Research on intelligence, dating back to Spearman’s 1904 article, “‘General Intelligence,’ Objectively Determined and Measured,” has been an area of keen interest to psychologists and the general public. Books such as Herrnstein and Murray’s *The Bell Curve* (1994) have created controversy, consternation, and commitment among different constituencies. Few areas of psychology—indeed few areas of scientific inquiry—have created such intense debate.

Intelligence, also called general mental ability (GMA) and cognitive ability, is of keen interest to industrial and organizational (I-O) psychology because it is an excellent predictor of two of the most important and often-studied variables: training proficiency and job performance. Dramatic gains in performance in training and on the job can result from using preemployment selection tests to identify and hire highly intelligent job applicants.

Psychologists and others who study human behavior are also interested in intelligence because it is related to many aspects of people’s lives. In addition to performance in the workplace, intelligence is related to academic performance (Kuncel, Hezlett, & Ones, 2001), occupational attainment (Jencks et al., 1979), many facets of everyday life (Gottfredson, 1997), health (Deary, Batty, & Gottfredson, 2005), and even mortality (Deary et al., 2005).

Because intelligence pervades so many aspects of human lives, it has been studied extensively. This chapter begins by reviewing factor analytic studies investigating the latent structure of intelligence. This line of research dates back to Spearman and is called the psychometric approach to the study of intelligence. Some of the most eminent and controversial psychologists of the 20th century have worked in this area, including Thurstone, Burt, Guilford, Thompson, Vernon, and Cattell. In a work of remarkable scholarship, John Carroll (1993) reanalyzed 461 correlation matrices from this literature using a single methodology to provide a coherent and compelling account of the factor analytic findings.

Information processing approaches to understanding intelligence constitute a second line of research summarized here. This work is characterized by carefully controlled experimental investigations of how people solve problems. In the psychometric literature, item responses are often aggregated up to subtest or total test scores prior to analysis; in contrast, information-processing research often decomposes item responding into more basic elemental components and processes to understand intelligence.

Neuropsychological approaches to the study of intelligence form a third area of research summarized in this chapter. Neuropsychology attempts to link the brain and behavior and thereby provide a deeper understanding of intelligence. Until recently, many of the most important findings in this area resulted from case studies of individuals with tragic brain damage. Advances in methods for imaging brain activity, such as functional magnetic resonance imaging (fMRI) and positron-emission tomography (PET), allow investigations of site-specific activation when individuals solve problems of a particular type. This research is exciting because it has the potential for connecting what is known about the latent structure of cognitive ability from psychometric research with the underlying hardware of the brain.
Intelligence and the Workplace 185

After summarizing the psychometric, information-processing, and neuropsychological approaches to understanding and explaining intelligence, the relation of intelligence and performance is described. Two lines of research are briefly summarized: laboratory studies of skill acquisition and meta-analytic studies summarizing correlations of intelligence with training and work performance.

Extensions of intelligence to social and emotional functioning are also reviewed. These types of intelligence—if they can properly be called intelligence—seem to have the potential for predicting and explaining at least some parts of a broadened criterion space. Finally, some common fallacies concerning intelligence are described. These fallacies have been highly persistent over time and resistant to empirical findings.

In sum, this chapter reviews psychometric approaches, information-processing models, and neuropsychological findings concerning intelligence as well as social and emotional intelligence. Although this chapter is primarily about intelligence, its discussion is framed by the enlarged criterion space that is of growing importance to I-O psychologists.

GENERAL MENTAL ABILITY

Psychometric Approaches to Intelligence

During the past century, the psychometric approach to intelligence has been the focus of a tremendous amount of research. Obviously, it is impossible to provide a comprehensive review of a century’s research in this chapter. More detail can be found in Carroll’s (1993) book, which provides a fascinating review, summarizing substantive findings, methodological advances, and the personal perspectives of key figures. In this chapter, the contributions of Spearman, Thurstone, Vernon, Guilford, Cattell, and Carroll are described.

Factor Fractionation

Before reviewing findings from the psychometric approach, it is important to highlight a point made by Truman Kelley in 1939 and often repeatedly by Lloyd Humphreys. Kelley stated that “evidence of existence of a factor [should] be not cited as evidence that it is important” in his famous “Mental Factors of No Importance” paper (Kelley, 1939, p. 141). Humphreys (1962) wrote that “test behavior can almost endlessly be made more specific, . . . factors [of intelligence] can almost endlessly be fractionated or splintered” (p. 475). With the advent of confirmatory factor analysis (CFA; Jöreskog, 1966) and convenient software implementations such as the LISREL computer program (Jöreskog & Sörbom, 1996), this problem has been exacerbated. In samples exceeding a few hundred, CFA can be likened to an electron microscope in that it can reliably determine the number of factors that are required to reproduce a correlation matrix, a number often substantially exceeding that expected on the basis of substantive theory.

How can researchers avoid extracting and interpreting “factors of no importance”? In factor analytic studies of test batteries of the sort pioneered by Thurstone (1938), there does not appear to be any way to differentiate substantively important factors from inappropriately splintered factors. Thus, research of a different kind is needed in which the pattern of relations with important criterion variables is examined. When a factor is fractionated, this research asks whether the newly split factors (a) correlate meaningfully with other important variables such as one or more of the dimensions of job performance, (b) exhibit a pattern of differential relations with such variables, and (c) increase our ability to understand and explain these variables. Vernon (1950) emphasized that “only those group factors shown to have significant practical value in daily life are worth incorporating in the picture” (p. 25). McNemar (1964), Lubinski and Dawis (1992, pp. 13–20), and Lubinski (2000) further elaborated on the pitfalls of factor fractionation and the importance of examining the scientific significance of factors.

For example, suppose a large sample completes an algebra test. It is likely that CFA could be used to demonstrate that a word-problem factor can be differentiated from a calculation factor (i.e., a factor determined from items that ask examinees to solve quadratic equations, solve two equations in two unknowns, etc.). Although statistically separable and likely to be correlated with performance on tasks requiring mathematical skill, the word-problem factor and the calculation factor would be highly correlated (probably in excess of 0.95), would have very similar correlations with other variables, and would not have a multiple correlation with any important criterion variable higher than the simple correlation of the original algebra test. Thus, there is little reason to fractionate the original algebra factor.

Spearman

Although Galton, Wundt, and others had studied intelligence previously, it is probably fair to say that contemporary theories of intelligence and corresponding
methodologies for research originated with Charles Spearman. Spearman was an Englishman who studied experimental psychology with Wundt. After completing his doctorate, Spearman returned to England and made many important contributions until his death in 1945.

Substantively, Spearman is best known for his two-factor theory of intelligence. Actually, this theory postulated two types of factors, not two factors. The first type is the general factor, which Spearman labeled \( g \), and the second type consists of specific factors. Spearman used the general factor as the explanation of why students’ grades in the classics were correlated with grades in other courses such as math and music. Indeed, much of Spearman’s research was directed to documenting the pervasive influence of the general factor. Specific factors were used to explain why performance in different domains had less than perfect correlations; performance in a given domain was influenced by general ability as well as domain-specific ability.

Spearman believed that general intelligence involved three fundamental processes, which he called the apprehension of experience, the eduction of relations, and the eduction of correlates. To educe means “to draw out; elicit” or “to infer from data; deduce” (Neufeldt, 1997, p. 432). The legacy of Spearman can be seen in the inductive and deductive reasoning factors found in Carroll’s (1993) reanalysis of cognitive ability correlation matrices.

Spearman also made important methodological contributions to the study of intelligence. In his 1904 paper, he examined the “hierarchy of the intelligences” (pp. 274–277) and provided a means for determining the “intellective saturation” of a variable, which was defined as the “extent to which the considered faculty is functionally identical with General Intelligence” (p. 276). These saturations are essentially factor loadings; later, Spearman introduced a method for computing the loadings on a single general factor (Hart & Spearman, 1912).

The law of tetrad differences (Carroll, 1993, attributes this term to a paper by Spearman & Holzinger, 1925) was introduced to test the two-factor model. Let \( r_{ij} \) denote the correlation between tests \( i \) and \( j \). Suppose the general factor is the sole reason that a set of variables have nonzero correlations and the loading of test \( i \) on the general factor is denoted \( \lambda_i \). Then the correlation \( r_{ij} \) should equal the product of \( \lambda_i \) and \( \lambda_j \) (plus sampling error). Consequently, for any four variables the tetrad difference,

\[
\text{Tetrad Difference} = r_{13}r_{24} - r_{23}r_{14} = (\lambda_1\lambda_3)(\lambda_2\lambda_4) - (\lambda_2\lambda_3)(\lambda_1\lambda_4)
\]

should differ from zero only due to sampling error. Investigating tetrad differences, to which Spearman devoted great effort, is akin to the modern analysis of residuals. Computer programs such as LISREL (Jöreskog & Sörbom, 1996) provide a matrix of residuals, which are obtained by subtracting the matrix of correlations reproduced on the basis of the parameters estimated for a hypothesized model from the original correlation matrix.

As described later, subsequent researchers have developed models of intelligence that incorporate additional factors. In fact, Spearman’s focus on a single ability may seem odd because there are measures of so many different abilities currently available. To provide a perspective for Spearman’s interest in a single dominant ability (and to illustrate later theories of intelligence), it is instructive to consider the correlations among a set of cognitive ability tests. Table 8.1 presents the correlations of 10 subtests that constituted the Armed Services Vocational Aptitude Battery (ASVAB) along with their internal consistency reliabilities. These correlations, provided by Ree, Mullins, Mathews, and Massey (1982), were obtained from a large sample (2,620 men) and have been corrected to estimate the correlations that would have been obtained from a nationally representative sample.

The ASVAB subtests assess a rather wide range of abilities. Arithmetic Reasoning and Math Knowledge measure quantitative reasoning; Word Knowledge and Paragraph Comprehension assess verbal ability; General Science is largely a measure of science vocabulary; Auto-Shop Information, Mechanical Comprehension, and Electronics Information assess technical knowledge required for increasingly sophisticated military occupational specialties; and Numerical Operations and Coding Speed assess very simple skills (e.g., \( 7 + 9 = ? \)), albeit in a highly speeded context. Although it is not surprising that the quantitative reasoning tests correlate highly (\( r = 0.79 \)) and the verbal tests correlate highly (\( r = 0.82 \)), the magnitude of the quantitative–verbal correlations is surprisingly large (RS between .60 and .70). Indeed, the quantitative–verbal correlations are only about 0.10 to 0.20 smaller than are the within-trait correlations. Moreover, the technical tests have remarkably high correlations with the verbal and quantitative skills (e.g., Word Knowledge correlates 0.67 with Mechanical Comprehension), and even the speeded tests have sizable correlations with the power tests (all correlations greater than 0.40).

