Despite its avowed goal of understanding individual behavior, the field of behavior analysis has largely ignored the determinants of consistent differences in level of performance among individuals. The present article discusses major findings in the study of individual differences in intelligence from the conceptual framework of a functional analysis of behavior. In addition to general intelligence, we discuss three other major aspects of behavior in which individuals differ: speed of processing, working memory, and the learning of three-term contingencies. Despite recent progress in our understanding of the relations among these aspects of behavior, numerous issues remain unresolved. Researchers need to determine which learning tasks predict individual differences in intelligence and which do not, and then identify the specific characteristics of these tasks that make such prediction possible.

Key words: intelligence, learning, three-term contingency, individual differences, processing speed, working memory, humans

The surprising and consistent empirical finding in psychometric intelligence research is that people who do well on one mental task tend to do well on most others, despite large variations in the tests’ contents... This was Spearman’s (1904) discovery, and is arguably the most replicated result in all psychology. (Deary, 2000, p. 6)

Historically, the study of individual differences has been an area of research relatively separate from experimental psychology. While experimental psychology has focused on the processes that determine performance in specific experimental situations, the field of individual differences has studied the stable differences among people, particularly those that generalize across diverse situations. The behavioral differences that have received the most attention in this regard have been personality traits and cognitive abilities. Behavior analysis has largely ignored such differences, other than those that are explainable in terms of reinforcement history. This disregard of individual differences is puzzling, given that behavior analysts emphasize that their research focuses on the behavior of individuals rather than on group averages. After all, it is the differences among individuals that distinguish them from the average.

The disregard of individual differences also is surprising because the agenda of behavior analysis includes the analysis of behavior in educational settings, and individual differences in performance are among the most salient aspects of behavior in such settings. Careful programming of environmental contingencies often can improve educational accomplishment, although the fact remains that individuals vary widely in how effectively they deal with academic topics. Some learn and understand complex material with relative ease, whereas others labor to succeed and nevertheless frequently fail in their efforts. Given the prominence of individual differences at every level of educational endeavor, it is surprising that behavior analysts have made so little effort to understand them.

Individual differences in educational performance are strongly related to differences in intelligence, a major focus of individual-differences research. ‘Intelligence’ has multiple meanings—so many in fact, that one of the most prominent researchers in the area has argued that the term should be abandoned (Jensen, 1998). Lurking within this diversity of meaning, however, are important facts that pose serious explanatory challenges to any approach to psychology that aspires to encompass the field’s most basic empirical phenomena. The purpose of this article is to describe some of the most essential of these findings, and to consider their implications for a functional analysis of behavior.
Numerous tests have been developed to help assess intelligence, including tests of vocabulary, short-term memory span, analogical reasoning, story construction from pictures, etc., with such diversity seemingly belying the usefulness of intelligence as an explanatory construct. Should the diversity of the tests putatively measuring intelligence be taken as evidence that, rather than there being one fundamental ability that distinguishes among people, there is a diverse set of cognitive skills in which people may differ? Would it be better, then, to consider these skills as separate behaviors with their own individual controlling variables? The few behaviorists who have addressed the topic of intelligence (e.g., Staats & Burns, 1981) have tended to adopt just such an approach, and nonbehavioral proponents of this view have had considerable popularity (e.g., Gardner, 1983). Such an interpretation certainly has intuitive appeal, but there is one very large elephant in the room that cannot be ignored: Performances on all of these tests, despite the obvious differences between the tests themselves, are significantly correlated with each other.

A specific example of such correlations is provided in Table 1, which shows the correlation matrix for the 14 subtests of one of the most widely used intelligence tests, the Wechsler Adult Intelligence Scale (WAIS-III; Psychological Corporation, 1997). Only 2 of the 91 correlations among the subtests are below .30. Most (74 of the 91) are in the range of .40–.70. For example, consider the correlations of the Block Design subtest with the other 13. On each trial of the Block Design test, an individual is shown a pattern and then must reconstruct that pattern using a set of colored cubes, and the total score is based on both time and accuracy. As may be seen in Table 1, the correlations between Block Design and each of the other subtests are surprisingly similar; all but one are in the .40–.70 range (the one exception being the correlation between Block Design and Digit Span).

The issue raised by the pattern of correlations shown in Table 1 is why the correlations among subtests are all so similar. For example, Block Design is as strongly related to Vocabulary as it is to the other subtests that do not involve verbal materials. It is clear that the various tests share a great deal of common variance, despite large differences in their structure and in the stimuli of which they are composed. In the intelligence literature, this common variance often is taken as evidence of general intelligence, or simply \(g\).

