User analysis in HCl—the historical lessons from individual differences research

Andrew Dillon

School of Library and Information Science, Indiana University, Bloomington, Indiana, 47405, USA. email: adillon@indiana.edu.

Charles Watson

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana, 47405, USA

User analysis is a crucial aspect of user-centered systems design, yet Human–Computer Interaction (HCI) has yet to formulate reliable and valid characterizations of users beyond gross distinctions based on task and experience. Individual differences research from mainstream psychology has identified a stable set of characteristics that would appear to offer potential application in the HCI arena. Furthermore, in its evolution over the last 100 years, research on individual differences has faced many of the problems of theoretical status and applicability that are common to HCI. In the present paper, the relationship between work in cognitive and differential psychology and current analyses of users in HCI is examined. It is concluded that HCI could gain significant predictive power if individual differences research was related to the analysis of users in contemporary systems design. ©1996 Academic Press Limited

1. Introduction

The current emphasis on user-centered design for interactive technologies (see, e.g. Norman & Draper, 1986; Karat, 1991) places great emphasis on understanding the user in attempting to develop more usable artifacts. To this end, design teams are urged to perform user and task analyses at the earliest stages of product development and to consider the nature of the users' cognitive and physical predispositions and abilities. These use characteristics are correctly seen as important in constraining the available design options and, if attended to, in increasing the likelihood of producing a usable application.

In current situations user analysis often means distinguishing users broadly in terms of expertise with technology, task experience, educational background, linguistic ability, gender and age. Nielsen (1993) for example, proposes a three-dimensional analysis of users that draws distinctions in terms of domain knowledge, computing experience, and application experience. Booth (1989) offers a more detailed list of user characteristics that clusters user variance into five broad dimensions: user data, job characteristics, background, usage constraints, and personal traits.

Of necessity, such user analyses are highly context sensitive and offer little potential for generalization, never mind agreement across proponents. Design

teams must repeat such data gathering for every new design process and there has been little progress over the last 10 years in characterizing users into groups that offer predictive power in terms of likely response to new technologies. Of course, one can be sure that there are differences between experienced and inexperienced users, or that computing expertise is likely to make some users more tolerant of, or efficient with, a given system than others, but the HCI community is less able to indicate in advance how user differences relate to interface design, or what general form of system might best suit particular user groups beyond guidelines such as "provide menus for novices". This would seem to be an important avenue for research in HCI.

In the present paper we will attempt to demonstrate that user analysis in HCI could gain substantive insight from examining the history of work on the psychology of individual differences. This work, ongoing for almost a century, has evolved into a form that appears to be applicable to practical design issues and offers the HCI community an opportunity to develop a framework within which generalizable findings regarding users could inform both systems design and user training at the earliest stages.

These issues are increasingly being raised in the design of interactive computing systems. Egan (1988) reported differences between users of the order of 20:1 for common computing tasks such as programming and text editing (a far greater range than usually found in human factors work). Furthermore, he pointed out that such differences could be understood and predicted as well as modified through design. The ever-widening user population resulting from the diffusion of technology means it is no longer enough to base ergonomic inputs on generic models of the users, or to assume training on computerized tasks can be standardized across user populations. Here the differential approach must be married with the traditional physical and cognitive ergonomic approaches if worthwhile influence is to result.

The potential payoff for the human factors profession is enormous. For example, as Overmier, Montague and Jenkins (1989) point out: the US Navy teaches over 7000 courses per annum to about 900 000 students, a significant proportion of these computer-based. With failure rates for unselected samples often reaching 50%, any theory-based knowledge of the influence, in this case, of computer-based instruction type on individuals of varying abilities, becomes critically important.

The study of individual differences is as old as psychology itself, and one may wonder how it has remained so marginal to mainstream HCI which is usually receptive to psychological theory. It is clear that the domain of differential or correlational psychology has largely evolved separately from the experimental tradition so much so that writers as early as Boring (1929) and as recent as Jenkins (1989) refer regrettably to the two approaches as distinct disciplines within psychology, each concerned with different issues and employing different methods, often with little interaction between them. Most noticeable is the difference in what Thorndike (1954) termed the "laboratory values" of these two approaches. The differential approach is characterized by large sample sizes and the rigorous application of multivariate or factor-analytic techniques in the search for identifiable patterns of differences within the samples. For the differentialist, variations from the mean are thought to reflect latent mental structures or "factors" required to

perform a task. The experimentalist, however, is less concerned with sample sizes and typically assumes relative homogeneity among subjects of whatever ability is required to perform a task, often relegating inter-subject differences into the category of error variance. Where individual differences are of interest, for contemporary experimentalists at least, process (how a psychological event occurs) rather than structure (what psychological factors are employed) in task performance is its most important determinant.

The two approaches can be bridged. Adams (1989) argues that the field of learning has always attempted to draw on both, if only as a result of the continuous debates about the relationship between intelligence and learning (see, e.g. Gagne, 1967). As will be shown, some investigators are beginning to take an interest in individual differences in human information processing and are applying differential concepts to experimental investigations (e.g. Dillon & Schmeck, 1983; Sternberg, 1985; Kyllonen, 1993). To date, the HCI community, in drawing on psychology, has adopted the experimental rather than the differential tradition. The present paper will argue that HCI could benefit greatly from attending to differential work.

This review examines the dominant themes of current differential psychology and discusses efforts to apply an experimental approach to the study of individual differences in cognitive abilities. It will then look at attempted and potential applications of this work to problems of personnel and user selection. On the basis of these reviews it proposes analysing individual differences in cognitive terms so that both user analysis and computer systems design may be better supported by human factors psychologists. In part, this review will take the reader outside of mainstream HCI concerns but in so doing it will seek to demonstrate how such a historical perspective gained from theoretical psychology can have relevance to HCI problems.

2. Differential psychology

In very general terms the modern history of psychological studies of individual differences began with an attempt to identify a general factor (such as intelligence), which gave way to a search for multiple factors (or mental abilities), which in turn waned in favor of contemporary attempts to reduce the exploding number of multiple-factor theories to a more manageable set of core abilities. (In the 18th century, the concept of ''faculty'' developed in psychology, and phrenology had anticipated this search, but this review will be limited to the era of ''scientific'' psychology.)

Initial interest focused on mental tests, in laboratories, such as Wundt's in Leipzig in the late 1800s. Here, researchers were interested in differences between individuals in terms of learning ability and sought to associate these differences with such indirect measures as reaction time, finger tapping speed, tactile sensitivity and keenness of vision. The goal was to identify lower-level sensory and motor capabilities that would explain differences in higher-level learning and other mental tasks.