Table 8.2 contains the factor loadings obtained when a single common factor (i.e., Spearman’s two-factor model) is fit to the ASVAB correlation matrix using maximum likelihood estimation as implemented in LISREL.
TABLE 8.1 Correlation Matrix of ASVAB Form 8A Subtests

<table>
<thead>
<tr>
<th>Subtest</th>
<th>AR</th>
<th>MK</th>
<th>WK</th>
<th>PC</th>
<th>GS</th>
<th>AS</th>
<th>MC</th>
<th>EI</th>
<th>NO</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic Reasoning (AR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Knowledge (MK)</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Knowledge (WK)</td>
<td>0.70</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.92)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paragraph Comprehension (PC)</td>
<td>0.70</td>
<td>0.60</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Science (GS)</td>
<td>0.71</td>
<td>0.65</td>
<td>0.83</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto-Shop Information (AS)</td>
<td>0.60</td>
<td>0.52</td>
<td>0.68</td>
<td>0.63</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanical Comprehension (MC)</td>
<td>0.69</td>
<td>0.64</td>
<td>0.67</td>
<td>0.64</td>
<td>0.71</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronics Information (EI)</td>
<td>0.68</td>
<td>0.61</td>
<td>0.76</td>
<td>0.69</td>
<td>0.78</td>
<td>0.79</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerical Operations (NO)</td>
<td>0.59</td>
<td>0.58</td>
<td>0.52</td>
<td>0.55</td>
<td>0.48</td>
<td>0.40</td>
<td>0.45</td>
<td>0.46</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Coding Speed (CS)</td>
<td>0.52</td>
<td>0.51</td>
<td>0.48</td>
<td>0.49</td>
<td>0.43</td>
<td>0.42</td>
<td>0.45</td>
<td>0.46</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

(Note: Internal consistency reliabilities (KR-20) appear in the diagonal within parentheses; internal consistency reliabilities were not computed for speeded tests.

TABLE 8.2 Factor Loadings and Residuals for Spearman’s “Two-Factor” Model Fitted to the ASVAB

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Factor Loadings</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR</td>
<td>MK</td>
</tr>
<tr>
<td>Arithmetic Reasoning (AR)</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Math Knowledge (MK)</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>(0.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Knowledge (WK)</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>(−0.03)</td>
<td>−0.05</td>
<td></td>
</tr>
<tr>
<td>Paragraph Comprehension (PC)</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>(0.01)</td>
<td>−0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>General Science (GS)</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>(−0.02)</td>
<td>−0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Auto-Shop Information (AS)</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>(−0.06)</td>
<td>−0.08</td>
<td>−0.02</td>
</tr>
<tr>
<td>Mechanical Comprehension (MC)</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>(0.01)</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>(−0.05)</td>
<td>−0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Electronics Information (EI)</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>(−0.04)</td>
<td>−0.04</td>
<td>−0.01</td>
</tr>
<tr>
<td>Numerical Operations (NO)</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>(0.09)</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>(−0.02)</td>
<td>−0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Coding Speed (CS)</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>(0.05)</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>(−0.02)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>(−0.07)</td>
<td>−0.03</td>
<td>−0.01</td>
</tr>
</tbody>
</table>

(Jöreskog & Sörbom, 1996). Table 8.2 also contains the residuals. Residuals are obtained by using the estimated factor loadings to compute the fitted correlation matrix (i.e., the correlations expected from the estimated factor loadings). The fitted correlations are then subtracted from the observed correlations to produce the residuals. For example, Table 8.2 shows that the factor loadings of Arithmetic Reasoning and Math Knowledge were estimated to be 0.83 and 0.75. For this single common factor model, the expected correlation is therefore 0.83 × 0.75 = 0.62. The fitted correlation is then subtracted from the actual correlation, 0.79 − 0.62, to obtain a residual of 0.17, which is shown in Table 8.2.

As reported in Table 8.2, all of the tests have large loadings; the two speeded subtests have loadings of about 0.6, whereas the eight power tests have loadings of about 0.8. Note the large positive residuals between Arithmetic Reasoning and Math Knowledge and between Numerical Operations and Coding Speed and the more moderate positive residuals among the three technical tests. The correlations among the three verbal tests have been reasonably well modeled (residuals of 0.08, 0.05, and 0.00) by estimating their loadings as quite large (0.89, 0.84, and 0.88). Thus, the general factor in this solution appears strongly related to verbal ability, with mathematical and technical abilities also highly related.

Fit statistics for the solution shown in Table 8.2 indicate substantial problems. The root mean squared error of approximation (RMSEA; Steiger, 1990) is 0.19; the adjusted goodness of fit statistic is 0.67; and the non-normed fit index is 0.83. All three of these indices, as well as the matrix of residuals, indicate that Spearman’s two-factor model is unable to account for the correlations among the ASVAB subtests. Instead, a consideration of the content of the subtests suggests that four factors are required to describe adequately the correlations in Table 8.1 (i.e., factors representing quantitative, verbal, technical, and speed abilities).

Nonetheless, it is clear that a single general factor explains much of the association seen in Table 8.1. In fact, Spearman’s response to the residuals in Table 8.2 may well have been “swollen specifics.” That is, Spearman might have argued that including two measures of a single skill (e.g., Arithmetic Reasoning and Math Knowledge) in a test battery causes the quantitative specific factor falsely to appear to be a general factor.
TABLE 8.3 Factor Loadings and Residuals for Four Correlated Factors Fitted to the ASVAB

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Factor Loadings</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q</td>
<td>V</td>
</tr>
<tr>
<td>AR</td>
<td>0.93</td>
<td>—</td>
</tr>
<tr>
<td>MK</td>
<td>0.85</td>
<td>—</td>
</tr>
<tr>
<td>WK</td>
<td>0.92</td>
<td>—</td>
</tr>
<tr>
<td>PC</td>
<td>0.86</td>
<td>0.03</td>
</tr>
<tr>
<td>GS</td>
<td>0.90</td>
<td>—</td>
</tr>
<tr>
<td>AS</td>
<td>0.91</td>
<td>—</td>
</tr>
<tr>
<td>MC</td>
<td>0.85</td>
<td>0.06</td>
</tr>
<tr>
<td>EI</td>
<td>0.85</td>
<td>—</td>
</tr>
<tr>
<td>NO</td>
<td>0.85</td>
<td>—</td>
</tr>
<tr>
<td>CS</td>
<td>0.75</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: Q = quantitative; V = verbal; T = technical; S = speed. Omitted factor loadings were fixed at zero.

Thurstone

Louis Leon Thurstone’s *Primary Mental Abilities* (1938) monograph stands as a landmark in the study of intelligence. A total of 218 college students completed 56 tests during five 3-hour sessions. The tests were carefully selected, and detailed descriptions of the items were provided in the monograph. A dozen factors were extracted and rotated, and seven primary factors were clearly interpretable: spatial, perceptual, numerical, verbal relations, word fluency, memory, and inductive reasoning.

In his study of cognitive abilities, Thurstone made many methodological innovations that contributed to the development of factor analysis. These innovations, developed over a period of years, were summarized in his *Multiple Factor Analysis* (Thurstone, 1947) text, which more than a half-century later continues to provide a remarkably lucid account of factor analysis. Central to his approach was the use of multiple factors, interpretable due to the “simple structure” of factor loadings, to explain the correlations among a set of tests. To obtain these interpretable factors in an era when calculations were performed by hand, Thurstone devised a computationally simple method for extracting factors. He clearly articulated the distinctions between common variance, specific variance, and error variance and provided means to estimate a variable’s communality (i.e., its common variance). When factors are extracted according to algebraic criteria (e.g., Thurstone’s centroid method or principal axes), Thurstone maintained that the resulting factor loading matrix is not necessarily psychologically meaningful. Consequently, he developed orthogonal and oblique rotation methods to facilitate interpretation. Simple structure, which Thurstone used to guide rotation, is now used as the principal model for the relation of latent (the factors) and manifest (i.e., the tests) variables.

For a battery of psychological tests, it is ordinarily impossible to obtain simple structure when the latent variables are required to be uncorrelated. For this reason, Thurstone introduced the idea of correlated factors and used such factors when rotating to simple structure. In LISREL terminology, Thurstone treated his tests as manifest variables (Xs) and used exogenous latent factors (ξs) to explain the correlations among the manifest variables. The results of this analysis are a factor loading matrix (Λx in LISREL notation) and a matrix (Φ) of factor correlations. Table 8.3 provides the factor loading matrix and residuals obtained by using LISREL to fit four correlated factors to Table 8.1; the factor correlations are given in Table 8.4.

Fitting four correlated factors to the ASVAB correlations shown in Table 8.1 is much more satisfactory. The RMSEA is 0.093; the adjusted goodness of fit is 0.90; and the nonnormed fit index is 0.95.

In this formulation of factor analysis, a general factor is not needed to describe the pervasive relations between manifest variables (and will not emerge in a factor analysis if Λx is specified to show simple structure) because the factor correlations in Φ explicitly model the associations of the latent variables. Note that the factor correlations shown in Table 8.4 are all large and positive. Interestingly, Carroll (1993) noted that “an acrimonious controversy between Spearman and his ‘British’ school, on the one

---

### Table 8.4 Correlations of Four Factors Fitted to the ASVAB

<table>
<thead>
<tr>
<th>Factor</th>
<th>Q</th>
<th>V</th>
<th>T</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative (Q)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Verbal (V)</td>
<td>0.83</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Technical (T)</td>
<td>0.80</td>
<td>0.90</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Speed (S)</td>
<td>0.76</td>
<td>0.68</td>
<td>0.62</td>
<td>—</td>
</tr>
</tbody>
</table>
hand, and Thurstone and his ‘American’ school, on the other” (p. 56) arose about the existence of a general factor. Carroll feels “fairly certain that if Spearman had lived beyond 1945, it would have been possible for him and Thurstone to reach a rapprochement” (p. 56).

It was not until 1957 that Schmid and Leiman showed the algebraic equivalence of correlated primary factors and a representation with a second-order general factor and orthogonal first-order factors. When viewed from the perspective of structural equation modeling, it is easy to see that the debate between advocates of a general factor and advocates of correlated primary factors was pointless. When $\Phi$ contains many large positive correlations between factors, the question is not whether a general factor exists but rather whether a single general factor can account for the factor correlations. To examine this question within the LISREL framework, the tests can be taken as endogenous manifest variables ($Y$’s); primary factors are taken as endogenous latent variables ($\eta$’s); and the issue is whether paths (in the $\Gamma$ matrix) from a single exogenous latent factor ($\xi$, i.e., the general factor) to each $\eta$ can account for the correlations between tests loading on different factors. With a large battery of the sort analyzed by Thurstone (1938), more than a single general factor may be required to model adequately the correlations in the $\Phi$ matrix.

Fitting this model to the ASVAB data yields estimates of paths from the second-order general factor $\xi$ to the endogenous Quantitative, Verbal, Technical, and Speed factors of 0.88, 0.96, 0.92, and 0.73. The factor-loading matrix $\Lambda_{v}$ is virtually identical to the factor-loading matrix ($\Lambda_{s}$) shown in Table 8.3. The residuals are also similar, except that rather large residuals remain between the Quantitative subtests and Speed subtests. For example, the residual between Math Knowledge and Numerical Operations was 0.13. Consequently, the fit statistics dropped slightly: the RMSEA is .11; the adjusted goodness of fit is 0.88; and the nonnormed fit index is 0.94. These results clearly show that the issue is not whether a general factor exists, but instead whether a model with a single general factor can account for the correlations among Thurstonian primary factors. The models described by Vernon (1950) and Carroll (1993) suggest that for large batteries of tests that sample diverse abilities the answer will ordinarily be negative.

**Vernon**

Philip E. Vernon, a junior colleague of Spearman, developed a model that addressed the main weakness of his senior mentor. Specifically, the law of tetrad differences fails for the correlation matrix presented in Table 8.1 and for almost any test battery unless the tests have been very carefully selected so that their tetrad differences vanish. A theory of intelligence that satisfactorily describes only some (very carefully selected) sets of tests is not satisfactory, and Spearman was criticized for this problem.