The amount of shared variance, and the relative contributions of the various subtests to it, can be measured using principal components analysis or PCA. In PCA, the first principal component or general factor, which is often taken as the operational definition of \(g\), is simply the linear combination of the standardized scores on the subtests that accounts for the greatest amount of variance. If one thinks of each individual’s subtest scores as a point in a space with as many dimensions as there are subtests, then PCA weights these scores so as to minimize the sum of the squared perpendicular distances from the data points to the line corresponding to the

### Table 1
Correlation Matrix for the 14 Subtests from the WAIS-III.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>V</th>
<th>S</th>
<th>A</th>
<th>DS</th>
<th>I</th>
<th>C</th>
<th>LN</th>
<th>PC</th>
<th>DC</th>
<th>BD</th>
<th>MR</th>
<th>PA</th>
<th>SS</th>
<th>OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarities</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arithmetic</td>
<td>.60</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>.45</td>
<td>.40</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>.77</td>
<td>.70</td>
<td>.63</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>.75</td>
<td>.70</td>
<td>.57</td>
<td>.39</td>
<td>.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter-Number Seq.</td>
<td>.50</td>
<td>.46</td>
<td>.55</td>
<td>.57</td>
<td>.47</td>
<td>.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture Completion</td>
<td>.47</td>
<td>.48</td>
<td>.40</td>
<td>.30</td>
<td>.46</td>
<td>.46</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit-Symbol Coding</td>
<td>.44</td>
<td>.40</td>
<td>.43</td>
<td>.36</td>
<td>.38</td>
<td>.37</td>
<td>.44</td>
<td>.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>.50</td>
<td>.52</td>
<td>.54</td>
<td>.36</td>
<td>.48</td>
<td>.49</td>
<td>.43</td>
<td>.52</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>.54</td>
<td>.52</td>
<td>.58</td>
<td>.42</td>
<td>.53</td>
<td>.52</td>
<td>.47</td>
<td>.48</td>
<td>.40</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture Assembly</td>
<td>.53</td>
<td>.52</td>
<td>.44</td>
<td>.33</td>
<td>.54</td>
<td>.50</td>
<td>.39</td>
<td>.49</td>
<td>.37</td>
<td>.49</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbol Search</td>
<td>.48</td>
<td>.48</td>
<td>.52</td>
<td>.41</td>
<td>.46</td>
<td>.44</td>
<td>.49</td>
<td>.49</td>
<td>.65</td>
<td>.53</td>
<td>.48</td>
<td>.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Assembly</td>
<td>.44</td>
<td>.47</td>
<td>.39</td>
<td>.26</td>
<td>.40</td>
<td>.45</td>
<td>.29</td>
<td>.52</td>
<td>.33</td>
<td>.61</td>
<td>.49</td>
<td>.46</td>
<td>.47</td>
<td></td>
</tr>
</tbody>
</table>

Note: Labels in the top row correspond to the letters printed in bold in the Subtest Column.
component axis. Importantly, the weight given a specific subtest in this linear combination, which is frequently referred to as its loading on \( g \), corresponds to the correlation between individuals’ scores on that subtest and their component scores, calculated as the weighted sum of their scores on all of the subtests.

For the WAIS-III correlation matrix shown in Table 1, the first principal component accounts for 51% of the total variance. Table 2 shows the loadings of the various WAIS-III subtests on the first principal component. As can be seen, the highest loading is for the Vocabulary subtest (.83), and the lowest is for Digit Span (.60), indicating that relative to the other subtests, the first principal component (or \( g \)) explains the most variance in Vocabulary scores and the least variance in Digit Span scores. Many researchers in the area of individual differences agree that the fundamental theoretical question about the nature of intelligence is how \( g \) should be conceptualized so as to capture the fact that the majority of variance is shared by such a diverse collection of subtests, while at the same time accounting for the differences among the subtests in terms of their contributions to this shared variance.

What does this mean from the standpoint of behavior analysis? Simply put, it means that an individual’s behavior (i.e., a person’s test performance relative to that of other individuals) is consistent from subtest to subtest. In principle, an above-average overall score on an intelligence test could indicate that an individual is far above average on a few subtests and below average on the rest, yet this is relatively rare. The universally positive correlations among the various subtests mean that people who are above average overall tend to be above average on all of the subtests and people who are below average overall tend to be below average on all of the subtests.