It soon became clear that such measures alone were unsatisfactory (just as contemporary usability evaluations in HCI would be seen if they were based solely

on input device speed for example) and Ebbinghaus (1895) and Wissler (1901) concluded on the basis of empirical studies that the psychomotor and sensory tests of the time bore little or no relationship to learning ability of students (though more recent approaches emphasize the importance of processing speed, as measured by reaction time, in successful prediction of higher-level performance as will be discussed later). Similarly, Binet argued for measurement of more complex activities such as memory and comprehension and developed his influential test of intelligence (Binet & Simon, 1905) utilizing items that had a higher degree of fidelity to the tasks students would be required to perform in educational settings (a crucial insight and one which Dunnette (1976) argued has been largely ignored by more recent differential psychologists attempting to construct elaborate theories of individual differences—a point discussed later in terms of current HCI applicability).

Adopting a more strictly statistical approach to the measurement of intelligence, Spearman (1904) developed correlational tests to determine the relationships among a variety of variables, such as school grades and teacher ratings, and concluded that all intellectual activities share a single common factor (the general factor—"g") accompanied by a number of specific factors ("s") that influence an individual's ability on particular tasks. Correlations between performance on any two tasks were thus influenced by "g" and whatever "s" factors were important for successful performance on each of the tasks. The legacy of this distinction still runs through contemporary work in which specific factors have come to be named and groups of these have been linked (e.g. Carroll, 1993).

The work of Thurstone and his followers in the decades surrounding World War II (e.g. Thurstone, 1938) led to the proposal and empirical corroboration of approximately seven aptitudes, or primary mental abilities, that can characterize individuals, the most commonly cited and tested being: verbal comprehension; word fluency; arithmetic ability; spatial relations; memory span and duration; perceptual speed; and inductive reasoning. To some extent all of these factors, or derivatives of them, are still employed in individual differences work [e.g. see Tirre and Pena (1993) for work on verbal comprehension, Lohman and Kyllonen (1983) for work on spatial relations or Dempster (1981) on memory span]. The rejection of a unitary concept termed ''intelligence'' in favor of a more detailed account of a range of specific abilities or aptitudes has been virtually complete (though it is debatable that any one, even Binet, really assumed that a unitary property, even ''academic ability,'' existed as such).

After Thurstone, the classification of human abilities produced a bewildering array of (generally factor-analysis-based) aptitudes. The result, by the mid 1960s, was myriad dimensions or constructs purporting to distinguish people in terms of abilities, cognitive style, personality and so forth, with little apparent unity between theories and increasingly fine distinctions being drawn between constructs.

Probably the most ambitious theory of this period was Guilford's (1967) Structure of Intellect (SI) Model. He proposed and tested for over 30 years the theory that human cognitive abilities could be described on three dimensions (cubic form): operations (cognition, memory, divergent production, convergent production and evaluation); content (figural, symbolic, semantic and behavioral); and products (implications, transformations, systems, relations, classes and units). Any one individual can be described as having an aptitude or ability according to his success

or failure in dealing with tasks involving particular contents, operations and products e.g., a person competent in cognitive operations involving semantic products such as classes or relations might be said to have "high verbal ability".

This model posits the existence of up to 120 distinct aptitudes (patterns of strengths on these cognitive dimensions), and represents the culmination of the multiple-aptitude approach. By the 1970s Guilford and colleagues claimed the successful identification of 98 of the (by then postulated) 120 aptitudes (Guilford & Hoepfner, 1971), and this number had grown to 105 (by now out of 150) by the early 1980s (Guilford, 1982). Critics of the theory however were less impressed and numerous analyses of its empirical basis drew attention to the form of factor analysis used by Guilford and his team (e.g. Carroll, 1972). In particular, they were accused of overly-subjective identification of factors. Horn and Knapp (1973) demonstrated that using Guilford *et al.*'s method of factor identification it is quite possible to support any theory, even those generated by grouping variables at random. Haynes (1970) performed a re-analysis of a subset of Guilford's data and reported a general factor with loadings of 0.30 or higher on 28 of the 34 tests he had chosen to represent 17 of the supposedly most clearly established SI factors, suggesting that a more conservative interpretation of the data would significantly reduce the number of factors identified. One might see interesting parallels in the early literature on user categorization in HCI, where the distinctions between programmers and non-programmers, gave way to more elaborate job-based classifications before dissatisfaction with these reduced most discussions of user types to knowledge-based distinctions (Dillon, 1987).

In an attempt to lessen the confusion that resulted from the multitude of aptitudes that had been proposed, the Educational Testing Service (French, Ekstrom & Price, 1963) had assembled a "kit" of reference tests. This consisted of 24 aptitudes that were tested with two to five tests per aptitude. Summarizing the scene 10 years later, Dunnette (1976) argued that on the basis of more recent evidence, 10 of these aptitudes had stood the test: speed of closure, fluency, inductive reasoning, associative memory, memory span, number facility, perceptual speed, deductive reasoning, spatial orientation and verbal comprehension. He noted in conclusion how remarkable it was that "years of factorial research since Thurstone's seminal contribution have added only minor modifications to his list of Primary Mental Abilities" (Dunnette, 1976: p. 483). Current theorists (e.g. Sternberg, 1985) seem to agree that these represent relatively stable factors.

The 1970s and 1980s were a period of revision and meta-analysis within differential psychology. Researchers became increasingly sophisticated in the application of multivariate and factor analytic techniques (see, for example, Nesselroade & Cattell, 1988) and several re-analyses of the older data sets have been performed. Carroll (1993) reports results of a 10 year program of re-analysis that he undertook, stating that there are about 2500 data sets covering 10 000 variables in the literature of which he has re-analysed 463 sets. Such re-analysis, he claims, is needed not only to check published data but to provide a consistent basis for comparison (Carroll chose hierarchical exploratory factor analysis for his re-analysis).

Carroll's characterization of the field also suggests relatively little progress has been made over the years. Like others, he identifies the core concepts of Thurstone as key but contends that they have been accepted too uncritically "and used as basis

for repetitive, largely uninformative studies year after year, even up to the present'' (Carroll, 1989: p. 45). While there has been much critical debate on the use of factor-analytic techniques over the years, Sternberg (1985) sums up the arguments by saying that when attempting to isolate global, structural constellations of data then the techniques are sound but still subject to misuse. However, the argument goes, there has been a tendency when factor analysis is misused to blame the technique rather than the misusers.