Vernon (1950) acknowledged that “almost any specific factor (in Spearman’s sense) can be turned into a primary factor, given sufficient ingenuity in test construction” (p. 133) and warned against “highly specialized factors, which have no appreciable significance for everyday life [and] are not worth isolating” (p. 133). Such factors are sometimes called eye twitch factors (Charles L. Hulin, personal communication, August 21, 1977). Instead, Vernon argued that “factorists should aim not merely to reduce large numbers of variables to a few components that account for their intercorrelations, but also to reduce them to the fewest components which will cover most variance” (p. 133).

To this end, Vernon (1950) developed the hierarchical group-factor theory of intelligence illustrated in Figure 8.1. At the apex is general intelligence, $g$, which Vernon suggested would account for about 40% of the variance in the scores of a test battery. Vernon used $v:ed$ and $k:m$ to denote two “major group factors,” which collectively might explain approximately 10% of the variance in test scores. The construct $v:ed$ refers to a verbal-educational higher order factor, which explains the relations among reading comprehension, logical reasoning, and arithmetic reasoning after partialling out $g$, and $k:m$ refers to a major group factor defined by spatial and mechanical abilities. Vernon believed the minor group factors (reading comprehension, logical reasoning, spatial ability, etc.) explained about 10% of the variance in test scores, and he attributed the remaining 40% to specific factors and error of measurement.

Vernon’s model in LISREL notation appears very different from Thurstone’s simple structure. As shown in Table 8.3, each test loads on just one factor in an ideal simple structure. In Vernon’s model, each test would load on the general factor $g$ (denoted as $\xi_{1}$ in LISREL notation); $v:ed$ and $k:m$ would be latent variables ($\xi_{2}$ and $\xi_{3}$); the $m$ minor group factors would be latent variables denoted $\xi_{4}$ to $\xi_{m+3}$; and all latent variables would be uncorrelated. A test hypothesized to assess the first minor group factor within the $v:ed$ domain would have loadings estimated on three factors: $\xi_{1}$, $\xi_{2}$, and $\xi_{4}$ (assuming that the first minor group factor was denoted as the fourth factor). Although the factors in Table 8.5 are labeled according to Carroll’s (1993) conceptualization, they illustrate the pattern of
Table 8.5: Carroll's Factor Loadings for a Correlation Matrix Published by Schutz (1958) after Schmid–Leiman Transformation

<table>
<thead>
<tr>
<th>Test</th>
<th>g</th>
<th>Gc</th>
<th>Gf</th>
<th>Verbal</th>
<th>Numerical</th>
<th>Space</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Meaning</td>
<td>0.56</td>
<td>0.43</td>
<td>-0.01</td>
<td>0.53</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Odd Words</td>
<td>0.62</td>
<td>0.44</td>
<td>0.02</td>
<td>0.50</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Remainders</td>
<td>0.53</td>
<td>0.22</td>
<td>0.18</td>
<td>-0.01</td>
<td>0.64</td>
<td>0.04</td>
<td>-0.01</td>
</tr>
<tr>
<td>Mixed Arithmetic</td>
<td>0.56</td>
<td>0.25</td>
<td>0.16</td>
<td>0.02</td>
<td>0.62</td>
<td>-0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Hatchets</td>
<td>0.50</td>
<td>0.01</td>
<td>0.35</td>
<td>0.01</td>
<td>0.01</td>
<td>0.58</td>
<td>0.00</td>
</tr>
<tr>
<td>Boots</td>
<td>0.49</td>
<td>0.00</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
<td>0.00</td>
</tr>
<tr>
<td>Figure Changes</td>
<td>0.60</td>
<td>0.18</td>
<td>0.27</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.27</td>
</tr>
<tr>
<td>Mixed Series</td>
<td>0.65</td>
<td>0.21</td>
<td>0.26</td>
<td>-0.02</td>
<td>0.07</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>Teams</td>
<td>0.53</td>
<td>0.21</td>
<td>0.18</td>
<td>0.05</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Note: Salient loadings are bolded. g = general intelligence; Gc = crystallized intelligence; Gf = fluid intelligence.

Source: Adapted from Carroll (1993, p. 95).

An interesting effect is that if the loadings expected to be small in Table 8.5 are fixed at zero and the bolded loadings are treated as parameters to be estimated, a program such as LISREL is unable to obtain a maximum likelihood solution. Without further constraints, such a pattern of fixed and free loadings is underidentified (McDonald, personal communication, December 1, 2000). McDonald (1999, pp. 188–191) describes the constraints that must be implemented for factor loadings to be estimable. LISREL 8.30 does not allow such constraints; instead, CALIS (SAS Institute, 1990) can be used. The prepotency of g in Vernon’s model nicely explains the large correlations among all variables seen in Table 8.1. The quantitative, verbal, technical, and speed factors apparent in Table 8.1 would correspond to minor group factors in Vernon’s model and, as expected, clearly explain much less variance. The v:ed and k:m major group factors are not obvious in Table 8.1, presumably because the ASVAB battery of tests is too limited in scope.

Guilford

Factor fractionation was taken to an extreme in J. P. Guilford’s (1967, 1985) structure of intellect (SOI) model. Guilford factorially crossed contents (i.e., the type of information processed) with operations (i.e., the mental activity or process applied to the content) and products (i.e., the output of the operation) to arrive at SOI abilities. Contents included visual, auditory, symbolic, semantic, and behavior categories; operations included evaluation, convergent production, divergent production, memory, and cognition; and products included units,
classes, relations, systems, transformations, and implications (Guilford, 1967, 1985). This three-way classification can be represented as a cube with 5 rows, 6 columns, and 5 slabs, for a total of 150 primary abilities.

Guilford spent much of his career developing multiple measures of the various abilities defined by the SOI cube. Great energy and effort was devoted to this program of research. Carroll (1993) noted that “Guilford must be given much credit for conducting a series of major factorial studies in which hypotheses were to be confirmed or disconfirmed by successive studies in which new tests were continually designed to permit such testing of hypotheses” (p. 58).

However, there is much to criticize. For example, Guilford wrote that “any genuine zero correlations between pairs of intellectual tests is sufficient to disprove the existence of a universal factor like g” (1967, p. 56) and that of “some 48,000 correlations between pairs of tests, about 18% were below 0.10, many of them being below zero” (1985, p. 238). The problem with Guilford’s argument is that eye-twitch factors are unlikely to correlate with other eye-twitch factors, so zero correlations between measures of obscure abilities are neither surprising nor particularly meaningful. Moreover, as noted previously, an important desideratum in evaluating psychometric factors is their practical significance. Research with broad abilities such as the ASVAB’s verbal, quantitative, and technical abilities has found that they add little incremental validity to that provided by g when predicting training performance (Ree & Earles, 1991) and job performance (Ree, Earles, & Teachout, 1994); it appears unlikely that the factors identified by Guilford would meet with more success.

A more fundamental criticism of the SOI model lies in its factorial combination of content, operation, and product to characterize human abilities. There is no a priori reason why the mind should be well described by factorially crossing these three factors. Indeed, new statistical methodologies such as hierarchical regression trees (Breiman, Friedman, Olshen, & Stone, 1984) and neural networks (Freeman & Skapura, 1992) suggest the need for nonlinear approaches to understanding complex phenomena.

Cattell
Raymond B. Cattell was a student of Spearman in the 1930s (Carroll, 1993) and spent most of his career at the University of Illinois at Urbana–Champaign. In addition to his academic appointment, Cattell also founded the Institute for Personality and Ability Testing (IPAT) and made numerous contributions to the study of personality.

Cattell (1971) described a variety of influences that led to his (1941, 1943) notions of fluid and crystallized intelligence, often denoted Gf and Gc. Among these were his consideration of the correlations of Thurstone’s (1938) primary factors, which he felt revealed more than one general factor, as well as the different kinds of abilities assessed by culture-fair tests (i.e., perceptual) and traditional intelligence tests (e.g., verbal comprehension).

Cattell (1971) wrote that “fluid intelligence shows itself in successfully educating complex relations among simple fundamentals whose properties are known to everyone” and that Gf “appears to operate whenever the sheer perception of complex relations is involved” (p. 98). Thus, Gf reflects basic abilities in reasoning and related higher mental processes (e.g., inductive reasoning). On the other hand, crystallized intelligence reflects the extent of an individual’s base of knowledge (vocabulary, general information). Cattell wrote that this crystallized intelligence operates “in areas where the judgments have been taught systematically or experienced before” (p. 98).

Cattell (1971) described an interesting mechanism that explains why cognitive ability tests have large positive correlations. Cattell suggested that individuals are born with “a single, general, relation-perceiving ability connected with the total, associational, neuron development of the cortex” (p. 117). This ability is what Cattell viewed as fluid intelligence. Through experience, individuals learn facts, relationships, and techniques for solving problems. This pool of acquired knowledge, which depends on opportunity to learn, motivation, frequency of reward, and so forth, is what Cattell viewed as crystallized knowledge. Cattell’s investment theory hypothesizes that “as a result of the fluid ability being invested in all kinds of complex learning situations, correlations among these acquired, crystallized abilities will also be large and positive, and tend to yield a general factor” (p. 118). However, correlations of measures of fluid and crystallized intelligence will not be perfect because of the various other factors affecting crystallized intelligence.

Carroll
John B. Carroll (1993) conducted a massive review and reanalysis of the factor analytic literature. He first compiled a bibliography of more than 10,000 references and identified approximately 1,500 “as pertaining to the correlational or factor analysis of cognitive abilities” (p. 78). Ultimately, 461 data sets were selected on the basis of being well suited to factor analysis (e.g., at least three tests were included as measures of each factor that was
hypothesized; a reasonable representation of factors was included; the sample of individuals was broad).

One of the problems in comparing factor analytic results from different researchers lies in their use of different statistical methods. The seriousness of this problem can be seen in the acrimonious debate between the British and American researchers. To allow valid comparisons across studies, Carroll (1993) used a single, consistent methodology, which he carefully described in his book (pp. 80–101). Exploratory factor analysis (EFA) provided the fundamental basis for Carroll’s analysis.

Carroll decided to use EFA to “let the data speak for themselves” (p. 82). Because EFA results are often unstable and sampling variability can play an unacceptably large role in samples of moderate size (i.e., a few hundred; Idaszak, Bottom, & Drasgow, 1988), CFA has largely replaced EFA. However, CFA requires the researcher to specify, prior to beginning the analysis, the pattern of fixed (at zero) and free (to be estimated) factor loadings as well as any higher order structure. Thus, to use CFA to reanalyze, say, Thurstone’s (1938) correlation matrix, the researcher would need to specify the pattern of fixed and free loadings for tests such as Block-counting, Lozenges, and Flags. The contents of such tests are not apparent from their names, and the traits they assess are not obvious. Of course, careful consideration of the contents of each test would allow tentative hypotheses to be made, but application of CFA to all of Carroll’s 461 sets of tests would have been incredibly difficult and impossibly time consuming. Consequently, EFA was the only viable option for this massive reanalysis.