In fact, this consistency in individual behavior is the heart of the matter—it is what is responsible for the universally positive correlations among the subtests and the relative similarity of their loadings on the first principal component or \( g \). These three findings—individual consistency, positive intercorrelations, and roughly similar subtest loadings—are essentially one, and each of the three findings follows mathematically from either of the other two. Together they represent the essence of the puzzle of intelligence, which is, “Why do some people tend to be good at lots of things while others tend to be bad at lots of things? And why does this tendency to be consistently good or bad apply more to some things than others?” The “things” referred to here are exemplified by the various subtests, but these subtests are of interest because they presumably stand for many other behaviors in everyday life, such as various tasks that need to be done, and problems that need to be solved.

**VARIETIES OF INTELLIGENCE**

One way to approach the issue of the consistency in individual differences in performance is to compare subtests with high \( g \)
loadings with subtests that have low \( g \) loadings, and try to identify the critical dimensions along which these subtests differ. Interestingly, the subtests of the WAIS-III that have the highest \( g \) loadings (i.e., Vocabulary, Similarities, Information, Comprehension, and Arithmetic) are those that tap previously acquired knowledge and skills. Tests that tap previously acquired knowledge and skills are said to reflect crystallized intelligence. In contrast, tests that are designed to be as free as possible from prior knowledge, and depend only on current, on-line processing, are said to be tests of fluid intelligence, the prototypical example being the figural analogies of Raven’s Progressive Matrices (Raven, Raven, & Court, 2000).

When diverse batteries of subtests are subjected to factor analysis, typically two factors emerge, one a fluid factor and the other a crystallized factor, as indicated by the nature of the subtests that load on these factors (Horn & Cattell, 1966). The distinction between crystallized and fluid intelligence is supported by their different functional properties, especially with respect to the differential effects of adult age. Whereas fluid intelligence begins its decline in the 20s, crystallized intelligence shows relatively little decline in healthy adults until they reach their 70s, and some tests of crystallized intelligence (e.g., vocabulary tests) even show a slight increase over this same period (for a review, see Deary, 2000, Chapter 8). The two categories of intelligence are differentially sensitive to brain damage of various sorts, with little impairment typically evident for crystallized intelligence but major deficits for fluid intelligence. This pattern has been observed, for example, in patients with white matter lesions (Leaper et al., 2001) and in those with frontal lobe lesions (Duncan, Burgess, & Emslie, 1995), as well as in patients with Huntington’s Disease, Parkinson’s Disease, and mild Alzheimer’s Disease (Psychological Corporation, 1997).

The distinction between fluid and crystallized intelligence is only one of several different partitions of the total variance across different intelligence tests. Other schemes have identified other broad categories of variance (e.g., verbal/educational vs. spatial/mechanical), sometimes with additional, somewhat less broad categories such as retrieval ability and processing speed. The specific structure provided by factor analysis is somewhat arbitrary because it reflects the specific assortment of tests that are included in any given analysis. In addition, the more tests included in a given test battery, the greater the number of covariance clusters that can be identified, with each cluster signifying an ability that is partially distinct from other abilities. But regardless of the complexity of the covariance partitions and the number of factors that emerge as a result, there are always positive correlations between the factors, just as there are universally positive correlations between the individual tests and between the subtests (Jensen, 1998), so that one can always identify a common factor that accounts for a major portion of the total variance (usually more than 50%).

Given that \( g \) can be extracted from any array of separate tests, a critical issue is how \( g \) factors extracted from separate test batteries are related. If the nature of the \( g \) factor were to depend on the specific composition of the test battery from which it is extracted, then \( g \) would be arbitrary and of much less theoretical interest. A recent study (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004) administered three extensive, independently developed intelligence test batteries to 436 individuals and examined the correlations among \( g \) factors extracted from the three different batteries: (1) the Comprehensive Ability Battery, consisting of 14 different subtests (Hakstian & Cattell, 1975); (2) a version of the Hawaii Battery, which included the Raven matrices in shortened and modified form as well as 16 other subtests (DeFries et al., 1974); and (3) the original version of the WAIS, which consisted of 11 subtests (Wechsler, 1955). The \( g \) factors from the three test batteries correlated .99 (!), providing strong support for the hypothesis that the shared variance captured by \( g \) represents a fundamental fact of human abilities and is not an arbitrary result of the composition of specific tests.

EXPLAINING THE SHARED VARIANCE

Is Speed What Is Shared?

One major hypothesis regarding the consistency in individual performance that underlies \( g \) has been that it reflects individual differences in processing speed (Jensen, 1998). There is growing evidence of individual consistency in speed of responding across very diverse tasks: Some people are consistently fast on most
tasks, whereas others are consistently slow (Hale & Jansen, 1994; Myerson, Hale, Zheng, Jenkins, & Widaman, 2003; Zheng, Myerson, & Hale, 2000). This consistency is reminiscent of that observed on intelligence tests and suggests that individual differences in the speed with which people process information could affect performance on many different psychological tests and everyday tasks.