The Carroll re-analysis groups factors into a three-level hierarchy that is undoubtedly the most representative summary of current thought. The top level is general intelligence (g). Under this, Carroll (1993) distinguishes eight ability categories:

- crystallized intelligence†
- fluid intelligence
- general memory and learning
- broad visual perception
- broad auditory perception
- Broad retrieval ability
- broad cognitive speed
- processing speed.

Each of these general ability types is further subdivided to provide specific first order factors of type usually found in any one investigation, e.g. memory is divided into associative memory, episodic memory, memory span and visual memory. Carroll admits that many other factors could be added (particularly at the first-order level) and the connections between factors at different levels are more complicated than his classification seems to indicate. However as a general synopsis of research to date this is probably the least contentious classification in the literature, and it has been highly praised by many reviewers as the likely standard conceptualization for the foreseeable future (see, e.g. Kline, 1994).

In summary, it appears that theorists differ primarily in terms of the number of posited factors and the organization of these into hierarchic, cubic, radial or linear form. The interpretive procedures used in ''factoring'' data appear largely responsible for these differences. All assume that abilities are determined by ''latent'' mental factors best revealed by identifying individual differences in test performance. As Carroll (1993) shows, many of the differences betwen theorists can be reduced by providing a consistent basis for comparison (i.e. a standard factor analytic technique and conservative interpretations of the resultant data). An obvious implication for HCI research is to avoid overly fine distinctions based on plausible task or knowledge based distinctions and to seek robust cognitive and psychophysiological constructs on which to categorize users. A useful starting point might be the extension of Carroll's hierarchy to the analysis of user performance with interactive systems.

---

† The distinction between crystallized and fluid intelligence dates back to Cattell (1943) who posited that crystallized intelligence reflected the extent to which an individual had profited from education, training and cultural influences, and fluid intelligence reflected basic abilities in reasoning and higher cognitive processes.

3. Cognitive science and individual differences

As with experimental psychologists, cognitive scientists originally tended to be more interested in general models of processes rather than in individual differences, maintaining the separation of approaches to psychology discussed earlier. However, during the last 20 years there has been growing interest in estimating the variability among people in terms of cognitive process, following the work of Hunt, in particular, in the mid 1970s (e.g. Hunt, Lunneborg & Lewis, 1975; Hunt, 1978) which tried to explain differences in verbal intelligence in information processing terms, specifically speed of access to lexical information in long-term memory (LTM). Hunt and his colleagues empirically demonstrated an interaction between treatment effects and certain characteristics of subjects as measured by aptitude tests and thereby raised the issue of how to investigate or control for inter-subject differences in experimental studies of cognition.

Since then, individual differences in information processing have been investigated for numerous constructs and processes: associative learning (Kyllonen, Tirre & Christal, 1991), name retrieval in verbal processing (Perfetti, Goldman & Hogaboan, 1978), visual information processing (Chiang & Atkinson, 1976), automatic vs. controlled processing (Ackerman & Schneider, 1985), and identification speed in memory span tasks (Dempster, 1981) to name but a few. The list of variables studies is ever-increasing and it would be pointless to list them here. Reviews of some of this work can be found in Lohman (1993), and its relevance to HCI is argued cogently in Egan (1988). Suffice it to say that for most components of information processing that have been subjected to differential investigations, individual differences have been observed.

The cognitive approach to individual differences postulates likely subsidiary processing stages in some "total" activity, devises appropriate experimental tasks to measure these and correlates overall task performance measures with the tests of ability. The major difference between the cognitive approach and earlier differential approaches to individual differences is the former's emphasis on dynamic rather than on structural differences between individuals. In many ways this is a natural extension of earlier factor-analytic work, identifying differences and attempting to account for them in information-processing rather than structural terms alone.

The cognitive approach can be classified in terms of levels of processing involved, which broadly run from concern with information processing speed on simple tasks to a concern with accuracy on complex problem-solving tasks. Sternberg (1985), for example, distinguishes between research on information-processing aspects of ability in terms of four speed measures employed: pure speed (e.g. simple reaction time) for which correlations are at best, low (e.g. Jensen, 1982); choice speed (e.g. choice reaction time) for which correlations with typical intelligence measures of about -0.30 have occasionally been shown (e.g. Nettelbeck, 1982); speed of lexical access (e.g. letter-comparison) which seems to reliably correlate approximately -0.30 with intelligence (e.g. Hunt *et al.,* 1975); and speed of reasoning processes (e.g. syllogistic reasoning) which Sternberg further breaks down into performance and executive processes in reasoning, both of which he claims on the basis of his own work (e.g.

Sternberg & Gardener, 1983) to correlate with intelligence at levels between -0.33 and -0.70 in various studies.

The speed-based classification proposed by Sternberg (1985) is not exhaustive as it offers little scope for the inclusion of work for which speed is of less or little concern [e.g. in the study of individual differences in learning and solving mathematical problems, Mayer (1982)]. In cases such as these, it is the breakdown of task performance and the classification of errors that can lead to the identification of likely processing components. Strength or weakness in these components' operations is more useful than speed in distinguishing individuals (even if the operation of these components may ultimately be distinguished in temporal terms). This is the approach adopted for studying differences in a variety of domains such as deductive reasoning (Johnson-Laird, 1985) or analogic problem solving (Whitely, 1980). Kyllonen's (1993) four-sources model seeks to explain variance in cognitive task performance in terms of working memory capacity, information processing speed, and the breadth of procedural and declarative knowledge. Such a model offers a potentially unifying perspective on individual differences in cognitive task performance and appears to be more applicable to HCI concerns.

4. Personality and cognitive style

Personality, defined loosely as traits or stable tendencies to respond to certian classes of stimuli or situations in predictable ways (see, for example, Cattell, 1965; Cronbach, 1984), has a long history of research and suffers many, if not more of the same problems that afflicted ability-differences research, namely the explosion of postulated personality models, factors or traits as they are normally termed. Cattell (1965), for example, proposed the existence of 16 distinct traits such as "cool–warm", "practical–imaginative", "shy–bold", "submissive–dominant", etc. which he derived through factor analysis of rater intercorrelations. The California Psychological Inventory (Gough, 1958) suggests there are 22 basic traits such as "dominance", "sociability" etc. Eysenck (1947) claims that only two orthogonal dimensions truly differentiate individuals: introversion–extraversion, and neuroticism–emotional stability. In later years, he added a third dimension dealing with psychoticism. More recent work (Eysenck, 1982) argues that these dimensions, coupled with work on heredity and psychophysiology, provide a potentially unifying paradigm for the study of personality.