Carroll’s analysis included some of the most reliable and trustworthy procedures developed in the long history of EFA. For example, the number of factors was determined in part by Montanelli and Humphreys’s (1976) parallel analysis, which compares the eigenvalues of a correlation matrix (with squared multiple correlations on the diagonal) to the eigenvalues of a correlation matrix for random data simulating the same number of people and variables. The parallel analysis criterion suggests extracting a factor only when the eigenvalue of the real data exceeds the corresponding eigenvalue of the random data.

Varimax (Kaiser, 1958) was used for orthogonal rotation, and Tucker and Finkbeiner’s (1981) direct artificial personal probability function rotation (DAPPF) was used for oblique rotation; in my experience, these rotation methods are the best available. When DAPPF produced correlated first-order factors (which Carroll reports was usually the case), the resulting factor correlation matrix was factor analyzed to produce second-order factors. When the second-order factors were also correlated, a third-order factor analysis was performed; no higher order analysis was needed (Carroll, 1993, p. 89). When second-order or third-order factors were obtained, Carroll performed a Schmid–Leiman (1957) transformation.

Carroll (1993) noted that the “Schmid-Leiman transformation can be thought of as one that redistributes variances from correlated factors to orthogonal factors” (p. 90) and demonstrates the equivalence of Thurstonian correlated factors with Vernon’s hierarchical representation. When a test battery allowed a third-order analysis, each test obtained a loading on the third-order factor, loadings on each second-order factor, and loadings on each first-order factor. Table 8.5, adapted from Carroll’s (1993, p. 95) Table 3.2, illustrates the end result of a reanalysis. Note that all nine tests have sizable loadings on the general factor g; four tests have moderate-sized loadings on crystallized intelligence Gc; five tests have moderate loadings on fluid intelligence Gf; and each test has a loading on its first-order common factor.

Reminiscent of Vernon’s (1950) hierarchical model shown in Figure 8.1, Carroll’s (1993) three-stratum model is shown in Figure 8.2. At the apex is general cognitive ability. Whereas Vernon had two broad factors (v:ed and k:n) at the second level, Carroll obtained many more; eight of the most important are shown in Figure 8.2, and several others appear in Carroll’s Table 15.14 (pp. 620–622). Following Carroll (see p. 625), the distance between g and each second-order factor (e.g., Gf) in Figure 8.2 reflects the approximate strength of relationship, with shorter distances indicating stronger association. Table 8.6 lists some of the first-order factors that define the second-order factors.

The second-order factor most strongly related to g is fluid intelligence, Gf. It is defined by the first-order factors of induction, deduction, and quantitative reasoning. Carroll (1993) stated that it is “concerned with the basic processes of reasoning and other mental activities that depend only minimally on learning and acculturation” (p. 624).

Also closely related to g is crystallized intelligence, Gc. Carroll (1993) found many first-order factors related to Gc, including verbal ability, reading comprehension, and lexical knowledge. From the first-order factors that Carroll found to be related to Gc, this factor could have been labeled Verbal Ability. Cattell’s (1971) investment theory would predict a much wider array of first-order factors lying beneath Gc, including perhaps science knowledge,
Adapted from Carroll (1993, p. 626).

Figure 8.2  Carroll’s three-stratum theory of intelligence

TABLE 8.6  Some First-Order Factors Identified by Carroll (1995)

<table>
<thead>
<tr>
<th>Second-Order Factor</th>
<th>First-Order Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Power Factor</td>
</tr>
<tr>
<td>Gf</td>
<td>Deduction (0.41)</td>
</tr>
<tr>
<td>Gc</td>
<td>Verbal ability (0.49)</td>
</tr>
<tr>
<td></td>
<td>Memory span (0.36)</td>
</tr>
<tr>
<td>Visual Perception</td>
<td>Visualization (.57)</td>
</tr>
<tr>
<td>Auditory Perception</td>
<td>Hearing threshold</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Originality (0.40)</td>
</tr>
<tr>
<td>Cognitive Speed</td>
<td></td>
</tr>
<tr>
<td>Processing Speed</td>
<td></td>
</tr>
</tbody>
</table>

Note: Median loadings of tests on the third-order g are provided in parentheses when available.

mechanical knowledge, and knowledge of other subjects taught in high school. A general-knowledge first-order factor did occasionally appear under Gc.

Actually, the empirical distinction between Gf and Gc was not sharp and clear in several data sets. Carroll (1993) obtained a second-order factor in some cases that was a combination of the first-order factors that usually define Gf and Gc, such as verbal ability, deduction, and quantitative reasoning. This combination may be the result of inadequately designed test batteries and the vagaries of sampling. It would be interesting to use CFA methods on these data sets to determine whether a latent structure that makes a sharp distinction between Gf and Gc first-order factors fits significantly worse than does the combination structure obtained by Carroll.

Carroll (1993) also identified a second-order memory factor. First-order factors lying beneath this second-order factor include memory span, associative memory,
free recall memory, and meaningful memory. Carroll suggested that the latent structure of memory has been understudied, noting that “our database does not include enough information to clarify the true structure of memory and learning abilities at higher strata” (p. 605). In their paper “Reasoning Ability Is (Little More Than) Working-Memory Capacity?” Kylonen and Christal (1990) certainly argued for the importance of memory, but Carroll found the median loading of memory span factors on g to be a less promising 0.36. The distinction between short-term memory and working memory (Engle, Tuholski, Laughlin, & Conway, 1999)—working memory involves Baddeley’s (1986) central executive—appears to be the critical distinction.

Visual perception is another second-order factor obtained by Carroll (1993). First-order factors defining visual perception include, among others, visualization, spatial relations, closure speed, and flexibility of closure. Some of these first-order tests had relatively high median loadings on g (e.g., .57 for visualization, .45 for flexibility of closure, and .42 for closure speed), suggesting that some of these item types should be included in a broad test of general cognitive ability.

A rather small number of studies have investigated auditory perception, but Carroll was nonetheless able to identify a second-order factor for this domain. Prior to the widespread availability of multimedia computers, research investigating auditory perception had been more difficult than factorial studies of abilities that can be assessed via paper-and-pencil tests. Multimedia computerized tests of musical aptitude (Vispoel, 1999) and other auditory abilities can now be easily developed and administered, so research in this area is warranted.

Carroll (1993) found a second-order retrieval factor, which he described as the “capacity to readily call up concepts, ideas, and names from long-term memory” (p. 612). The first-order factor found most often beneath retrieval was ideational fluency. Tests used to assess this construct require examinees rapidly to list exemplars of some category. For example, examinees might be given three minutes to write as much as they can about a given theme, identify objects that are round, or enumerate things that might happen on an airplane trip. Another first-order factor in this domain is word fluency, which can be assessed by tests that give examinees a few minutes to list words that begin with the letter R, make as many anagrams as possible from a given word, unscramble words (e.g., “rabvle” is “verbal”), and so forth. Carroll found both ideational fluency and word fluency factors to have fairly large median loadings (0.38 and 0.43, respectively) on g.

The final two second-order factors shown in Figure 8.2 are cognitive speed and processing speed. First-order factors underlying cognitive speed include perceptual speed (which also sometimes appears under the second-order visual perception factor), numerical facility (e.g., the ASVAB Numerical Operations test), and the rate of test taking. The second-order processing speed factor includes first-order factors such as simple reaction time, choice reaction time, and semantic processing speed. The distance of these second-order factors from g in Figure 8.2 indicates their relatively weak association with general cognitive ability.

Summary and Critique of the Psychometric Approach

Humphreys (1984, p. 243) defined intelligence as an individual’s “entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that [is] available at any one period of time.” Factor analytic research has carefully analyzed the latent structure of this repertoire of knowledge, skills, and problem-solving strategies; the most important finding lies in the tremendously important general ability g. A handful of second-order factors are also necessary to model correlation matrices that show patterns of first-order factors more highly associated than expected on the basis of a single general factor. Thus, Gf, Gc, memory, visual perception, auditory perception, retrieval, and cognitive speed factors are required for adequately describing the broad structure of the repertoire. Countless first-order factors can be obtained, but they seem unlikely to explain additional variance in important workplace behaviors. Instead, their main use lies in helping to define and understand the higher order factors.

Cattell (1941) proposed investment theory to explain how crystallized skills and abilities develop over the life span. He envisioned Gf as one’s fundamental reasoning capability and believed that Gc grew as a function of one’s fluid intelligence and investment of time and energy. Of relevance to this conceptualization is Tuddenham’s (1948) comparison of White enlisted men’s intelligence in World Wars I and II. Using the Army Alpha test of intelligence, Tuddenham reported a gain of about one standard deviation in test scores over this period. Such an increase in scores is difficult to explain if the Army Alpha test is thought to assess fundamental reasoning capacity. The World War II men averaged about two years more education than the earlier sample (Tuddenham, 1948), so the increase can be explained if Humphreys’s definition of intelligence as a repertoire is used.
Flynn’s (1984, 1987) research is also relevant. The Flynn effect refers to large gains in intelligence test scores over time. Flynn compiled longitudinal results from 14 nations for tests with “culturally reduced content” that assess “decontextualized problem solving” (i.e., tests that generally fit better into the Gf category) and tests with greater verbal content (i.e., fitting better into the Gc category). Flynn (1987, p. 185) found “strong data for massive gains on culturally reduced tests,” and, for nations where such comparisons were possible, “gains on culturally reduced tests at twice the size of verbal gains.”

Thus, the view of Gf as one’s inherent reasoning ability is inconsistent with Flynn’s data (if we are willing to assume that there has not been a major change in the gene pool in the past half-century). Instead, Flynn’s findings appear to be more consistent with Humphreys’s (1984) definition of intelligence as an individual’s repertoire of knowledge, skills, and problem-solving strategies.

In addition to education, test scores can be affected by coaching. It is important to note that item types vary in their susceptibility to coaching. For example, it is difficult to develop effective coaching strategies for some item types (Messick & Jungeblut, 1981). On tests of verbal ability that use a synonyms or antonyms format, students must substantially increase their vocabulary to raise test scores, which is a very difficult task. Messick and Jungeblut reported that SAT–Verbal scores increase linearly with the logarithm of time spent studying; based on a variety of regression equations, they predicted a 7-point gain for 10 hours of SAT-V preparation, a 20- to 25-point gain for 100 hours of study, and a 30- to 35-point gain for 300 hours. Flynn’s (1984, 1987) results, however, suggest that tests with culturally reduced content are more coachable. Specifically, the simplest explanation for Flynn’s findings is that one can learn problem-solving strategies that substantially increase scores on culturally reduced tests. Indeed, test developers should conduct coachability studies of new item types to ensure that they are resistant to easily learned strategies for answering items.

Flynn (1987) concluded that “psychologists should stop saying that IQ tests measure intelligence” (p. 188). If we accept Humphreys’s definition, then Flynn’s results can be interpreted as providing compelling evidence that intelligence tests do measure intelligence, but some test formats are more coachable than others (i.e., scores are more affected by problem-answering strategies).

Flynn (1987) defined intelligence as “real-world problem-solving ability” (p. 188), a definition quite different from Humphreys’s definition. What would a test of real-world problem solving look like? Test development for such an assessment instrument could begin with interviews of a fairly large number of people who would be asked to describe situations where they had to solve important problems; this is essentially Flanagan’s (1954) method of critical incidents. Test development could then follow the approach described by Motowidlo, Dunnette, and Carter (1990) and used by Olson-Buchanan et al. (1998). The resulting test would likely be viewed as a situational judgment test and look much more like the assessments described later in this section on social and emotional intelligence; the test would not appear similar to the tests that usually define first-order factors beneath Gf, Gc, or memory.