The consistency in an individual’s response times may be seen in Figure 1, which shows individual performance on four quite different tasks: choice reaction time, letter classification, visual search, and abstract matching (Hale & Myerson, 1993). Each of the four panels depicts the mean response times of a different individual plotted as a function of the average response times for the whole group of 65 undergraduates who were tested on these tasks. As may be seen, the 2 slow individuals (top panels) were slower than average in all task conditions (i.e., their data points all lie above the diagonal representing average performance) and the 2 fast individuals (bottom panels) were faster than average in all task conditions (i.e., their data points all lie below the diagonal). In addition, the size of the difference between individual and average response times (represented by the vertical distance from the diagonal) increased in an orderly fashion as the difficulty of the task increased.

Fig. 1. Response times (RTs) for 4 individuals from Hale and Myerson (1993) plotted as a function of the group mean RT. Each panel presents data from a different individual. Data from the 5th, 10th, 55th, and 60th slowest subjects out of the 65 subjects (ranked based on their slopes) are shown in the upper left, upper right, lower left, and lower right panels, respectively. Within each panel, each data point represents performance in one experimental condition. The tasks, in order of increasing difficulty, were as follows: two-choice RT, inverted white triangles (one condition); letter classification, gray squares (three conditions); visual search, white upright triangles (four conditions); abstract matching, gray diamonds (two conditions). The dashed diagonal line represents equality; if an individual’s mean RT for a particular task condition did not differ from the corresponding group mean RT (as is approximately the case for the two-choice RT task for each of the individuals shown), then the data point for that condition would fall along the equality line.
Additional data showing the same pattern from 6 different individuals as well as the fast and slow quartiles of this sample may be seen in Figures 1 and 8 of Myerson et al. (2003). Comparable data from a separate study using seven different tasks are presented in Hale and Jansen (1994). Taken together, these results strongly support the idea that an individual’s speed is a general characteristic of that individual, and has equivalent, multiplicative effects on the time required for any information processing task, regardless of the task being performed (for more details and a formal model, see Myerson et al., 2003).

The potential implications of such general characteristics for differences in intelligence may be seen in Figure 2, which depicts performance by university students and by students at a vocational college, groups that differed in average academic aptitude, plotted as a function of the average response times calculated across both groups combined (Vernon & Jensen, 1984). Not only was the higher ability group faster on all of the tasks (choice reaction time, memory scanning, same/different word judgment, and synonym/antonym judgment), but also the size of the difference from average increased linearly as a function of task difficulty, just as in the Hale and Myerson (1993) data seen in Figure 1. These data show relatively little evidence of specific strengths or weaknesses on particular tasks or in particular task conditions. Clearly, results like these are difficult to reconcile with the goal of the once-popular chronometric approach to individual differences, which strove to identify distinct cognitive processes through componential analysis of response times, and then relate individual differences in these components to differences in higher cognitive abilities (e.g., Sternberg, 1977).

One of the major proponents of the importance of processing speed for higher cognitive abilities has been Salthouse (1996), who offered two reasons why speed should be so important. First, if you are slow to process information, and you cannot control the rate at which it is presented, then you are likely to miss information, some of which may be needed for the behavior in which you are engaged. Second, coordination between two different tasks is likely to be impaired if you are slow, because you may take so long on one task that you forget information that is needed to perform the other task.

It should be noted that the purely behavioral analyses of processing speed data presented here, as well as previous analyses reported in this journal and elsewhere (e.g., Chen, Hale, & Myerson, 2003; Myerson, Adams, Hale, & Jenkins, 2003; Myerson, Hale, Hansen, Hirschman, & Christiansen, 1989; Myerson, Robertson, & Hale, 2007), aptly illustrate the potential utility for behavior analysis of studying individual differences. As Underwood (1975) famously argued, individual differences can be thought of as natural experiments, and the results of such natural experiments can help assess the validity of theoretical constructs and models. As Underwood noted, “The whole idea behind behavioral theory is to reduce the number of independent processes to a minimum; to find that performance on two
apparently diverse tasks is mediated at least in part by a single, more elementary, process is a step towards this long-range goal” (p. 133). Whereas cognitive psychologists typically assume that a set of diverse tasks of varying complexity, such as the various tasks that generated the data for the analyses presented here, involve different collections of processing components, the orderly linear relations between individual differences and group average response times on these tasks imply that the tasks primarily vary along a single dimension, which might be termed simply difficulty.

