kilo, Mike) would be scrolled upward and become visible.

The list on the right doesn't contain a slider-button, for the data fit the windowed listing perfectly: there are nine list entries for a window that has room for exactly nine items. The benefits of using such windowed listings, of course, lies in the fact that a GUI (as has already been mentioned) has limited real estate. As such, lists can contain hundreds of entries if needed with their accessing being effected via scrolling. Regardless of the amount of information contained in such listings, the window lists maintain a constant size. Thus, their placement in a GUI design can be standardized. Let us say, however, that we are dealing with a test candidate who is under much stress in the assessment situation, is running out of time, and has an impulsive personality. Let us further assume that he is viewing the list on the left. The only entries he sees are Alpha through India. Would it be possible, with all the pressure of the assessment situation, that he might forget to scroll the list to see what other entries are available for viewing? Might such an industry standard (or, surface topology) be potentially biasing for particular sub-populations of test candidates?

Summary

Regarding issues discussed here, there are many lessons for us to learn. First, it should be now clear that the role of the UI can not in any way be considered peripheral to the CBT process, for the UI can potentially play a significant role in every major phase of CBT-based assessment. Additionally, international ISO/EN/EC normative directives (which could only be very briefly discussed here regarding their impact upon CBT) hold particular significance, as they are legally binding for all organizational forms of industry, government, private practice – in brief: *anywhere* where PC technology is being used. As such, institutional, organizational and private-industry CBT practitioners and policy-makers are legally required to systematically evaluate their HW/SW systems regarding CBT ergonomics and user-friendless. In this regard, and on the basis of the previous discussion, it should be evident that only the professionally-trained specialist in user interface design issues – in unison with psychometric experts – is in a position to afford the expertise needed for advising the CBT policy-maker and/or practitioner regarding EN/EC directive conformance of a SW system (here, for CBT, see Booth 1988b).

As has been discussed, a particular problematic of CBT, regarding ergonomic directives, lies in what will be termed here as interpretational band-width: in many instances, there will be inevitable conflicts between ergonomic directive specifications and psychometric mandates respective to the attaining of reliable and valid psychometric measurement. The maintaining of a delicate balance between meeting the 'spirit of the law' of such normative directives – while meeting one's goals of psychometric

© Blackwell Publishers Ltd 1998

reliability and validity will, undoubtedly, constitute a central challenge for the CBT professional. It would behoove professional psychological organs such as the APA, BPS and BDP to offer the professional psychologist assistance in meeting such a challenge. (It also might be reflective of the current state-of-the-art regarding communication between international ergonomic norming agencies and psychological authorities and professional organs that, of the cited literature in the official EN ISO 92341-10 directive, the majority was over ten years old at the time of the directive's publishing. Articles contained in that literature listing are included, separately, in the following reference section.)

Regarding the attractiveness of such technology for psychometric practice, we have seen in the past that CBT hasn't always delivered on what it has promised. For example, adaptive testing (AT), although often cited as being an ideal candidate for CBT, has been shown to be particularly susceptible to what Kubinger (1989) has termed 'technically-linked presentation modality' effects; such modality effects center largely upon issues of the UI (see Booth 1991b). Nevertheless, improvements in PC technology, and evolving uses of it do seem to portend a wealth of newer psychometric applications and, in this regard, one could ask what the future holds for CBT. It is obvious that, in this respect, exiting and challenging days are ahead of us. Bartram, for example, in his discussion of 'distance asessment' (1997) points out that the 'logistics of assessment' will be radically changed by Internet technology in that '. . . an applicant for a job may reside and be assessed in Country A for a job which is in Country B, by assessors who work in Country C, for an organization which is based in Country D, using assessment procedures developed in Country E!'[8] Indeed, and as he further points out, professional associations and publishers on the national level will have to cooperate on an international basis. Globalization of industry is a fact. The same may hold for computer-based testing. Regardless of the platform topology, however, well-designed user interfaces will be a central aspect of their eventual success.