A different sort of criticism of the psychometric approach is made by those who wish to have an explicit definition of intelligence. Many psychometricians would agree that the first principal component of a broad battery of well-developed cognitive ability tests (e.g., the ASVAB) provides an excellent measure of intelligence. But the criticism is that the contents of the first principal component depend on the test battery, so, substantively, their meaning varies from battery to battery and therefore cannot be used as an unequivocal definition of intelligence.

It has been suggested that “g is to psychology what carbon is to chemistry” (Ree & Earles, 1993, p. 11). It would be nice if g could be specified with the precision of carbon (e.g., an element with six protons and six electrons, etc.) but, as the information processing approach to intelligence has found (see the next section), there does not appear to be one “thing” that constitutes intelligence. Instead, intelligence permeates behavior involving judgment, reasoning, and decision making. Thus, it is relatively straightforward to specify a process for developing a good measure of intelligence (i.e., use the first principal component of a broad battery of well-developed assessments of various types of reasoning and knowledge) but there is wide latitude for the contents of test battery.

**Information Processing Approaches to Intelligence**

Whereas the psychometric approach to the study of intelligence examines covariation among total test scores, the information processing approach decomposes responses to individual items into more elemental parts. Performance on these elemental parts can then be related to traditional measures of intelligence to identify the specific process or processes that constitute intelligence.

One of the most influential information processing conceptualizations is Sternberg’s (1977) componential model.
of intelligence. This approach begins with the *component*, which is defined as “an elementary information process that operates upon internal representations of objects or symbols” (p. 65). Sternberg noted that a “component may translate a sensory input into a conceptual representation, transform one conceptual representation into another, or translate a conceptual representation into a motor output” (p. 65).

Componential theories consist of two parts. First, the researcher must identify the elemental components required to perform a task; examples are given later. The researcher must also identify the processes by which the components are combined; this is often most easily described by a flowchart. The goal of the componential theory is to decompose response time (RT) to an item into its constituent parts; these parts include the time required to execute each component as influenced by the combination rules. For example, an item response might require $a$ ms for encoding, $d$ ms for responding, and the lesser of two processing times, $b$ and $c$. Thus, response time would be decomposed into $RT = a + \min(b, c) + d$.

Sternberg (1977) used a within-subject design to estimate the durations of the components for each respondent. These estimates are called the component scores. To evaluate a particular componential model, Sternberg examined the proportion of variance in response times accounted for by the model for each respondent. In one study, the best-fitting model accounted for 85% of the variance in response times for the most predictable respondent and 69% of the variance for the least predictable (the $R^2$ values were apparently not corrected for capitalization on chance).

To illustrate a componential model, consider an analogy $A$ is to $B$ as $C$ is to $D'$, which Sternberg (1977) denoted $(A:B::C:D')$. Sternberg’s model begins with encoding whereby an individual “identifies attributes and values of each term of the problem” (p. 135). Then, in successive steps, it is necessary to discover the rule relating $A$ to $B$, discover the rule relating $A$ to $C$, and then form a hypothesis about $D'$. Next, the match between a true–false alternative $D$ and the hypothesized $D'$ is evaluated, and finally the response is made. According to this model, the total time needed to solve the problem should equal the sum of the times needed to perform each step. Information processing models have been developed for a variety of tasks, including inductive reasoning (Pellegrino, 1985), deductive reasoning (Johnson-Laird, 1985), and verbal reasoning (Hunt, 1985).

Although important from the perspective of basic psychology, attempts to find a specific component that is strongly associated with intelligence (and that can therefore be interpreted as the essence of intelligence) have not been successful. Kyllonen and Christal (1990) wrote,

One of the hopes for this research was that complex cognitive abilities, such as reasoning ability, would be reducible to more elementary components, such as the inference component. Despite some successes (see Pellegrino, 1985, for a review), in one important sense this research can be looked upon as a modest failure. No one component was shown over different studies to be the essence of reasoning ability. (p. 427)

Thus, it appears that trying to derive the meaning of intelligence from a componential analysis of item responses can be likened to trying to learn about beauty by examining the Mona Lisa with a microscope; componential models provide little insight for understanding workplace behavior because they view intelligence from a distance that is too close.

Kyllonen and his colleagues have retained an information processing perspective but view intelligence from a distance better suited for understanding. The cognitive abilities measurement (CAM) project described by Kyllonen (1994) is grounded on his “consensus information-processing model” (p. 310) depicted in Figure 8.3. This model utilizes two long-term memories, one for procedural knowledge and one for declarative knowledge. The cognitive processing system retrieves information from these systems into working memory, where it is manipulated and a response is ultimately generated through the motor processing system. Clearly, Kyllonen takes a more molar view of intelligence than do the componential researchers.

Kyllonen and Christal (1989, 1990) suggested that performance on cognitive tasks is primarily a function of...
four of the components shown in Figure 8.3: procedural knowledge, declarative knowledge, cognitive processing speed, and working memory capacity. Certainly, greater amounts of declarative and procedural knowledge and faster cognitive processing should be related to superior performance. Kyllonen and Christal (1990) speculated that “the central factor is working-memory capacity. Working memory is the locus of both declarative and procedural learning . . . , and limitations in working memory are responsible for the difficulties of learning new facts (Daneman & Green, 1986) and procedures (Anderson & Jeffries, 1985)” (p. 392).

Baddeley’s (1986) definition of working memory capacity as the degree to which an individual can simultaneously store and manipulate information is central to Kyllonen and Christal’s (1990) research. This definition was used to develop several tests. For example, in the Alphabet Recoding test, examinees are given three letters (e.g., GVN is presented on a first computer-administered screen) and instructed to move forward or backward a certain number of letters (e.g., +2 on the second screen), and then type the answer (IXP). Interestingly, Kyllonen and Christal (1990) found strong relationships between their measures of reasoning ability and working memory capacity, with correlations estimated to be between 0.80 and 0.88 across four studies.

The work of Kyllonen and his colleagues has clear connections with the psychometric approach and Carroll’s (1993) three-stratum model. Kyllonen’s measures of reasoning ability might form a first-stratum factor lying beneath fluid intelligence, and his measures of working memory capacity appear to be related to the second-order memory factor. However, Baddeley’s (1986) conceptualization of working memory capacity is different from the digit span and free recall tests ordinarily used to define memory factors in psychometric studies in that he describes a central executive process responsible for controlled attention. Clearly, manipulating information held in short-term memory is cognitively challenging, and if tests of this sort are used to define a memory factor, it would be expected to be closer to Gf than a memory factor defined by tests such as digit span. It would be interesting to include several working memory capacity tests in a battery that used inductive, deductive, and quantitative first-order factors to identify second-order fluid intelligence as well as more standard first-order memory factors to define second-order memory; Kyllonen and Christal’s (1990) working memory capacity appears to be a combination of Gf and memory.

Summary of the Information-Processing Approach

This line of research has very carefully examined how people solve various types of questions. In effect, it identified the molecules of intelligence. Moreover, as illustrated by Sternberg’s (1977) large proportions of variance explained, information-processing models provide a substantially complete description of how examinees solve problems. However, no single element of the componential models has been found to be preeminent and consequently there is not a particular component or process that can be identified as intelligence. Recent research in this area has taken a more molar view. Kyllonen’s model, shown in Figure 8.3, for example, focuses on higher order constructs such as procedural knowledge, declarative knowledge, and working memory. It provides important insights and could guide the development of a variety of new tests.

Neuropsychological Approaches

Psychometric researchers view the brain as a black box whose functioning can be empirically investigated by examining the covariation in performance across diverse tasks. In contrast, neuropsychologists explicitly study the brain, functions of various parts of the brain, and interrelations of various functions. Although a detailed review of neuropsychological approaches to the study of intelligence is beyond the scope of this chapter, it is interesting and important to summarize some of the basic findings about the underlying hardware of the brain.

Parts of the brain are specialized for particular functions. In overview, the left side performs verbal information processing, and the right side processes visuospatial information and emotion. As an example of the specialization of the brain, different areas underlie the production and comprehension of speech. Paul Broca was a French neurologist who, in the 1860s, noticed that some patients could not produce speech but were able to understand speech (Banich, 1997). Broca performed autopsies on deceased patients and found damage to the left anterior hemisphere. Other patients with damage in the analogous location of the right hemisphere did not suffer a loss of fluent speech production (Banich, 1997). This inability to produce speech is called Broca’s aphasia.

Wernicke’s aphasia, in contrast, consists of loss of speech comprehension but fluent production of grammatically correct (but nonsensical) speech. It is caused by damage to the posterior left hemisphere (Banich, 1997). Again, damage to the mirror-image side of the right hemisphere does not cause this deficit.
Based on these anatomical findings, it seems plausible to hypothesize that the abilities to comprehend speech and produce speech are distinct and would be separable in carefully designed psychometric studies. To date, there has been little formal development of psychometric assessments of either speech production or comprehension. With the advent of multimedia computers, assessments of speech comprehension could be developed in a relatively straightforward manner. Examinees equipped with headphones could be presented with audio clips; after listening to the clip, multiple-choice questions could be presented either as audio clips or as text on the computer’s monitor.

Speech production is of course critically important in many occupations; its assessment is typically via unstructured interviews (or the job talk in academic circles). Computerized assessment of speech production is likely to become a reality within a few years; speech recognition software that converts speech to text (e.g., Dragon Dictate) could be linked with software used to grade essays (e.g., e-rater; Attali & Burstein, 2005) to produce virtually instantaneous scores.

Neuropsychological research provides important insights into our understanding of memory. Cohen (1997) pointed out that memory is not a unitary process. Rather, it must be thought of as a collection of memory systems that operate cooperatively, each system making different functional contributions and supported by different brain systems. Normal memory performance requires many of the brain’s various systems, which ordinarily operate together so seamlessly that intuitively appreciating the separate systems and the distinct contributions of each is difficult. (p. 317)

For example, working memory is not a unitary system. Cohen (1997) noted that auditory–verbal working memory can be severely compromised in some patients while their working memory for spatial relations and arithmetic remains perfectly intact. This has implications for developing assessments of working memory capacity; a richer assessment might be constructed by including items that tap into the different types of working memory. Although working memory has several distinct components, they all appear to be situated in the same part of the brain: the dorsolateral prefrontal cortex (Cohen, 1997).

Neuropsychological research clearly demonstrates the distinction between procedural memory and declarative memory that is part of Kyllonen’s (1994) information-processing model. Originally proposed by Cohen and Squire (Cohen, 1981, 1984; Squire & Cohen, 1984) and further elaborated by Cohen and Eichenbaum (1993), declarative memory accumulates facts and events and provides the means to learn arbitrary associations (e.g., people’s names, phone numbers); it is mediated by the hippocampal system, which includes the hippocampus, the amygdala, and the adjoining cortex. In contrast, skill acquisition and performance (e.g., riding a bicycle) are effected by procedural memory. Amnesia is caused by damage to the hippocampal system and affects declarative but not procedural memory. Thus, it is possible to teach patients with amnesia new skills; they do not have a conscious awareness of their recently acquired skills, but they can perform them (Cohen, 1997).