The finding that the size of individual differences varies as a simple multiplicative function of task difficulty is consistent with what would be expected if speed were the source of the shared variance on intelligence tests. Regardless of whether that claim is ultimately supported or not, however, we believe the finding stands on its own as revealing an important regularity in individual behavior, one that already has shed substantial light on the behavioral processes underlying changes observed in both child development (e.g., Fry & Hale, 1996) and aging (e.g., Myerson et al., 2003).

Or Is It Working Memory?

There can be little doubt that a substantial fraction of the variance in intelligence scores is related to measures of processing speed (for a review, see Vernon, 1983), but numerous investigators have questioned its adequacy as a complete account of g (e.g., Stankoff & Roberts, 1997). A popular alternative to processing speed as the major correlate of g is working memory capacity. Cognitive psychologists continue to disagree about specific aspects of the working memory construct, but it is generally assumed that information is maintained in a temporary memory buffer while it is being processed as well as while other information is being processed. Separate buffers for verbal and visuospatial information have been proposed, along with an executive function to organize and allocate limited attentional resources (Baddeley, 1986). Various other executive functions also have been proposed, such as updating and monitoring, switching between mental sets, and inhibiting competing responses (e.g., Miyake, Friedman, Emerson, Witzi, & Howerth, 2000).

Many researchers assume that competition between the maintenance and processing functions of the working memory system provides the basis for the differences between individuals (e.g., Engle, Tuholski, Laughlin, & Conway, 1999). More capable individuals are assumed to have greater working memory “capacity” and/or a more effective attentional or executive system that allows the memory system to be less disrupted when simultaneous processing is required. It even has been suggested that reasoning ability is little more than working memory capacity (Kyllonen & Christal, 1990), although this strong claim has been strenuously criticized (Ackerman, Beier, & Boyle, 2005).

The importance of working memory in everyday life is exemplified by its role in reading. To comprehend what one is currently reading, one must relate the present text to the portion of the material already read. To the extent that the ability to retain prior relevant information while processing new information is deficient, the reader will need to frequently recheck the earlier material, greatly impeding the efficiency of reading and thereby limiting the actual amount learned. Not surprisingly, tests of verbal working memory strongly predict various aspects of verbal ability, including vocabulary, verbal comprehension, and the ability to infer word meanings from context (Daneman & Carpenter, 1980; Daneman & Merikle, 1996).

To get a sense of what a working memory task is like, consider the “alphabet recoding” task used by Kyllonen and Christal (1990). Subjects are presented a series of three letters and then an instruction such as “+ 2” that tells them how to transform the original letters before reporting the transformed series. If the letters were “C-K-W,” for example, then the correct answer would be “E-M-Y.” To perform the task correctly, subjects must transform the first letter without forgetting the other two letters, then transform the second letter without forgetting either the previously transformed first letter or the untransformed third letter, and then finally transform the third letter without forgetting the previous two transformations.

The alphabet span task (Craik, 1986), in which subjects hear a list of words and then must recall them in alphabetical order, is another...
example of a working memory task involving transformations.

Dual-task procedures represent a more commonly used type of working memory test. With these procedures, to-be-remembered items are presented individually, alternating with other stimuli that must be responded to without forgetting any of the to-be-remembered items. One prominent example is the operation span task, in which words are presented alternately with simple arithmetic equations (e.g., $3 + 5 = 7$) that must be judged for correctness. After an entire list of words and equations has been presented, the subject must recall the words in order. As may be noted, competition and interference are what distinguish all these tasks from simple short-term memory span tasks. In the case of the operation span task, for example, the equations and the words compete to control recall responses, and the process of judging the correctness of the equations also may interfere with recall. (For more detailed information about this and other working memory task procedures, see Conway et al., 2005; and Waters & Caplan, 2003.)

A variety of specific executive functions have been proposed, but recent studies have failed to support the idea that they are part of a unitary executive system. For example, at least one of these putative executive functions, task switching, has been found to be unrelated to working memory span measures (Oberauer, Süß, Wilhelm, & Sander, 2007). Other failures to observe predicted covariation have been reported as well, not only between different putative executive functions (e.g., Miyake et al., 2000; Ward, Roberts, & Phillips, 2001) but also between different measures of the same function (e.g., Shilling, Chetwynd, & Rabbitt, 2002). Thus, rather than explaining individual differences in $g$, the concept of executive function seems both too fuzzy and too variegated to offer a gain in explanatory parsimony over the phenomena to be explained. Even Alan Baddeley, who pioneered the concept of a working memory system with a “central executive,” has explicitly stated that the “central executive” is merely a conceptual placeholder that serves as a reminder of what remains to be explained (Baddeley, 2001).