Notes

1 All translations from German are by the author.
2 The directive is binding for Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.
3 The use of square brackets is used here to denote GUI functions which are activated via buttons.
4 Personal communication; Claas Fock, Project Director, DELTA. Permission to use figs. 2 and 3 gratefully acknowledged.
5 Although the EN ISO 9241-10 directive lists application examples for each basic section, they are not individually numbered. Accordingly, and for purposes of this discussion, the number of an application example here pertains to its actual list position in the referenced directive section listing.
6 Personal communication.
7 Permission for the author to use a modified version of the original is gratefully acknowledged.
8 Personal communication

References

Alonso, D.L. and Norman, K.L. (1996) *Apparency of contingencies in single panel menus*. College Park, MD: University of Maryland, Department of Psychology. Document CAR-TR-824, May 1996.

Altmann, A. (1987) Direkte Manipulation: Empirische Befunde zum Einfluß der Benutzeroberfläche auf die Erlernbarkeit von Textsystemen. [Direct manipulation: Empirical findings regarding the impact of the user interface upon learnability of text systems.] In: Frese, M. (ed.). Themenheft: Mensch-Computer-Interaktion. *Zeitschrift für Arbeits- und Organisationspsychologie*, **36**(3), 127–136.

Ashworth, C.A. (1996) *GUI Users have trouble using graphic conventions on novel tasks*. URL http://www.acm.ord/sigs/sigchi/chi96/proceedings/int-post/Ashworth/Ashworth.html.

Bartram, Dave (1997) Distance Assessment: Psychological assessment through the internet. *Proceedings of the British Psychological Society*, Occupational Psychology Conference, Blackpool.

BfA (1995) DELTA. *Mehrplatzsystem zur Unterstützung der psychologischen Diagnostik und Beratung im Arbeitsamt*. [Multiple-station system for psychological diagnosis support and counseling at the Department of Employment.] Nürnberg, Bundesanstalt für Arbeit.

Booth, J.F. (1983) Pre-diagnostic clinical screening on the basis of hand-held computer-technology. *Mitteilungen und Nachrichten*, **100/111**, 136–148. Frankfurt, Deutsches Institut für Internationale Pädagogische Forschung.

Booth, J.F. (1984) Diagnostische Urteilsbildung: Grundlagen und angewandte klinische Forschung, Training und Praxis unter Anwendung des programmierbaren Taschenrechners. [Diagnostic decision-making: Basic and applied clinical research, training and practice using programmable hand-held computers.] In: Jäger, R.S. *et al*. (ed.). *Diagnostische Urteilsbildung in der Psychologie: Grundlagen und Anwendung* (203–276). Göttingen, Hogrefe.

Booth, J.F. (1986) Anwendungs-möglichkeiten für Personal-Computer in der eignungsdiagnostischen Praxis. [Personal computer applications in selection and placement.] *Psychologie und Praxis: Zeitschrift für Arbeits und Organisationspsychologie*, **30**, 98–104.

Booth, J.F. (1988a) Computerdiagnostik. [Computer-based diagnosis.] In Jäger. R.S. (ed.), *Psychologische Diagnostik. Ein Lehrbuch* (145–152). München, Psychologie Verlags Union.

© Blackwell Publishers Ltd 1998

Booth, J.F. (1988b) Kompetenz. [Competency.] in Jäger, R.S. (ed.), *Psychologische Diagnostik. Ein Lehrbuch* (104–113). München, Psychologie Verlags Union.

Booth, J.F. (1990) Computerdiagnostik: Errungenschaften und Möglichkeiten der Technologie. [Computer-based diagnosis: Achievements and potential of the technology.] *Diagnostica*, **36**, 180–203.