The executive functions, which “include the ability to plan actions toward a goal, to use the ramifications of behavior, and to make reasonable inferences based on limited information” (Banich, 1997, p. 369), are also studied by neuropsychologists. Banich noted that these activities are multifaceted and include the ability “to create a plan and follow through with it, to adapt flexibly, to sequence and prioritize, to make cognitive estimations, and to interact in a socially astute manner” (p. 370). Lezak (1995, pp. 43-44) provided a vivid description of a once-successful surgeon who suffered hypoxia as he was having minor facial surgery. His reasoning ability was spared (he continued to score high average to very superior on intelligence tests), but he was utterly unable to plan. He ultimately worked as a truck driver for his brother; after each individual delivery, it was necessary for him to call his brother for instructions about his next destination. The executive functions are typically compromised by damage to the prefrontal cortex.

In the past, conducting neuropsychological research was very difficult because it had been limited to observing patients with brain damage. In many cases, it was not possible to understand fully the nature and extent of brain damage until an autopsy was performed following a patient’s death. Recently developed brain imaging methods have been embraced by neuropsychologists because they allow direct, immediate observations of brain functioning. By tracking blood flow, researchers can see the parts of the brain that are active when specific activities are performed. PET examines brain activity via a radioactive agent, and fMRI examines changes in neuronal activity by using a contrast agent to track blood flow.

An example of this research is provided by Duncan et al. (2000), who used PET to examine brain activity while research participants performed tasks with high factor loadings on g (called high g tasks) and tasks with matching content but low factor loadings on g (low g...
tasks). The high g tasks were associated with increased activity of a specific area, namely, the lateral frontal cortex. For tasks with verbal content, the high g task was associated with increased activity in the left lateral frontal cortex relative to a matching low g task. In contrast, there was increased activity in both hemispheres’ lateral frontal cortex for tasks involving spatial content. Duncan et al.’s (2000) study is important because it found that the brain performed intellectual activities in relatively specific sites, rather than in multiple diffuse areas.

**Summary of the Neuropsychological Approach**

Until recently, conducting research linking psychometric theories of intelligence with the brain has been difficult if not impossible. Now PET and fMRI provide methods for imaging that make such research possible. As these methods become more available to researchers, it is likely that many important studies will be conducted.

Connections with neuropsychology deepen and enrich our understanding of intelligence. For example, inclusion of procedural and declarative memories in Kyllonen’s (1994) model has been shown to have an anatomical justification. Duncan et al.’s (2000) research has identified specific sites for reasoning and demonstrates that reasoning about verbal and spatial material involves different parts of the brain.

The executive functions identified in neuropsychological research suggest important directions for research by test developers. Situational judgment tests (discussed later) seem to provide a means for assessing executive functions, but to date they have not been developed with this in mind. Methods for assessing the executive functions are needed, as is research examining the relation of executive functions and job performance.

**INTELLIGENCE AND PERFORMANCE**

Two streams of research are important for understanding the relation of intelligence and performance. First, the topic of learning and skill acquisition has been of interest to psychologists since the beginning of psychology as a discipline. This research has ordinarily utilized laboratory studies of “subjects” learning relatively narrow tasks. In the other stream of research, job and training performance have been related to various measures of intelligence and aptitude in field studies. Across the entire gamut of predictors of job performance, Schmidt and Hunter (1998) noted that there have been “thousands of research studies performed over eight decades and involving millions of employees” (p. 271).

**Laboratory Studies of Skill Acquisition**

Ackerman’s 1987 literature review and series of experiments reported in a 1988 article provide a definitive picture of skill acquisition. Ackerman (1988, pp. 289–290) noted that skill acquisition is usually described as consisting of three phases (although different researchers use various terms for the phases). In the first phase, sometimes termed the **declarative stage**, heavy cognitive demands are made on the learner as he or she begins to understand and perform the task; responses are slow, and many errors occur. The next phase is sometimes called the **knowledge compilation stage**. Here, strategies for performance are developed, and responses become faster and with fewer errors. Finally, in the **procedural stage**, fast and accurate responses become highly automatic responses.

Schneider and Shiffrin (1977) defined **automatic processing** as “activation of a learned sequence of elements in long-term memory that is initiated by appropriate inputs and then proceeds automatically—with subject control, without necessarily demanding attention” (p. 1) and contrasted it with **controlled processing**, which “requires attention, is capacity-limited (usually serial in nature), and is controlled by the subject” (p. 1). The declarative stage of skill acquisition requires controlled processing, whereas automatic processing is used in the procedural stage.

Schneider and Shiffrin (1977) and Ackerman (1987, 1988) identified an important characteristic that affects skill acquisition. **Consistent tasks** are characterized by “invariant rules for information processing, invariant components of processing, or invariant sequences of information processing components that may be used by a subject to attain successful task performance” (Ackerman, 1987, p. 4). **Inconsistent tasks** are tasks where invariant rules or components do not exist. The key point is that skill acquisition for consistent tasks goes through the three stages just described and that the final stage is characterized by automatic processing; inconsistent tasks interrupt this process and always require controlled processing.

In a series of eight experiments, Ackerman (1988) showed that human ability requirements differ across the stages of skill acquisition and across the two types of tasks. Controlled processing is resource intensive; intelligence, as a measure of cognitive resources, is strongly correlated with performance in the declarative stage of skill acquisition. For consistent tasks, intelligence becomes less important as performance becomes automated. Perceptual speed, which is relatively unimportant for controlled
processing, becomes more strongly related to performance during the compilation stage but ultimately diminishes in importance. When performance becomes highly automated, it is primarily influenced by an individual’s psychomotor ability; psychomotor ability is much less important for performance in earlier stages that demand controlled processing. This pattern of relationships suggests that performance in assembly-line jobs would initially be related to workers’ cognitive ability, but g would quickly diminish in importance, and psychomotor abilities would ultimately determine performance.

In contrast, inconsistent tasks always require controlled processing, and cognitive ability consequently remains highly correlated with performance regardless of practice. In many managerial and technical jobs, individuals face continuously changing problems and issues. Here, Ackerman’s findings imply that general cognitive ability is always an important determinant of performance.

### Intelligence and Performance: Training and Job Criteria

The relation of intelligence and performance on the job and in training has been studied extensively for much of the past century. This literature was so vast and the effects of sampling variability so pernicious that the findings were essentially incomprehensible until statistical methods for aggregation across studies were introduced by Frank Schmidt and John Hunter (1977). Their meta-analytic procedure, which they termed *validity generalization*, provides a means for combining results across studies to estimate a population mean correlation between intelligence (or some other type of predictor) and a measure of job or training performance. In addition to minimizing the effects of sampling (because results of many studies can be combined), validity generalization allows corrections for range restriction and unreliability in job performance ratings. The method also allows researchers to estimate the population standard deviation of the validity coefficient; that is, after correcting for the effects of sampling, range restriction, and criterion unreliability, to what extent does the intelligence–job performance correlation vary across settings? A population standard deviation of zero implies that the relation of intelligence and job performance is invariant across settings and organizations.

Ones and colleagues (Ones, Viswesvaran, & Dilchert, 2005a, 2005b; Ones, Dilchert, Viswesvaran, & Salgado, 2009) have provided excellent summaries of the findings of the numerous meta-analytic studies investigating the relation of training and job performance criteria with intelligence. Concerning the prediction of training performance, Ones et al. (2005b) presented the results of 17 meta-analytic studies that consistently showed criterion-related validities (corrected for range restriction on cognitive ability and unreliability in the assessment of training performance) in the 0.5 to 0.6 range. Berry and Sackett (2009) predicted another important learning outcome, undergraduate grade-point averages (GPAs), for more than 150,000 students attending 41 colleges and universities. After controlling for differences in course selection, Berry and Sackett found a correlation of 0.672 between SAT total score—a good measure of intelligence—and GPA.

Ones and colleagues also summarized findings about the prediction of job performance. Remarkably, these results are based on a literal mountain of data: over 22,000 primary studies of over 5 million job applicants and employees. They concluded, “Data are resoundingly clear: GMA [general mental ability] is the most powerful individual-differences trait that predicts job performance across situations, organizations, and jobs” (Ones et al., 2005a, p. 450), with estimates of its criterion-related validity in the range of 0.5. Three conclusions are incontrovertible: In the United States, (a) the criterion-related validity of GMA increases as task and job complexity increases (Ones et al., 2005a); (b) the validity of GMA does not vary substantially or systematically across organizational settings (i.e., validity is general; Ones et al., 2005b); and (c) validity does not vary across subgroups that have been compared (men and women, White, African American, and Hispanic; Ones et al., 2005b). A tentative conclusion is that GMA is a powerful predictor of performance across cultures and nations (Ones et al., 2009); the research base supporting this conclusion is somewhat limited. Meta-analyses of primary studies conducted in the European Community (Salgado et al., 2003), the United Kingdom (Bertua, Anderson, & Salgado, 2005), and Germany (Hulsheger, Maier, & Stumpp, 2007) uniformly find that GMA is a powerful predictor of performance, but research in many other countries and cultures is still lacking.

It is important to note that empirical studies using intelligence to predict job performance will not ordinarily obtain correlations of approximately 0.5 even when large samples are obtained; instead, it is much more likely that a correlation of 0.25 will be observed. This will occur because job performance is always measured with error, which will reduce the correlation. Moreover, there is usually at least indirect selection on intelligence due to direct selection on other preemployment procedures (e.g., interviews) used in the hiring process. R. L. Thorndike’s
Intelligence and Performance: More Than $g$?

Ree and Earles (1991) and Ree et al. (1994) examined the extent to which specific abilities assessed by the ASVAB provide validity incremental to that of general cognitive ability for predicting job and training performance. These researchers used the first principal component from the ASVAB as their measure of $g$; they reported that other plausible methods for estimating $g$ from the ASVAB tests correlated in excess of 0.996 with the first principal component. The remaining principal components served as the measures of specific abilities. This partitioning of variance is useful because the measures of specific variance are orthogonal to the measure of $g$ and, moreover, because all of the specific variance is utilized.

Ree and his colleagues first computed the simple correlation between their measure of $g$ and the job or training school criterion measure and corrected for restriction of range. Next, the validity of the total test battery was estimated via multiple regression (with a multivariate correction for restriction of range), and then the multiple correlation was adjusted for capitalization on chance. Finally, the difference between the adjusted multiple correlation and the simple correlation was computed; it represents the incremental validity provided by the specific knowledge, skills, and abilities assessed by the ASVAB.

As a summary of their basic findings, Ree et al. (1994) reported that the simple correlation of $g$ (corrected for range restriction and averaged across occupations) with various measures of job performance was about 0.42. After a multivariate correction for range restriction and correction for capitalization on chance, the multiple correlation of the ASVAB battery with job performance averaged about 0.44. Thus, the incremental validity of the specific abilities assessed by the ASVAB was 0.02, which led to the remarkable conclusion that predicting job performance is “not much more than $g$,” to quote the article’s title.

There are at least three limitations regarding Ree et al.’s (1994) conclusion. First, the ASVAB does not provide reliable assessments of some of the various second-stratum factors of Carroll’s (1993) model depicted in Figure 8.2. Reliable and valid measures of these factors, as well as important first-order factors, would need to be included in the type of study conducted by Ree et al. before concluding that no specific cognitive ability adds to the predictive power of $g$. Second, Ree et al. considered only measures of cognitive ability; Schmidt and Hunter (1998) provided estimates of incremental validity for predicting job and training performance from other types of measures such as work samples, integrity tests, and the
personality trait of conscientiousness. Incremental validities large enough to have practical importance were found for several of the measures. Third, the criterion measures used by Ree et al. might best be described as assessments of task performance; measures of other important criteria in the enlarged criterion space were not included.