Despite these conceptual and empirical problems with the executive aspects of the working memory construct, several features seem interpretable within a behavior-analytic account, and measures based on these features may eventually be shown to be critical predictors of individual differences in $g$. For example, Engle and his colleagues recently proposed that the reason that working memory performance predicts fluid intelligence is not primarily because memory per se is used to solve reasoning problems; rather, it is because working memory tasks measure how well an individual’s attentional processes function under conditions of interference, distraction, or conflict” (Kane, Conway, Hambrick, & Engle, 2007; Unsworth & Engle, 2005; see also Hasher, Lustig, & Zacks, 2007). This account may be compared to a description of working memory tasks from a behavioral perspective. We would say that briefly presented stimuli must continue to control appropriate recall responses even after those brief stimuli are replaced by other stimuli that control different responses; on some tasks, previously presented stimuli must also compete with self-generated stimuli resulting from covert responses (e.g., transformations). In other words, working memory tasks involve competition for control (or “attention”) among stimuli, most of which are no longer present, and it is plausible that individuals will differ substantially in the outcome of such competition, perhaps in a way that predicts their performance on intelligence tests.

An analysis of individual differences in working memory in terms of individual differences in stimulus control seems achievable and clearly relevant to understanding the nature of intelligence. Recent theoretical accounts, however, have retreated from the strong view (e.g., Kyllonen & Chrystal, 1990) that fluid intelligence and working memory capacity are essentially isomorphic, in part because studies have shown that working memory tasks are far from perfectly correlated with intelligence (e.g., Ackerman et al., 2005). This means that much of the variance in $g$ must be explained by factors other than working memory. The results of our own research in this area, discussed in the following section, suggest that competition among stimuli for control plays a role in human learning as well as in working memory, and that this broader role provides a fuller (albeit far from complete) account of individual differences in $g$ (Tamez, Myerson, & Hale, 2008; Williams & Pearlberg, 2006).
THE ROLE OF LEARNING

A remarkable feature of most recent attempts to identify the basic underlying dimensions that account for individual differences in general intelligence has been the neglect of associative learning. This neglect seems strange given the obviously important role played by associative learning in the acquisition of knowledge and skills that underlie performance on tests of crystallized intelligence (e.g., the Vocabulary and General Information subtests of the WAIS-III). In part, the neglect of associative learning may be a consequence of the view that knowledge and skill acquisition are the result of an inferential process rather than the result of forming associative connections. For example, the meaning of a new word is commonly inferred from the context in which it is encountered, rather than being learned from flash cards or by looking it up in the dictionary. Regardless of how learning is conceived, however, crystallized intelligence obviously reflects prior learning, and presumably also reflects individual differences in the efficiency of the learning process.

In addition, much recent research on intelligence seems out of touch with the goal of predicting educational achievement, which was the original purpose of intelligence tests and is still their major application. Recently, however, Luo, Thompson, and Detterman (2006) examined the extent to which basic information-processing tasks could replace conventional intelligence tests as predictors of children’s performance on academic achievement tests. Luo et al. used multiple regression and structural equation models to analyze data from two large, representative samples: a normative sample of nearly 5,000 children, ages 6–19 years, that had been used to standardize the multi-faceted Woodcock-Johnson Cognitive Abilities and Achievement Tests, and a separate sample of more than 500 children, ages 6–13 years, from the Western Reserve Twin Project. For both samples, analyses showed that fluid intelligence tests could be replaced as predictors of academic achievement by measures of processing speed and working memory. In contrast, tests of crystallized intelligence could not be replaced because they explained a substantial amount of the variance in academic achievement that was not accounted for by the information-processing measures. In the Woodcock-Johnson normative sample, for example, a combination of crystallized intelligence and basic information-processing abilities accounted for more than one-half of the variance in academic achievement, of which approximately 40% was attributable to crystallized intelligence alone and approximately 45% was common to both crystallized intelligence and information processing ability (Luo et al., 2006). If a major goal of intelligence testing is the prediction of academic achievement, then assessment of learning ability, which is a major determinant of crystallized intelligence, would appear to be important for achieving that goal.