Personality measures retain a degree of acceptability in personnel selection work and user analysis that is difficult to justify empirically although recent re-analyses in this field have led to the proposal of the "Big Five" personality factors: neuroticism, extraversion, openess to experience, agreeableness, and conscientiousness (Digman, 1990). A recent review of this work by Landy, Shankster and Kohler (1993) concluded that it is still too early to draw any reliable inferences on this characterization, though it hardly inspires confidence that Hough (1992) suggests that these five factors really ought to be extended to nine.

"Cognitive style" refers to relatively stable patterns of information processing that are displayed by an individual. In a sense it can be seen as the cognitive-psychological, or more accurately, information-processing equivalent of personality. Popular in the 1960s and 70s, this field again was dogged by the proposal of

innumerable style dimensions such as "holism–serialism (Pask, 1976), "field dependence–independence" (Witken, Moore, Goodenough & Cox, 1977) and "reflective–impulsive (Kagan, Rosman, Day, Albert & Phillips, 1964), which have intuitive appeal as representing the thought patterns of individuals but are hard to distinguish from many existing personality constructs and seem to add little to what can be predicted about performance solely on the basis of aptitudes (Vernon, 1972). Indeed, cognitive styles such as "visualizer–verbalizer" have proved of little use in predicting user performance with various user interfaces (see, e.g. Booth, Fowler & Macaulay, 1987).

However, before dismissing style (or personality) as a concept it is worth considering possible causes of this poor showing. These include the possibility that: (i) the dimensions identified thus far are superficial and need refining to determine true information processing differences; (ii) individuals are capable of manifesting several styles even presumably contradictory ones, depending on the circumstances; and (iii) specific styles may be highly correlated (positively or negatively) with specific tasks. To date though, there is little empirical support for any of these hypotheses, but they do suggest potentially fruitful lines of further research.

5. Psychomotor differences and skill acquisition

In parallel with much of the research on information processing, research on psychmotor performance was also addressing individual differences. Adams (1987) traces this research back to Thorndike's (1908) interest in the variance of performance as a function of practice. Thorndike's main finding was that variance did not decrease with training or practice, a claim that naturally enough caused a stir amongst educationalists. Subsequent work (e.g. Kincaid, 1925) cast doubt on this finding but the real surge of interest in motor performance occurred during the years leading up to World War II when the problem of pilot selection, a task with a considerable motor component, became acute.

Research up to this point seemed to indicate that motor skills were highly task specific (e.g. Seashore, 1928), representing "~s" rather than "~g" factors in Spearman's terms. However, as part of the selection drive during the war, military researchers worked on the development of what became known as the Complex Co-ordination Test Battery (see Adams, 1987) that was considered useful enough to be employed for pilot selection [a meaningful distinction thus being drawn between theoretical import and practical relevance, an issue of frequent concern in contemporary HCI work, see, e.g. Landauer (1995)].

Subsequent research on differential motor skills has been strongly influenced by Fleishman and colleagues (e.g. Fleishman & Hempel, 1954; Fleishman, 1972). Their usual approach involved extensive training on the Complex Co-ordination Test (or similar) as the criterion, with printed and motor tests used as a reference battery. Performance on training tasks at different levels of learning was usually correlated with test scores and the results were factor analysed to identify particular abilities in motor performance. This led to the proposal of several psychomotor aptitudes such as coordination, spatial relations etc., some of which were highly criterion-task specific. Despite criticism of some of their findings for methodological or analytical weakness (Adams, 1987), Fleishman and Hempel (1954) did clearly identify an

interesting pattern of skill development that continues to be studied. They showed that as motor skill was acquired, the learner progressed through a stage of predominantly perceptual or cognitive processing to one dominated by psychomotor ability i.e. proficiency in early learning phases is related most closely to factors such as perceptual speed of spatial awareness. With increased practice, skill is determined less by these and more by motor constraints.

Most recently, as witnessed in intelligence testing research, Ackerman (1987, 1988) has performed extensive re-analyses of some of Fleishman's data as well as carrying out new studies to examine these changes from an information processing perspective. On the basis of this work, he proposes that skill acquisition occurs through three stages, typically described as cognitive, associative and automatic (utilizing the model of Fitts & Posner, 1967), each of which is influenced by different abilities. For Ackerman (1987), stage 1 is reliant mostly on general intelligence or ability, stage 2 on perceptual speed and stage 3 on psychomotor ability. Individual differences in skill acquisition are therefore determined at different stages of learning by differences in one or other of these three types of ability.

Ackerman's work marks an interesting and fruitful marriage of information processing to differential psychology. In terms of the debate over variance in performance with practice, this model suggests that, given "task consistency" (i.e. low attentional demands and the prospects of automaticity occurring), general ability will cease to be a major determinant of performance as skill is acquired. If psychomotor differences are the final determinant of the level of asymptotic performance, then inter-person variance should diminish over time, assuming that such differences themselves are small. For inconsistent tasks (i.e. always requiring attentional resources), general ability-performance relationships will account for most of the variance in performance. Studies of individual-difference variance and the effect of practice or training need to be based therefore on a detailed understanding of the task being performed [precisely the argument Dunnette (1973) made earlier, and a core perspective in HCI practice, see, e.g. Shackel & Richardson (1991)].

6. Applications of differential psychology

According to both common sense and the assumptions of most employers, people vary, thus any group of workers performing virtually any type of task are not likely to be performing equally effectively at all times. This is presumably the founding principle of personnel selection, i.e. recruit and retain the best people for the job. However, one could extend this argument to include the design of the technology that an organization uses, since different systems will be usable to varying extents by different users, and this is a clear link to the application of this work in HCI.

Hull (1928) is generally credited with the first study of variance in worker performance. He calculated the best-to-worst ratios for a variety of workers in terms of their output and reported ratios in the range of 1.5:1 to 4:1. His general conclusion was that the best workers were typically twice as productive as the worst. Tiffin (1943) calculated the distribution (as opposed to Hull's reliance on the range) of output on a variety of jobs and found Hull's estimate to generally hold true. Interestingly, he also reported an increase in performance with experience (as

expected) but a decrease in the variance across the sample over time, a contrary finding to Thorndike's (1908) claim that individual differences did not diminish over time but which would now be consistent with the Ackerman–Dunnette theories of differential performance.

As with Carroll's recent attempts to categorize abilities through re-analysis, work on individual differences in worker performance has focused on re-analysis of earlier studies in the light of new or refined statistical procedures. Schmidt and Hunter (1983) summarized data from 17 published studies of performance variability and confirmed Hull's ratio. They define "best" as the 95th performance percentile and "worst" as the 5th, and report the best to be twice (however defined) the worst with little or no quality vs. rate trade-off.