It appears premature to conclude unequivocally that there is “not much more than g.” Nonetheless, the work of Ree and his colleagues as well as numerous other practitioners who have used test batteries assessing cognitive abilities to predict task performance demonstrate that a search for incremental validity in this context is unlikely to be successful. Instead, to obtain incremental validity, it is probably necessary to use individual differences outside the cognitive domain to predict some measure of performance other than task performance.

SOCIAL AND EMOTIONAL INTELLIGENCE

Lezak’s (1995) report of a surgeon who became a truck driver needing special assistance despite intact reasoning abilities demonstrates that more than g is required for successful job performance. The executive functions summarized by Banich (1997) suggest several types of assessments that might be related to job performance. This section addresses the usefulness of interpersonal skills (i.e., “social intelligence”) and emotional intelligence as predictors of performance in the workplace.

To understand the relation of social and emotional intelligence with job performance, it is important to think carefully about the aspects of job performance for which incremental validity might be obtained. In this regard, Borman and Motowidlo’s (1993) distinction between task and contextual performance is important. Borman and Motowidlo (1997) argued that contextual activities are important because they contribute to organizational effectiveness in ways that shape the organizational, social, and psychological context that serves as the catalyst for task activities and processes. Contextual activities include volunteering to carry out task activities that are not formally part of the job and helping and cooperating with others in the organization to get tasks accomplished.

Thus, a major part of contextual performance appears to be intrinsically social in nature.

Is contextual performance important in the workplace? Motowidlo and Van Scotter (1994) conducted a study to address this question. Using a sample of 421 U.S. Air Force mechanics (which is not a job where one would expect contextual performance to be especially salient), Motowidlo and Van Scotter obtained job performance ratings from three different supervisors. One rated overall job performance; one rated task performance; and one rated contextual performance. Contextual performance was assessed by a 16-item scale; these items asked supervisors how likely it was that the mechanic would perform various contextual behaviors, including “cooperate with others in the team,” “persist in overcoming obstacles to complete a task,” “defend the supervisor’s decisions,” and “voluntarily do more than the job requires to help others or contribute to unit effectiveness” (p. 477). The ratings of task performance correlated 0.43 with overall performance; remarkably, the contextual performance measure correlated 0.41 with overall performance. Thus, in a prototypical blue-collar job, contextual performance and task performance appear to be equally important components of overall job performance. Similar results have been reported by Borman, White, and Dorsey (1995); Dunn, Mount, Barrick, and Ones (1995); Ferris, Judge, Rowland, and Fitzgibbons (1994); and Werner (1994).

Collectively, these findings show that cooperating with others and helping coworkers are important in virtually every job. The extent to which an employee actually enacts such behaviors appears likely to be a function of both willingness to help and the capability (a) to recognize situations where one should help others or defend the organization and (b) to know what steps to take. These capabilities appear to have a knowledge component; consequently, social and emotional intelligence may be related to the performance of some aspects of behavior.

Measurement of Social Intelligence

Two distinct conceptual approaches to the measurement of social intelligence have been taken, although both seem to have originated with E. L. Thorndike’s (1920) definition of social intelligence as “the ability to understand and manage men and women, boys and girls—to act wisely in human relations” (p. 228). In the first line of research, instruments explicitly intended as measures of social intelligence were developed. One of the earliest measures was the George Washington University Social Intelligence Test developed by Moss, Hunt, Omwake, and Ronning (1927). The other line of research consisted of situational judgment tests (SJTIs) that are intended primarily to predict job performance. Early examples include the How Supervise? test (File, 1945; File & Remmers, 1948) and the Supervisory Judgment Test (Greenberg, 1963).
In addition to the conceptual approaches, two technological approaches have been taken. For most of their history, social intelligence tests and SJTs utilized a paper-and-pencil format. Increasingly, SJTs use video assessments presented via computer (Olson-Buchanan et al., 1998).

**Definition of Social Intelligence**

Walker and Foley (1973) provided a review of research on social intelligence during the 50 years following E. L. Thorndike’s 1920 paper. They noted that the two key elements of E. L. Thorndike’s definition were “the ability to (a) understand others and (b) act or behave wisely in relating to others” (p. 842). They further noted that O’Sullivan, Guilford, and deMille (1965) viewed social intelligence as the ability to understand other people’s feelings, thoughts, and intentions. Walker and Foley also cited Flavell, Botkin, and Fry (1968) as providing “the single most extensive analysis and investigation of the development of various aspects of social-cognitive functioning” (p. 844). Flavell et al. argued that effective social interacting requires five steps. First, an individual must recognize the existence of other people’s perspectives (i.e., an individual needs to realize that others may perceive a particular situation very differently than he or she does). Second, the individual must understand the need to consider other people’s perspectives. Third, the individual must have the ability to predict how others will perceive a situation. Fourth is the need for maintenance of perceptions of others’ perspectives when they conflict with one’s own views. The last step is the application of this understanding of others’ views to determine one’s behavior in a particular situation.

**Explicit Measures of Social Intelligence**

It is important for measures of social intelligence to exhibit discriminant validity from other constructs. Unfortunately, Riggio (1986) noted that “difficulties in assessing social intelligence, particularly the inability to discriminate social intelligence from general intelligence, led to the demise of this line of research” (p. 649). Riggio’s (1986) Social Skills Inventory represents a more recent attempt to develop a measure of social intelligence. It utilizes a “typical performance” format rather than a “maximal performance” format. For example, an item is “At parties I enjoy speaking to a great number of different people” (p. 652). Thus, distinguishing the Social Skills Inventory from cognitive ability is unlikely to be a problem; however, establishing discriminant validity vis-à-vis personality is clearly important.

Riggio (1986) viewed social intelligence as “not a single entity but, rather, a constellation of many more basic skills” (p. 650). The Social Skills Inventory includes six of these more basic skills: emotional expressiveness, the ability to communicate one’s affect and attitudes; emotional sensitivity, the ability to “decode others’ emotions, beliefs, or attitudes, and cues of status-dominance” (p. 650); social expressiveness, the ability to express oneself verbally and initiate conversations; social sensitivity, the ability to understand others’ verbal statements and recognize social rules and norms; emotional control, “the ability to regulate emotional communications and nonverbal displays” (p. 650); and social control, an individual’s social self-presentation skill. Riggio reported high internal consistency and test-retest reliabilities for his instrument and generally satisfactory results in an exploratory factor analysis. However, some of the subscales of the Social Skills Inventory had large correlations with scales of the 16 Personality Factor (16 PF) instrument (Cattell, Eber, & Tatsuoka, 1970). For example, the Social Control scale correlated 0.69 with the 16 PF Shy-Venturesome scale and −0.78 with the Social Anxiety scale.

**Definition of Emotional Intelligence**

There are two principal approaches to defining emotional intelligence (EI). The ability model posits emotional intelligence as

- the ability to engage in sophisticated information processing about one’s own and others’ emotions and the ability to use this information as a guide to thinking and behavior. That is, individuals high in EI pay attention to, use, understand, and manage emotions, and these skills serve adaptive functions that potentially benefit themselves and others. (Mayer, Salovey, & Caruso, 2008, p. 503)

The four-branch model of EI (Mayer & Salovey, 1997) hypothesizes four facets to the EI construct:

- “Perceiving emotions accurately in oneself and others;
- Using emotions to facilitate thinking;
- Understanding emotions, emotional language, and the signals conveyed by emotions; and
- Managing emotions so as to attain specific goals.” (Mayer et al., 2008, p. 507)

The second approach to defining EI has been termed the mixed model. This approach defines EI in terms
of traits such as assertiveness, optimism, impulsiveness, and so forth. Bar-On (1997), for example, defined EI as “an array of noncognitive capabilities, competencies, and skills that influence one’s ability to succeed in coping with environmental demands and pressures” (p. 14).

Measures of EI

The Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) uses an ability-testing format (i.e., test takers’ responses are evaluated with a scoring key) to assess their four facets of EI. Due to this maximal-performance approach to measurement, low correlations of MSCEIT facets and total score with the Big Five personality dimensions have been observed. Roberts, Schulze, and MacCann’s (2008) meta-analysis found that the MSCEIT total score correlated 0.12 with Openness to Experience, 0.07 with Conscientiousness, 0.05 with Extraversion, 0.22 with Agreeableness, and −0.07 with Neuroticism. As an aspect of intelligence, EI would be expected to have positive correlations with other measures of intelligence, and Roberts et al. found that the MSCEIT Total score correlated 0.18 with fluid intelligence, 0.35 with crystallized intelligence, and 0.31 with a composite of fluid and crystallized intelligence. In sum, these correlations provide evidence of construct validity for EI because relatively small correlations were obtained with dimensions of personality and some moderate correlations were obtained with aspects of intelligence.

Measures based on the mixed model of EI are much more problematic. They rely on self-reports rather than objectively scored items and hence appear very similar to personality items. Roberts, MacCann, Matthews, and Zeidner (2010), for example, note:

The extent that self-report measures [of EI] correlate with personality and especially assessments of the Big Five personality factors is very high. . . . De Raad (2005) showed that 66% of items drawn from self-report inventories of EI could be classified under the Big Five Framework. . . . correlations between the Big Five and self-report measures have been found to be around 0.50–0.70 for at least one of the super-factors, with multiple correlations approaching 0.80, and near unity if corrected for attenuation. (p. 4)

A Theory for EI

Based on an extensive literature review and meta-analysis, Joseph and Newman (2010) proposed and tested their cascading model for EI. The core of the model hypothesizes that emotion perception precedes emotion understanding, which in turn precedes emotion regulation. Each of these constructs was hypothesized to have a different individual difference exogenous influence. Conscientiousness was hypothesized to influence emotion perception because “conscientious individuals may develop a heightened perception of self-conscious emotions as a sort of radar to detect when they have lost control of their behavior” (p. 58). Cognitive ability is hypothesized to influence emotion understanding because “individuals with high cognitive ability would acquire a stronger knowledge base associated with understanding one’s emotions” (p. 59). And emotional stability was hypothesized to drive emotion regulation because “neurotic individuals do not engage in effective emotion regulation strategies (i.e., reappraisal) as often as emotionally stable individuals” (p. 59). Emotion regulation was expected to predict job performance and to fully mediate the relation of emotion perception and emotion understanding with performance. However, conscientiousness, cognitive ability, and emotional stability were all expected to have relations with job performance that were partially mediated by the emotion variables.

Joseph and Newman (2010) fit a path model to a meta-analytically derived correlation matrix and found a good fit. The analysis revealed a 0.28 path from conscientiousness to emotion perception, a 0.35 path from cognitive ability to emotion understanding, and a 0.12 path from emotional stability to emotion regulation. Emotion perception led to emotion understanding (0.43 coefficient) and emotion understanding led to emotion regulation (0.53). Conscientiousness predicted job performance (0.22) as did cognitive ability (0.44), but not emotional stability. Finally, the path from emotion regulation to job performance was significant, but small (0.08). It should be noted that this part of the model was based on very little data: only 8 studies with a total N = 562. Clearly, there is a strong need for more primary research on the EJ–job performance relationship.

Situational Judgment Tests

Due to the historical difficulty in assessing social intelligence and the problematic literature on the measurement of EI, we turn to a discussion of SJTs because they appear to offer a useful approach to the measurement of such elusive constructs.