It is important to note that not all types of associative learning are related to intelligence, and in fact, early studies failed to find a meaningful relation between intelligence scores and rate of learning on a variety of associative learning tasks (Underwood, Borch, & Malmi, 1978; Woodrow, 1946). These findings contributed to the subsequent disregard of learning by intelligence researchers, but they also may be a major clue as to the role of learning ability in performance on intelligence tests. Recently, Williams and Pearlberg (2006) replicated the low correlation between some measures of verbal learning and intelligence. They found that both paired-associate learning and free-recall list learning had correlations below .20 with Raven’s Advanced Progressive Matrices (Raven, Raven, & Court, 1998). More complex learning tasks (e.g., learning to write computer programs; Kyloren & Stephens, 1990) have produced more substantial correlations with intelligence, but because of their very complexity, neither what is being learned on such tasks, nor how it is being learned, is clearly understood.

In contrast to the weak correlations observed between traditional verbal learning tasks and intelligence tests, Williams and Pearlberg (2006) have developed a novel verbal learning task that shows a surprisingly high correlation with scores on a test of fluid intelligence. In their new task, each subject learns a set of “three-term contingencies” modeled after the basic unit of operant conditioning. More specifically, subjects see 10 stimulus words presented one at a time on a computer screen. In the presence of each stimulus word, subjects first press the “A” key,
then the “B” key, and finally the “C” key, with each response producing a different outcome word. For example, given the stimulus word “LAB,” pressing the “A” key produces “PUN,” pressing the “B” key produces “TRY,” and pressing the “C” key produces “EGG.” Given a different stimulus word (e.g., “RUM”), the same set of three responses produces a different set of outcome words (e.g., A→FAT, B→CAN, C→TIC). In subsequent testing, subjects are presented with all 30 stimulus word–response combinations and, in each case, they have to try and provide the appropriate outcome word (e.g., LAB, A→ ? Correct response = PUN). When Williams and Pearlberg had college students perform this task, they found that students’ performance on the three-term learning task correlated strongly (r = .50) with their scores on the Raven’s Advanced Progressive Matrices.

Williams and Pearlberg conducted a follow-up study (unpublished) in which they compared their three-term learning task with a two-term associative learning task designed to mimic their three-term task as closely as possible. In the two-term learning task, ten stimulus words were each paired sequentially with three different words but without intervening responses to keys A, B, and C. During testing, the subject was asked to recall each of three words that had been associated with each stimulus. Despite the fact that both the type and the amount of material to be learned on this two-term task was similar to that on the three-term contingency task, the correlation between learning rate and Raven scores was only about .25, approximately half that obtained with the three-term task. This difference could not be attributed to differences in task difficulty. These findings clearly demonstrate that learning ability is an important component of fluid intelligence, but they also raise an important question: How can simply adding the middle (response) term of the three-term contingency lead to such a substantial increase in the correlation between learning and intelligence test scores?

One way to approach this question is to see how performance on the three-term learning task relates to performance on other types of basic information-processing tasks. Williams and Pearlberg (2006) originally reported that the three-term contingency task was not significantly correlated with working memory and processing speed tasks, but in subsequent research they have found occasional significant correlations between three-term contingency learning and some working memory tasks (but as yet no significant correlation with any speed-of-processing task has been observed). Importantly, the correlations between working memory and Raven scores have always been smaller than the correlations between three-term verbal learning and Raven scores.

Tamez et al. (2008) recently replicated the substantial correlation between Williams and Pearlberg’s (2006) three-term contingency learning and Raven scores, and extended this finding to an essentially similar three-term contingency task that used visual-spatial patterns as stimuli. Moreover, as in the research by Williams and Pearlberg, Tamez et al. found that the three-term verbal learning task correlated more strongly with Raven’s Advanced Progressive Matrices than any of the three working memory tasks used in their study. Of the three, the operation span task, which is becoming a standard for assessing working memory capacity (Conway et al., 2005), correlated most highly with Raven scores (r = .395). However, this correlation was still less than that between the three-term verbal learning task and Raven scores (r = .489). In addition, multiple regression analyses revealed that performance on the three-term learning task accounted for all of the variance in Raven scores explained by operation span as well as additional variance unique to three-term contingency learning.

Thus, learning, or at least learning on tasks like the three-term contingency learning task developed by Williams and Pearlberg (2006), is among the very best predictors of performance on fluid intelligence tests. From one perspective, this should not be surprising. Ever since Binet and Simon (1905, 1916) developed the first standardized intelligence test, a major purpose of such tests has been to aid in the identification of children who were likely to have problems learning in regular schools. It follows that if one develops an intelligence test that can predict learning, then given the nature of correlations, learning should also predict performance on that test.