Recent work has also attempted to quantify the value of good performers to an organization. This has traditionally been seen as difficult to estimate reliably ever since Brogden (1950) examined a formula for estimating the return on selection that was based on test validity and the standard deviation of productivity. This issue has been raised in contemporary HCI work since it is assumed that users' abilities to utilize new technology is positively correlated with job performance and some researchers seek to demonstrate the cost-benefits of well-designed technology to an organization (see, e.g. Chapanis, 1991).

Accountancy-based analyses (e.g. Roche, 1965) examined unit output per worker and related this to unit costs and profits, concluding that there would be a possible 3.7% profit increase in using appropriate selection tests in recruitment for the company concerned. Cronbach and Gleser (1965) criticized this finding as simplistic, failing to take account of type and value of component in calculating worker output (not all outputs were of equal value or could be produced as quickly) and thereby pooling rates inappropriately.

A series of studies by Schmidt and colleagues (e.g. Schmidt & Hunter, 1983) has produced what appears to be a more acceptable method, termed the Rational Estimate. They argued that the best judges of a worker's productivity were their immediate supervisors and elicited information from these on the performance of a variety of workers. Raters had to estimate in cash terms the productivity of poor, average and good workers (designated the 15th, 50th and 85th percentile performers). Pooling ratings from supervisors and assuming a normal distribution of abilities, these data yielded an estimate of the standard deviation of productivity in cash terms. Over several studies they found sufficiently consistent results to propose the rule that the standard deviation (S.D.) of workers' productivity is between 40–70% of salary. If the best are really considered to be two standard deviations better than the worst then this puts a very high price on poor selection. Schmidt and Hunter (1981) estimated that poor selection by the US Federal Government with 4 million employees, was then costing $16 billion per annum.

Critics of the Rational Estimate argue that it is inherently subjective, the raters are asked to put cash terms on worker productivity purely on the basis of their own judgement. Proponents counter that the approach merely asks raters to estimate productivity in terms of how much they would have to pay to an outside firm to acquire the service the worker provides, a not wholly satisfying response. But additional evidence supporting the Rational Estimate is emerging. Bobko and Karren (1982) for example, used the Rational Estimates approach to rate 92

insurance salesmen and deduced a S.D. (worker productivity) of $56 000. Actual sales figures were then used and these showed a S.D. of $52 000—a very close match to the estimated figure. Ledvinka and Simonet (1983) also provide supporting evidence, reporting that productivity in their sample could be put into objective cash terms consistent with the ''40–70% of salary'' rule. Such an approach has clear resonance with the concerns of the HCI community to justify their involvement in the design process. As well as performing the cost–benefit analyses advocated by Nielsen (1993) amongst others, an individual differences based approach of this kind could emphasize the financial value of designing the most appropriate interface for a given user population.

7. Use of tests in selection

Use of tests for selection began in military selection and placement in World War I, before spreading to commercial organizations in the 1920s (see, e.g. Freyd, 1923). Developments since then have been tied largely to the theoretical and methodological approaches of differential psychology described above, though Guion (1976) suggests that testing and selection procedures were considered so valid by the middle of the century that psychologists were no longer required to administer them, personnel managers were sufficient [an issue surely worthy of further examination by those within the human factors profession concerned with technology transfer, and ''giving HCI away,'' Diaper (1989)].

Doubts about testing both in terms of culture bias and predictive validity reached a head in the mid 1960s with the passing of the Civil Rights Act in America prohibiting the use of selection procedures that might be biased against minorities, and the publication by Ghiselli (1966) of a damning review of predictive validities of tests. In the case of bias, it is possible to devise culture-fair tests but given the wording of the law in the US, even a valid and fair test can be seen as being used to bias selection of candidates. This issue will not be discussed further here except to observe that it provided a climate in which certain findings were readily seized on and frequently distorted.

Ghiselli (1966) compared the results of published studies of the validity of ability tests. In effect he compiled mean validity coefficients for studies using either trainability or proficiency (performance at the actual task) as a criterion for three types of clerical jobs. This sample of data covered the use of 25 tests. He reported a range of validity coefficients from 0.01 (for interest and proficiency in recording clerks) to 0.58 (for cancellation/perceptual accuracy and training success in recording clerks), with a mean validity coefficient of 0.30 for all tests. This finding was a disappointment to advocates of testing in selection, suggesting as it does that tests only account for 9% of the variance in trainability or proficiency.

These results led some researchers to assume that the validity of tests varied as a result of situational variables in the work organization, resulting in the postulation of numerous moderator variables such as organizational climate, worker motivation, leadership style etc. which were presumed to ''moderate'' the extent to which test-based estimates of aptitude correlated with eventual job performance. As moderator variables failed to be reliably demonstrated (see, e.g. Schmidt, Hunter & Caplan, 1981), a more pessimistic view suggested that the prediction of validity for

any specific test and organization was too complex, therefore a local validation study would always be required in order to determine the most appropriate test for any given situation—the "situation-specificity hypothesis" (Ghiselli, 1973). Again, this emphasis on context specificity has obvious parallels with current emphases in usability testing (Dumas & Redish, 1993).

Schmidt and Hunter (1977) however argued that the problem with most meta-analytic studies is that they accept the variance across studies at face value and ignore the possible contamination effect of sampling error. Given the small sample sizes typical of most studies (mean of $n = 68$ according to a survey of 1500 studies by Lent, Aurbach & Levin, 1971), correlations can be very unstable. Furthermore, criterion reliability varies, i.e. if proficiency is estimated by supervisor ratings, then it is important to realize that the ratings themselves, though ecologically valid, have a reliability that varies around a mean of $r = 0.60$ (Cook, 1988). Other possible limitations in validity coefficients include: restricted range (scores may be limited by job restrictions on performance), test reliability, criterion contamination ("objective" raters of performance may actually know the subjects' test scores) and other experimenter errors (e.g. typographical or computational errors).

Schmidt and Hunter (1977) proposed a number of formulas to correct for several of these possible errors in validity generalization, which have been modified and adjusted slightly over the following years (see, e.g. Hunter, Schmidt & Pearlman, 1982). Hunter and Hunter (1984) used such formulae in a review involving 23 new meta-analyses of data on validity studies (including a re-analysis of Ghiselli's meta-analysis). In short, they found that true mean validity for ability tests is generally much higher than Ghiselli's estimate of 0.30, and is in fact 0.53. This is the figure they cite for general ability composite measures and the prediction of success on jobs for which employees will be recruited at entry level and trained after hiring. Where the predictor is used to make cases for promotion, and current performance is the criterion then they argue that prediction based on a work sample test may be equivalent ($r = 0.54$). Combining predictors (as is typical) is mathematically likely to increase total validity at most by the square of the validity of the less valid predictor. Even then, the various predictors must be weighted or else the combined predictors may in fact reduce the total validity that would be obtained using only the best single prediction.