Booth, J.F. (1991a) Die Anwenderschnittstelle – Schlüssel zum anwenderfreundlichen und validen computergestützten Testen. [The User Interface: Key to user-friendly and valid computer-based testing.] In Schuler, H. and. Funke, U. (eds.), *Eignungsdiagnostik in Forschung und Praxis. Psychologische Information für Auswahl, Beratung und Förderung von Mitarbeitern*. Reihe: Beiträge zur Organisationspsychologie, Band **10**, 71–77. Stuttgart, Verlag für Angewandte Psychologie.

Booth, J.F. (1991b) The key to valid computer-based testing: The User Interface. *Revue Européenne de Psychologie Appliquée [European Review of Applied Psychology]*, **41**(4), 281–293.

Booth, J.F. (1992) Computerdiagnostik. [Computer-based diagnosis.] In Jäger, R.S. (ed.), *Psychologische Diagnostik. Ein Lehrbuch*, **2**, veränd. Auflage (186–197). München, Psychologie Verlags Union.

Booth, J.F. (1997a) User Interface issues in computer-based testing: Three research perspectives. Invited presentation, *International Test Commission Symposium*, Fifth European Congress of Psychology (6–11 July, 1997), Dublin, Ireland.

Booth, J.F. (1997b) *Computer-based psychological assessment and emerging technological trends: Paradigms, problematics, perspectives*. Invited presentation, NATO Partnership for Peace Workshop – Psychological Readiness for Multinational Operations: Directions for the 21st Century (7–9 July, 1997, Heidelberg, Germany). Hosted by the US Army Medical Research Unit-Europe and Walter Reed Army Institute of Research.

Daniels, J. (1962) *Figure Reasoning Test*. Nottingham, The Russell Press.

De Kleer, J. and Brown, J.S. (1983) Assumptions and ambiguities in mechanistic mental models. In Gentner, D. and Stevens, A.L. (eds.), *Mental models*, 155–190, Hillsdale, NJ, Erlbaum.

DIN (1988) *DIN 66 234 Teil 8. Bildschirmarbeitsplätze. Grundsätze ergonomischer Diagloggestaltung*. [*DIN 66 234 Part 8. Visual display unit working stations. Fundamentals of ergonomic dialog design*.] Deutsches Institut für Normung e. V. Berlin, Beuth.

DIN (1996a) *DIN EN ISO 9241–10. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 10: Grundsätze der Dialoggestaltung. (Mai.)* [*DIN EN ISO 9241–10. Ergonomic requirements for office work with visual display terminals*.] Deutsches Institut für Normung e. V. Berlin, Beuth.

DIN (1996b) *DIN EN ISO 9241–10. Ergonomic requirements for office work with visual display terminals. Dialog principles*. Deutsches Institut für Normung e. V. Berlin, Beuth. July.

Dutke, S. (1987) Lernen an einem textkommunikationsgerät: Wissenserwerb und Handlungsfehler. [Learning with a text-communications device: Knowledge acquisition and operational errors.] In: Frese, M. (ed.). Themenheft, Mensch-Computer-Interaktion. *Zeitschrift für Arbeits- und Organisationspsychologie*, **31**, 100–107.

Dutke, S. (1994) *Mentale Modelle: Konstrukte des Wissens und Verstehens. Kognitionspsychogische Grundlagen für die Software-Ergonomie*. [Mental models: Constructs of knowledge and understanding. Cognitive psychology foundations for software ergonomics.] Stuttgart, Verlag für Angewandte Psychologie.

Dzida, W. (1988) Modellierung und Bewertung von Benutzerschittstellen. [Modelling and evaluation of user interfaces.] *Software Kurierer*, **1**, 13–27.

EVADIS (1997) *Der ISO 9241-Evaluator*. [The ISO 9241 Evaluator.] URL http://zeus-gmd.de/ie/mmk/evadis/gruenes_blatt/gruenes_blatt.html. April, 1997.

Ford, J.K. and Kraiger, K. (1995) The application of cognitive constructs and principles to the instructional systems model of training: Implications for needs assessment, design, and transfer. In: Cooper, C.L. and Robertson, I.T. (eds.), *International Review of Industrial and Organizational Psychology*. Volume **10**. New York, John Wiley & Sons, 1–48.