From another perspective, however, the relation of learning ability to intelligence test scores is surprising, at least with respect to fluid intelligence tests. After all, tests of fluid intelligence such as Raven’s Progressive Matrices were developed originally with the goal of
measuring cognitive ability independent of past learning. Such a goal may be unattainable, however, if individuals taking fluid intelligence tests are actually learning the correct solution strategies as they proceed, as some researchers recently have suggested (e.g., Carlstedt, Gustafsson, & Ullstadius, 2000; Verguts & De Boeck, 2002). Indeed, our research on three-term learning suggests that learning ability is a major component of fluid intelligence (Tamez et al., 2008; Williams & Pearlberg, 2006).

The specific roles played by learning ability in determining individual differences in performance on intelligence tests remain to be determined. Nevertheless, Williams and Pearlberg’s (2006) finding that only some types of learning tasks are significantly correlated with intelligence scores raises the possibility that it is the structure of what is to be learned that determines the strength of such correlations. For example, consider the structure of what must be learned on the three-term contingency task of Williams and Pearlberg: There are 10 stimulus words, each of which is associated with three different key-press responses, which in turn are each associated with 10 outcome words to be recalled. There presumably are also associations between stimulus words and outcome words in addition to the associations between stimulus–response combinations and outcome words. When a stimulus word and key press are specified, an individual performing the three-term task must contend with multiple competing associations. Out of all these various associations, only one unique stimulus–response–outcome combination is correct.

Individual differences in the ability to learn which word to recall in the face of such competing associations may well be what underlie differences on intelligence tests. Building on the idea that fluid intelligence tests also involve learning (e.g., Carlstedt et al., 2000; Verguts & De Boeck, 2002), we suggest that the reason that learning on the three-term task is related to performance on tests of fluid intelligence is because these fluid tests also involve learning in the face of stimuli competing for control, and the reason that learning ability is related to performance on tests of crystallized intelligence is that these crystallized tests assess the results of past learning in the face of such competition. In fact, learning in the presence of competing stimuli may be an important part of what glues the various components of g together and gives rise to the consistency of individuals’ behavior across different tests and in quite different situations.

Interestingly, Unsworth and Engle (2007) recently suggested that the ability to efficiently constrain searches of long-term memory is a critical aspect of working memory function, and in their view, this ability may underlie the correlation between working memory and intelligence. Although the terminology is quite different (e.g., attention control vs. stimulus control) and the emphasis is on using knowledge rather than on its acquisition, the view expressed by Unsworth and Engle is similar to our own. At this point in time, the preceding ideas are clearly hypotheses in need of further evaluation. However, they exemplify our belief that in order to shed light on what scores on intelligence tests mean, and why individuals show such consistent performance on the subtests of which these tests are composed, what will be needed is a clear determination of the critical dimensions that govern when tests of learning ability predict scores on intelligence tests, and vice versa.

FINAL COMMENTS

Astute application of behavioral principles may facilitate the development of expertise in specific learning situations, but it remains to be established whether general behavioral principles can provide insight into the fact that some people consistently perform better than others in situations that would be considered intellectually demanding. In principle, the observed covariance in performance level across very diverse tasks that characterizes g poses the same sort of issues as the covariance in different measures of other intervening variables. For example, a motivational construct like hunger refers to the fact that some people consistently perform better than others in situations that would be considered intellectually demanding. In principle, the observed covariance in performance level across very diverse tasks that characterizes g poses the same sort of issues as the covariance in different measures of other intervening variables. For example, a motivational construct like hunger refers to the fact that diverse food-related behaviors covary in strength. Similarly, Skinner’s (1969) definition of the operant entails that seemingly different movement patterns are the same response unit to the extent that they are functionally equivalent and thus covary in strength. Like these other integrative concepts, the covariance implicit in the concept of g can be studied productively using functional analysis to determine the natural lines of
fracture. Processing speed and working memory capacity, which are currently the predominant integrative constructs for explaining g, also are amenable to such a behavioral analysis. Indeed, new behavioral principles governing the absolute size of individual and age differences on processing speed tasks have already begun to emerge (Chen et al., 2007; Hale & Jansen, 1994; Myerson et al., 2003).

Recent findings relating intelligence scores to learning rate suggest that a behavior-analytic approach has great promise for understanding individual differences in intelligence. These findings present us with an opportunity to identify the specific features of learning that are most relevant to intelligent behavior. It remains to be seen how much of the general factor in intelligence scores can be explained by differences in learning, but given the importance of g for so much of everyday life (e.g., Gottfredson, 1997; Herrnstein & Murray, 1994; for a recent review, see Lubinski, 2004), behavior analysts surely will be motivated to undertake the relevant research. There is no justification for leaving the study of intelligence to others.

REFERENCES


Received: January 4, 2008
Final acceptance: May 5, 2008