One implication of all these analyses and re-analyses is that measures of ability can account for approximately 25% of variance in performance. They are not unduly limited by situation specificity and thus can be used for most selection applications with appropriate caution. If used in combination with other sources of information such as previous work experience, education, task knowledge, etc., they should add important data to the decision making process.

8. Implications of this work for HCI

The present review has ranged over a century of work in differential and experimental psychology in the search for clues this work may offer HCI researchers interested in better analysing users. From this perspective there appears now to be little need for further number-crunching exercises to search for factors underlying

general abilities. The practical implications of 100 years of differential psychology seem to the present authors to be the following.

(1) A core number of basic abilities have been reliably and validly identified, though their names and precise relationships to each other are still in question. They include the original list proposed by Thurstone, as modified recently by Carroll into his three-stratum model. For the foreseeable future we envisage this classification to remain stable.

(2) These cognitive abilities influence the performance of specific tasks in predictable ways.

(3) Understanding these abilities and differences can be fruitfully tackled within an information-processing perspective that emphasizes the changing role of various processing stages in skill acquisition and task type.

(4) The reliability and validity of user and task analyses for systems design could be improved on the basis of this work. As a result, user analysis could support the design of more cognitively compatible interfaces or lead to the development of more appropriate training packages for users.

(5) This theoretical analysis of user differences would support greater generalization of findings across HCI applications. This would enable research to build on a scientific base and advance user classifications from their current reliance on experience or job-based criteria.

The last three points emphasize that an understanding of what sets certain user types apart in cognitive terms can constrain the number of potential design solutions. Given the current reliance on broad measures such as task experience, technical skills, domain knowledge or even age in typical user analysis (e.g. Greene, Gomez & Devlin, 1986; Nielsen, 1993) there is a definite opportunity for more rigorous theoretical and data-driven approach here. While common sense variables are likely to remain important in establishing context, generalizability can only be obtained with a more consistent basis of comparison. Psychological measures of individual differences provide one such basis.

This work is only beginning to be addressed in the HCI community but it seems to be insightful (e.g. Vicente & Williges, 1988). Allen (1994) examined the relationship between two cognitive abilities (logical reasoning and perceptual speed), and user performance with an information retrieval system designed to present items in one of two different ways (rank-ordered or non-rank-ordered output). While he reported no effect for perceptual speed, logical reasoning (as tested by the Diagramming Relationships Test; Ekstrom *et al.,* 1976) interacted significantly with interface type, leading Allen to conclude that users with low logical reasoning ability would have an increased chance of identifying relevant information if they used a system designed to provide rank-ordered rather than traditional non-rank-ordered output. Extending this analysis, one could estimate the distribution of logical reasoning ability across a population and so devise appropriate interface design for specific contexts, e.g. public libraries, medical research centers, etc.

Sein, Olfman, Bostrom & Davis (1993) examined the abilities of users with either high or low visual ability to learn to use three different software applications (email, modeling software and operating systems). In so doing they manipulated training and interface design (command language and direct manipulation). Using a standard test for visual ability [the VZ2 paper folding test of Ekstrom *et al.* (1976)], they

found that "high visuals" indeed learned faster on all applications. Perhaps most interestingly though, from an HCI perspective, was their finding that with appropriate training and interface design, the gap between the two user groups could be reduced and even reversed. In particular, they emphasized the role of direct manipulation interfaces in increasing the users' ability to visualize the system's activities.

What is interesting from both these studies is the explicit mapping of individual differences to interface characteristics. Both show that even though there are differences amongst users that predict performance with interactive systems, appropriate design of the interface and/or training can reduce these differences. Sein *et al.* (1993) for example, showed that direct manipulation interfaces could reduce the requirement for the user to form a detailed internal representation of the system state during interaction and that such a lessening in requirement, aided low-visual-ability users to perform as well as high-visual-ability users.

Such an approach to the study of computer users is both theoretically possible and practically important. The prediction of 25% of variance on the basis of ability alone is useful and encouraging but should not be considered the end of the story. Much as the situation-specificity requirement of recruitment testing has been modified, with the identification of information processing distinctions among users, the complete reliance on in-site usability testing might be lessened. Linking specific physical, perceptual and cognitive abilities to specific tasks analysed in ergonomic terms would thus further extend our predictive abilities, after all, selection in industry rarely relies solely on test scores but combines these data with other sources of information (job history, experience, references, etc.) to build up a composite view of candidate employees. The present authors see the analysis of individual differences similarly as one means of gaining better understanding of the users of interactive technologies that would be combined with other relevance sources of information of the kind outlined by Booth (1989). The authors contend that psychological knowledge in this domain is approaching a level of sophistication that can have significant impact on real-world issues. The natural reticence of the human factors profession towards theoretical perspectives (see, e.g. Barber 1988) may be justifiable on the basis of previous failures in transfer, but should not prevent attempts based on mature findings and pragmatic theories rooted in real-world concerns.

References

ACKERMAN, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General,* **117,** 288–318.

ACKERMAN, P. L. (1987). Individual differences in skill learning: an integration of psychometric and information processing perspectives. *Psychological Bulletin,* **112,** 3–27.

ACKERMAN, P. & SHNEIDER, W. (1985). Individual differences in automatic and controlled

information processing. In R. DILLON, ED. *Individual Differences in Cognition,* Vol. 2. Orlando, FL: Academic Press, pp. 35–66.

ADAMS, J. A. (1989). Historical background and appraisal of research of individual differences in hearing. In R. KANTER, P. L. ACKERMAN & R. GUDECK, Eds. *Abilities, Motivation and Methodology: The Minnesota Symposium On Learning and Individual Differences.* Hillsdale, NJ: LEA, pp. 3–20.

ADAMS, J. A. (1987). Historical review and appraisal of research on the learning retention and transfer of human motor skills. *Psychological Bulletin,* **101,** 41–74.

ALLEN, B. (1994). Cognitive abilities and information system usability. *Information Processing and Management,* **30,** 177–191.

BARBER, P. (1988). In favour of theory. *Ergonomics,* **31,** 871–872.

BINET, A. & SIMON, T. (1905). New methods for diagnosing the intellectual level of non-normals. *L'Anne Psychologique,* **11,** 191–244.

BOBKO, P. & KARREN, R. (1982). The estimation of standard deviation in utility analysis. *Proceedings of the Academy of Management,* **422,** 272–276.