FRT (1997) *The Figure Reasoning Test – Form PC-B*. Test manual. Heidelberg, John Booth Consulting.

Gentner, D.R. and Grudin, J. (1996) Design models for computer-human interfaces. *Computer*, **29**(6), 28–35.

Gibson, H.B. (1975) Relations between performance on the Advanced Progressive Matrices and the EPI in high-intelligence subjects. *British Journal of Social and Clinical Psychology*, **14**, 363–369.

Golightly, D. (1997) *Harnessing the interface for domain learning*. URL http://www.acm.org/sigs/sigchi/chi96/proceedings/doctoral/Golightly/dag_txt.h-tm (Accessed: 17 Dec 97).

Hamborg, K.-C. (1996) Zum Einfluß der Komplexität von Software-Systemen auf Fehler bei Computernovizen und Experten. [On the influence of complexity of software systems upon errors of computer novices and experts.] *Zeitschrift für Arbeits- und Organisationspachologie*, **40**, 3–11.

Hayes, F. and Baran, N. (1989) A guide to GUIs. *Byte*, **7**, 250–267.

Heineken, E. and Kuckhoff, A. (1992) Kognitive Kompatibilität und mnestische Stabilität von Menüstrukturen. [Cognitive compatibility and mnestic stability of menu structures.] *Zeitschrift für Arbeits- und Organisationspsychologie*, **36**(3), 127–136.

Horn, R. (in preparation) Computerized tests: An alternative for the assessment of personality traits.

Jorna, R.A. (1988) Comparison of presentation and representation: Linguistic and pictorial. In van der Veer, G.C. and Mulder, G. (eds.), *Human-computer interaction: Psychonomic Aspects*. Berlin, Springer-Verlag, 172–185.

Kandogan, E. and Shneiderman, B. (1997) Elastic Windows: Evaluation of multi-window operations. In *CHI 97 Proceedings, Atlanta GA*, 22–27 March 1997, ACM New York, 250–257.

Kubinger, K.D. (1989) Aktueller Stand und kritische Würdigung der Probabilistischen Testheorie. [Present status and critical evaluation of

© Blackwell Publishers Ltd 1998

probabilistic test theory.] In Kubinger, K.D. (ed.). *Moderne testtheorie*. Weinheim, Beltz.

Lohman, D.F. (1989) Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, **59**, 333–373.

Neubauer, A. and Urban, E. (1991) Der Vergleich von computergestützter Testdarbeitung und Standardvorgabe am Beispiel von Ravens Advanced Progressive Matrices. [A comparison between computer-aided test presentation and standard test presentation using Ravens Advanced Progresive Matrices.] In Schuler, H. and Funke, U. (eds.), *Eignungsdiagnostik in Forschung und Praxis. Psychologische Information für Auswahl, Beratung und Förderung von Mitarbeitern*. Reihe, Beiträge zur Organisationspsychologie, Band **10**, 81–84. Stuttgart, Verlag für Angewandte Psychologie.

Norman, K.L. (1991) *The psychology of menu selection: Designing cognitive control at the Human/Computer Interface*. Norwood, NJ, Ablex Publishing Corporation.

Rauterberg, M. (1992) An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. *Behaviour & Information Technology*, **11**, 227–236.

Rauterberg, M. (1993) AMME: an Automatic Mental Model Evaluation to analyse user behavior traced in a finite, discrete state space. *Ergonomics*, **36**(11), 1369–1380.

Rauterberg, M. (1996a) How to measure cognitive complexity in human-computer interaction. In Trappl, R. (ed.), *Cybernetics and Systems '96*, **2**. Vienna: Austrian Society for Cybernetic Studies, 815–820, ISBN 3 85206 133 4.