SJTs present descriptions of workplace situations and ask the respondent either (a) what he or she would do (behavior tendency instructions); or (b) the effectiveness of the response options (knowledge instructions; see McDaniel, Hartman, Whetzel, & Grubb, 2007). They
are often developed by interviewing job incumbents and asking about critical incidents (Flanagan, 1954). Information gleaned from these interviews is then transformed into the items constituting an SJT. As a result of this process, the items on SJTs are viewed as interesting and face valid by job applicants and employees (Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). McDaniel and Nguyen (2001) provided a summary of the constructs assessed in SJTs as well as test development procedures.

McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001) described the history of SJTs. These authors noted a flurry of activity in the 1940s but less emphasis during the ensuing decades. Motowidlo et al.’s (1990) “low-fidelity simulation” appears to have reenergized work in this area. For example, a recent book edited by Weekley and Ployhart (2006) presents a great deal of information about SJTs.

Relation of Situational Judgment Tests to Job Performance

There has been much research examining the relation of SJTs to measures of job performance. For example, cross-validation samples demonstrated that Weekley and Jones’s (1997) video-based SJTs were substantially related to job performance. In preparation for their meta-analysis, McDaniel et al. (2001) identified 102 validity coefficients for 39 different SJTs based on data from 10,640 research participants. They then conducted a meta-analysis, correcting for range restriction and unreliability in measures of job performance. After these corrections, McDaniel, Morgeson et al. estimated the population mean correlation of SJTs with job performance to be 0.34.

Due to the large number of correlations identified by McDaniel et al. (2001), they were able to examine the moderating effect of g on the SJT–job performance relationship. High-g tests were defined as SJTs with mean correlations with g in excess of 0.50; medium-g SJTs had correlations with g between 0.35 and 0.50; and low-g SJTs had correlations below 0.35. McDaniel et al. (2001) estimated the mean population validity of high-, medium-, and low-g SJTs to be 0.41, 0.18, and 0.34. Although confidence intervals for these point estimates were not provided, it is unlikely that the high-g and low-g validities differ significantly from one another.

The implication of McDaniel, Morgeson et al.’s meta-analysis is that researchers can build predictive SJTs that are more or less related to general cognitive ability. Of course, the incremental validity of an SJT will be greater when it has a smaller correlation with g. In some cases it may be very useful, however, to construct an SJT with a very large correlation with g. For example, in a tight labor market, applicant reactions to selection procedures can be very important. It is unlikely that applicants for senior executive positions would enjoy taking a test like the Wonderlic, and consequently they might drop out of the recruitment process. However, a senior executive might be intrigued (and hence remain a job candidate) by an SJT that is fundamentally a cognitive ability test in disguise. Consistent with McDaniel et al.’s (2001) meta-analysis, Weekley and Jones (1999) concluded that “SJT represents a method and not a construct” (p. 695).

Linking Situational Judgment Tests and Social Intelligence

Chan and Schmitt (1997) conducted a study that examined how administration medium affected an SJT’s correlation with g (and the resulting Black–White score difference). Two forms of SJTs were developed; the forms had identical content, but one was administered via paper and pencil, and the other used a video-based administration. The paper-and-pencil version was found to correlate 0.45 with a measure of reading comprehension, but the correlation for the video version was just 0.05. Particularly noteworthy were the effect sizes for Black–White differences: −0.95 for paper-and-pencil administration versus −0.21 for video presentation. Thus, the paper-and-pencil form was moderately confounded with g and had substantial adverse impact; the video version was independent of g with little adverse impact.

Olson-Buchanan et al.’s (1998) video-based SJT also had near-zero correlations with measures of cognitive ability. It is notable in that it predicted overall job performance and managers’ skills at resolving conflict in the workplace. Their measures of cognitive ability also predicted overall job performance but did not significantly predict conflict resolution performance.

A hypothesis that explains the pattern of results obtained by McDaniel et al. (2001) and Olson-Buchanan et al. (1998) is that high-g SJTs predict job performance, and especially task performance, because of the strong g–job performance relationship. However, low-g SJTs may measure mainly social intelligence uncontaminated by g; they may have a stronger relationship with a measure of contextual performance because of its fundamental social nature. Clearly, further research is needed to understand why both high- and low-g SJTs have similar validities.
FALLACIES AND MISINFORMATION ABOUT INTELLIGENCE

Despite an extraordinarily large empirical literature investigating intelligence, some demonstrably incorrect beliefs have persisted. Carroll (1997) and Kuncel and Hezlett (2010) describe some of these fallacies. A few of them are described here.

A first fallacy is that cognitive ability tests do not predict academic performance or job performance. As noted in a previous section, there are mountains of data that disprove this belief. A more nuanced version of this fallacy is that intelligence tests really measure socioeconomic status (SES) and, if SES is controlled for, intelligence tests would not predict academic or job performance. Sackett, Kuncel, Arneson, Cooper, and Waters (2009) examined several very large data sets to investigate this hypothesis. They found that, for the SAT, SES was indeed related to test scores \( r = 0.42 \). After correcting for range restriction, the average (over 41 colleges and universities) SAT total score correlation with freshman GPA was \( r = 0.44 \). Clearly, intelligence tests measure something other than SES and controlling for SES does not diminish the intelligence–performance correlation. Moreover, Sackett et al.’s results suggest that critics of testing may have their causal arrow pointing in the wrong direction: The pattern of correlations they reported suggests that intelligence leads to SES rather than the reverse.

There is no dispute that the mean test scores of some minority groups (African American, Hispanic) are substantially lower than the majority White group. However, there is a persistent belief that cognitive ability tests underpredict the performance of these minority groups in college and on the job. Despite numerous studies, there is no evidence for underprediction (Linn, 1973; Sackett, Borman, & Connelly, 2008; Young, 2001): The overall regression line or the White group’s regression line accurately predicts (or overpredicts) minority group performance. A caveat on this conclusion is that the criterion measures used in this research would typically be classified as task performance or their analog for college (GPA). Much less research has investigated contextual performance, CWBs, and various dimensions of Bartram’s (2005) Great Eight.

Another fallacy is that above some threshold, differences in intelligence do not matter. David Lubinski and his colleagues have tracked the careers of several samples of gifted youths (top 1%) and compared those in the bottom quarter of this group (i.e., in the 99.00th to 99.25th percentiles) to those in the top quarter (99.76th percentile and above). Lubinski (2009), for example, found that individuals in the top quarter were over three times as likely to obtain a doctorate, five times as likely to publish a scientific paper, and three times as likely to receive a patent as individuals in the bottom quarter of the top 1%. Certainly, the extraordinarily gifted are higher performers than those who are “merely” highly gifted.

A related fallacy is that level of education explains differences in career success. For example, in a paper based on his 2009 American Educational Research Association Distinguished Lecture, Hauser (2010) claimed that “Among persons with equal levels of schooling, IQ has little influence on job performance” (p. 95). Park, Lubinski, and Benbow (2008) clearly refute this belief. For example, in an intellectually talented sample with terminal bachelor’s degrees, individuals in the top quarter on the SAT–Math were four times as likely to receive a patent and two times as likely to have a scientific publication as those in the bottom quarter. Among individuals with a terminal master’s degree, the top quarter had five times as many patents and 12 times as many scientific publications. For those with doctorates, the top quarter had three times as many scientific publications and almost five times as many patents. These differences, although large, suffer from range restriction: All of the individuals in the Park et al. study were in the top 1% of their age group on the SAT-M when they were tested (before age 13). Much larger differences in the performance of individuals with the same terminal degree would be expected if the full range of ability were included.

CONCLUSIONS

Psychometric, cognitive, and neuropsychological approaches to investigating intelligence provide complementary perspectives to this important area of human functioning. Convergent evidence across disciplinary lines greatly strengthens the confidence of our conclusions.

Carroll’s (1993) three-stratum model, depicted in Figure 8.2, represents a landmark accomplishment in the psychometric study of intelligence. It is a comprehensive elaboration of Vernon’s (1950) hierarchical model that summarizes and integrates literally hundreds of factor analytic studies.

Comparing Carroll’s (1993) model to Spearman’s original theory presented in 1904, it is interesting to see how far research has progressed. Spearman’s theory could adequately describe a correlation matrix if one test beneath each of the eight stratum II factors was included; if more
than one test beneath a stratum II (or stratum I) factor was included, Spearman’s theory is not supported. Nonetheless, for understanding performance in the workplace, and especially task performance and training performance, g is key. As demonstrated by Ree and colleagues (Ree & Earles, 1991; Ree et al., 1994), g accounts for an overwhelming proportion of the explained variance when predicting training and job performance.

The information-processing models for cognitive ability test items of Sternberg (1977), Hunt (1985), and others provided important information about what intelligence tests measure. Specifically, no one element of these componential models emerged as the fundamental process of intelligence, thus suggesting that intelligence should be viewed as a mosaic of microprocesses. For understanding and predicting job behavior, a more macro-level perspective better serves researchers. Kyllonen’s (1994) consensus information processing model provides a useful framework for understanding performance on cognitive ability tests. His demonstration of the importance of working memory should influence psychometric researchers. Moreover, computerized assessment greatly facilitates measurement of time-related phenomena such as working memory and should allow measures of working memory to be routinely included in test batteries. Baddeley’s (1986) research on the structure of working memory provides a solid conceptual foundation for developing test specifications for assessments in this area.

To date, there has been little interaction between researchers with psychometric and neuropsychological perspectives. In part, this has been due to the difficulty in measuring brain activity while performing psychometric tasks. The article by Duncan et al. (2000) demonstrates the value of such collaborations.

Research on social and emotional intelligence has had a dismal history. Measures of these abilities have been found to be either unreliable or confounded with cognitive ability. Nonetheless, neuropsychologists(e.g., Wendy Heller, personal communication, December 3, 2000) can describe individuals who are unable to keep jobs, who are unable to remain married, or who are unable to interact appropriately with others following head injuries despite intact cognitive abilities. Clearly, important abilities have been compromised in such individuals, but standard measures of cognitive skills are insensitive to the consequences of the injuries. Measures of social and emotional intelligence unconfounded with g and personality are needed.

Video-based SJTs may provide this type of assessment. It would be fascinating to use PET or fMRI to examine the locus of brain activity for individuals responding to the two versions of Chan and Schmitt’s (1997) SJT. One might hypothesize left lateral frontal cortex activity for the paper-and-pencil SJT because verbal reasoning is used to process the items. In contrast, the brain may be more active in the right hemisphere for the video SJT because this is where emotions are processed. Such results would explain why video-based SJTs have been found to be unrelated to cognitive ability by Chan and Schmitt (1997) and Olson-Buchanan et al. (1998).

In conclusion, despite a century of research on intelligence, much work remains. Is working memory as important as Kyllonen and Christal (1990) believe? How should assessments of working memory be constructed, and will they add incremental validity to predictions of important job behaviors? Will measures obtained from other tasks in the cognitive domain add incremental validity? Will video-based assessments finally provide a means for assessing social intelligence? What will brain imaging studies find when they examine individuals answering video-based assessments? Will video-based SJTs predict contextual job performance better than g? How should emotional intelligence be measured? Clearly, intelligence research represents an area with many important and exciting issues as yet unresolved.

REFERENCES


Intelligence and the Workplace 209


210 Personnel Psychology