BOOTH, P. (1989). *An Introduction to Human–Computer Interaction.* Hillsdale, NJ: LEA.

BOOTH, P., FOWLER, C. & MACAULAY (1987). An investigation into business information presentation at human–computer interfaces. In H. BULLINGER & B. SHACKEL, Eds. *Interact '87.* Amsterdam: North-Holland, Elsevier, pp. 599–604.

BORING, E. G. (1929). *A History of Experimental Psychology.* New York: Appleton-Century.

BROGDEN (1950). When testing pays off. *Personnel Psychology,* **2,** 171–183.

CARROLL, J. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. KANTER, P. L. ACKERMAN & R. GUDECK, Eds. *Abilities, Motivation and Methodology: The Minnesota Symposium On Learning and Individual Differences.* Hillsdale, NJ: LEA, pp. 43–67.

CARROLL, J. (1972). Stalking the wayward factors. *Contemporary Psychology,* **17,** 321–324.

CARROLL, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytical Studies.* Cambridge: Cambridge University Press.

CATTELL, R. B. (1965). *The Scientific Analysis of Personality.* Hammondsworth: Penguin.

CHAPANIS, A. (1991). The business case for human factors in informatic. In B. SHAKEL & S. RICHARDSON. *Human Factors for Informatics Usability.* Cambridge: Cambridge University Press, pp. 39–72.

COOK, M. (1988). *Personnel Selection and Productivity.* New York: Wiley.

CHIANG, A. & ATKINSON, R. (1976). Individual differences and interrelationships among a select set of cognitive skills. *Memory and Cognition,* **4,** 661–672.

CRONBACH, L. (1984). *Essentials of Psychological Testing,* 6th edition. New York: Harper & Row.

CRONBACH, L. & GLESER, G. (Eds.) (1965). *Psychological Tests and Personnel Decisions.* Urbana, IL: University of Illinois Press.

DEMPSTER, F. N. (1981). Memory span: sources of individual and developmental differences. *Psychological Bulletin,* **89,** 63–100.

DIAPER, D. (1989). Giving HCI away. In A. SUTCLIFFE & L. MACAULAY, Eds. *People and Computers V.* Cambridge: Cambridge University Press, pp. 109–117.

DIGMAN, J. (1990). Personality structure: emergence of the five-factor model. *Annual Review of Psychology,* **41,** 417–440.

DILLON, A. (1987). A psychological view of user-friendliness. In H. BULLINGER & B. SHACKEL, Eds. *INTERACT'97.* Amsterdam: North-Holland, Elsevier, pp. 157–163.

DILLON, R. & SCHMECK, R. (1983). *Individual Differences in Cognition.* New York: Academic Press.

DUMAS, J. & REDISH, J. (1993). *A Practical Guide to Usability Testing.* Norwood, NJ: Ablex.

DUNNETTE, M. (1983). Aptitudes, abilities and skills. In M. DUNNETTE, Ed. *Handbook of Industrial and Organizational Psychology.* New York: Wiley, pp. 473–520.

EBBINGHAUS, H. (1895). Uber eine neve methode zur prufung geistiger fahigkeiten und ihre Andwendung be:Schulkindern. *Zeitschrift Psychologie,* **13,** 401–459.

EGAN, D. (1988). Individual differences in human–computer interaction. In M. HELANDER,

Ed. *Handbook of Human–Computer Interaction.* Amsterdam: North-Holland; Elsevier, pp. 543–568.

EKSTROM, R., FRENCH, J., HARMAN, H. & DERMEN, D. (1976). *Kit of Factor Referenced Cognitive Tests.* Princeton, NJ: Educational Testing Service.

EYSENCK, H. J. (1947). *Dimensions of Personality.* London: Routledge & Kegan Paul.

EYSENCK, H. J. (Ed.) (1982). *A Model for Intelligence.* New York: Springer-Verlag.

FITTS, P. & POSNER, M. (1967). *Human Performance.* Belmont, CA: Brooks/Cole.

FLEISHMAN, E. (1972). On the relation between abilities, learning and human performance. *American Psychologist,* **27,** 1017–1032.

FLEISHMAN, E. & HEMPEL, W. (1954). Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika,* **19,** 239–252.

FRENCH, J., EKSTROM, R. & PRICE, L. (1963). *Kit of Reference Tests for Cognitive Factors.* Princeton, NJ: Educational Testing Service.

FREYD, M. (1923). Measurement of vocational selection: an outline of research procedure. *Journal of Personnel Research,* **2,** 215–249.

GAGNE, R. (1967). *Learning and Individual Differences.* Columbus, OH: Charles Merrill.

GHISELLI, E. (1966). *The Validity of Occupational Aptitude Tests.* New York: Wiley.

GHISELLI, E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology,* **26,** 461–477.

GOUGH, H. (1958). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

GREENE, S., GOMEZ, L. & DEVLIN, S. (1986). A cognitive analysis of database query production. *Proceedings of the Human Factors Society. 30th Annual Conference.* Santa Monica, CA: Human Factors Society, pp. 9–13.

GUILFORD, J. (1967). *The Nature of Human Intelligence.* New York: McGraw-Hill.

GUILFORD, J. (1982). Cognitive psychology's ambiguities: some suggested remedies. *Psychological Review,* **89,** 48–59.

GUILFORD, J. & HOEPFNER, R. (1971). *The Analysis of Human Intelligence.* New York: McGraw-Hill.

GUION, R. (1976). Recruiting selection, and job placement. In M. DUNNETTE, Ed. *Handbook of Industrial and Organizational Psychology.* New York: Wiley, pp. 777–828.

HAYNES, J. (1970). Hierarchical analysis of factors in cognition. *American Educational Research Journal,* **7,** 55–68.

HORN, J. & KNAPP, J. (1973). On the subjective character of the empirical base of Guilford's structure of intellect model. *Psychological Bulletin,* **80,** 33–43.

HOUGH, L. (1992). The "big five" personality variables—construct confusion: description versus prediction. *Human Performance,* **5,** 139–155.

HULL, C. (1928). *Aptitude Testing.* New York: Harrup.

HUNT, E. (1978). The mechanics of verbal ability. *Psychological Review,* **85,** 109–130.

HUNT, E., LUNNEBORG, C. & LEWIS, J. (1975). What does it mean to be high verbal? *Cognitive Psychology,* **7,** 194–227.

HUNTER, J. & HUNTER, R. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin,* **96,** 72–98.

HUNTER, J., SCHMIDT, F. & PEARLMAN, K. (1982). History and accuracy of validity generalization equations: a response to the Callender & Osborn reply. *Journal of Applied Psychology,* **67,** 853–858.