Rauterberg, M. (1996b) How to measure and to quantify usability attributes of man-machine interfaces. In: Tanie, K. (ed.), *Proceedings 5th IEEE International Workshop on Robot and Human Communication*, 11–14 November, Tsukuba Japan. IEEE Catalog Number: 96TH8179, ISBN 0-7803-3253-9, 262-267.

Raven, J.C. (1958) *Advanced Progressive Matrices*. London, Lewis.

SANUS (1996) *Richtlinie des Rates vom 29. Mai 1990 über die Mindestvorschriften bezüglich der Sicherheit und des gesundheitsschutzes bei der Arbeit an Bildschirmgeräten.* [*Guidelines of the commission of 29 May 1990 regarding minimal requirements concerning security and health protection in working with visual display terminals.*] URL http://sanus.uni-wuppertal.de/sanus/was_ist/90_290.html. Last modified 2 Sep 1996, Sassmannshausen, A. and Luter, B.

Seel, N. (1986) *Wissenserwerb durch Medien und 'mentale Modelle.'* [*Knowledge acquisition via media and 'mental models.'*] Unterrichtswissenschaft, **14**, 384–401.

Seymour, J. (1996) Why you don't use windows. *PC Magazine*, **10**, 91–92.

Shapiro, E. (1995) Needed: A GUI revolution. *Byte*, **2**, 238.

Shneiderman, B. (1987) *Designing the user interface. Strategies for effective human-computer interaction*. Reading, Massachusetts, Addison-Wesley.

Ulich, E. (1987) Zur Frage der Individualisierung von Arbeitstätigkeiten unter besonderer Berücksichtigung der Mensch-Computer-Interaktion. [On the question of individualization of work activities with particular consideration of man-machine interactions.] In Frese, M. (ed.), Themenheft: Mensch-Computer-Interaktion. *Zeitschrift für Arbeits- und Organisationspsychologie*, **31**, 86–93.

van der Veer, G.C. and Felt, M.A.M. (1988) Development of mental models of an office system: A field study on an introductory course. In van der Veer, G.C. and Mulder, G. (eds.), *Human-computer interaction: Psychonomic Aspects*. Berlin, Springer-Verlag, 250–272.

Zwingmann, B. (1996) *Statement. Arbeitsschutz an Bildschirmarbeitsplätzen: Die Bedeutung der EU-Richtlinien aus Arbeitnehmersicht. [Statement. Work safety with video display terminals: The significance of EC guidelines from the perspective of the employee.]* URL http://sanus.uni-wuppertal.de/de/sanus/zeitung/Z4/statemen.html. Last modified July 19, 1996 Th. Schnitger.

Literature cited in the EN ISO 9241-10 directive

AFNOR Z 67-110 (1988) *Information Processing, Ergonomics and Man-Computer Dialogue, Part 1: Concepts*. January.

Arnold, A.G. and Roe, R.A. (1989) Action facilitation: a theoretical concept and its use in user interface design. In *Work with Computers: Organizational, Management, Stress and Health Aspects*. Smith, M.J. and Salvendy, G. (eds.), Elsevier, Amsterdam.

Cakir, L.A. (1986) Towards an ergonomc design of software. *Behav. Inf. Technol.*, **5**, 63–70.

DIN 66 234-8, *VDU work stations – Part 8: Principles of ergonnomic dialogue design*.

Dzida, W., Herda, S. and Itzfeld, W. (1978) *User-perceived quality of interactive systems*. IEEE Trans. Softw. Eng., SE4.

Hacker, W. (1986) *Allgemeine Arbeits- und Ingeneiurpsychologie, Psychische Struktur und Regulation von Arbeitstatigkeiten*. Verlag Hans Huber, Bern.

Shneidermann, B. (1987) *Designing the User interface: Strategies for effective human-computer interaction*. Addison-Wesley.

Sperandio, L.J. (1987) Software ergonomics of interface design. *Behav. Inf. Technol.*, **6**, 271–278.

© Blackwell Publishers Ltd 1998