JENKINS, J. (1989). The more things change, the more they stay the same: comments from a historical perspective. In R. KANTNER, P. ACKERMAN & R. CUDECK, Eds. *Abilities, Motivation and Methodology: The Minnesota Symposium on Learning and Individual Differences.* Hillsdale, NJ: LEA.

JENSEN, J. (1982). Reaction time and psychometric g. in H. J. EYSENCK, Ed. *A Model for Intelligence.* Berlin: Springer-Verlag, pp. 93–132.

JOHNSON-LAIRD, P. (1985). Deductive reasoning ability. In R. STERNBERG & M. GARDNER, Eds. Unities in inductive reasoning. *Journal of Experimental Psychology* (*General*)*,* **112,** 173–194.

KAGAN, J., ROSMAN, B., DAY, D., ALBERT, J. & PHILLIPS, N. (1964). Information processing in the child: significance of analytic and reflective attitudes. *Psychological Monographs,* **78**. Washington: American Psychological Association.

KARAT, J. (1991). *Taking Software Design Seriously.* New York: Academic Press.

KINCAID, M. (1925). A study of individual differences in learning. *Psychological Review,* **32**, 34–53.

KLINE, P. (1994). Review of Carroll's human cognitive abilities. *Applied Cognitive Psychology,* **14**, 387–399.

KYLLONEN, P. (1993). Aptitude testing inspired by information processing: a test of the four-sources model. *The Journal of General Psychology,* **120**, 375–405.

KYLLONEN, P., TIRRE, W. & CHRISTAL, R. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General,* **120**, 57–79.

LANDAUER, T. (1995). *The Trouble with Computers: Usefulness, Usability and Productivity.* Cambridge, MA: MIT Press.

LANDY, F., SHANKSTER, L. & KOHLER, S. (1994). Personnel selection and placement, *Annual Review of Psychology,* **45**, 261–296.

LEDVINKA, J. & SIMONET, J. (1983). *The dollar value of JEPS at life of Georgia.* Working paper 83–134, College of Business Administration, University of Georgia, GA, USA. (cited in Cook, M., 1988).

LENT, R., AURBACH, H. & LEVIN, L. (1971). Predictors, criteria nad significant results. *Personnel Psychology,* **24**, 519–533.

LOHMAN, D. & KYLLONEN, P. (1983). Individual Differences in solution strategy or spatial tasks. In R. DILLON & R. SCHMECK, Eds. *Individual Differences in Cognition.* New York: Academic Press.

LOHMAN, D. (1993). Implications of cognitive psychology for ability testing: three critical assumptions. In M. RUMSEY, G. WALKER & J. HARRIS, Eds. *Personnel Measurement*: *Directions for Research.* Hillsdale, NJ: Erlbaum.

MAYER, R. (1982). Different problem-solving strategies for algebra word and equation problems. *Journal of Experimental Psychology*: *Learning, Memory and Cognition,* **8**, 448–462.

NIELSEN, J. (1993). *Usability Engineering.* Cambridge, MA: Academic Press.

NESSELROADE, J. & CATTELL, R. (1988). *Handbook of Multivariate Experimental Psychology,* 2nd Edition. New York: Plenum Press.

NETTELBECK, T. (1982). Inspection time: an index of intelligence. *Quarterly Journal of Experimental Psychology,* **34**, 299–312.

NORMAN, D. & DRAPER, S. (1986). *User–Centred systems Design.* Hillsdale, NJ: LEA.

OVERMIER, J., MONTAGUE, W. & JENKINS, J. (1989). Prolegomenon. In R. KANTER, P. L. ACKERMAN & R. GUDECK, Eds. *Abilities, Motivation and Methodology: The Minnesota Symposium On Learning and Individual Differences.* Hillsdale NJ: LEA, pp xix–xxi.

PASK, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology,* **46**, 128–148.

PERFETTI, C., GOLDMAN, S. R. & HOGABOAN, T. (1978). Reading skill and the identification of words in discourse context. *Memory & Cognition,* **7**, 273–282.

ROCHE, W. (1965). A dollar criterion in fiscal-treatment employee selection. In L. CRONBACH & G. GLESER, Eds. *Psychological Tests and Personnel Decisions.* Urbana, IL: University of Illinois Press.

SCHMIDT, F. & HUNTER, J. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology,* **62**, 529–540.

SCHMIDT, F. & HUNTER, J. (1981). Employment testing: old theories and new research findings. *American Psychologist,* **36**, 1128–1137.

SCHMIDT, F. & HUNTER, J. (1983). Individual differences in productivity: an empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology,* **68**, 407–414.

SCHMIDT, F., HUNTER, J. & CAPLAN, J. (1981). Validity generalization for jobs in the petroleum industry. *Journal of Applied Psychology,* **64**, 609–626.

SEASHORE, R. (1928). Individual differences in motor skills. *Journal of General Psychology,* **3,** 38–65.

SEIN, M., OLFMANN, L., BOSTROM, R. & DAVIS, S. (1993). Visualization ability as a predictor of user learning success. *International Journal of Man–Machine Studies,* **39,** 599–620.

SHACKEL, B. & RICHARDSON, S. (1991). *Human Factors for Informatics Usability.* Cambridge: Cambridge University Press.

SPEARMAN, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology,* **15,** 201–293.

STERNBERG, R., (Ed.) (1985). *Human Abilities: An Information-Processing Approach.* New York: W. H. Freeman & Co.

THORNDIKE, E. L. (1908). The effect of practice in the case of a purely intellectual function. *American Journal of Psychology,* **19,** 374–384.

THORNDIKE, R. (1954). The psychological value systems of psychologists. *American Psychologist,* **9,** 787–789.

THURSTONE, L. (1938). *Primary Mental Capabilities.* Chicago, IL: University of Chicago Press.

TIFFIN, J. (1943). *Industrial Psychology.* New York: Prentice-Hall.

TIRRE, W. & PENA, C. (1993). Components of quantitative reasoning general and group ability factors. *Intelligence,* **17,** 501–521.

VERNON, P. (1972). The distinctiveness of field independence. *Journal of Personality,* **40,** 366–391.

VICENTE, K. & WILLIGES, R. (1988). Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies,* **29,** 647–668.

WHITELY, S. (1980). Modeling aptitude test validity from cognitive components. *Journal of Educational Psychology,* **72,** 750–769.

WITKIN, H., MOORE, C., GOODENOUGH, D. & COX, P. (1977). Field dependent and field independent cognitive styles and their educational implications. *Review of Educational Research,* **47,** 1–64.