In: Sternberg, R. J., & Grigorenko, E. L. (2002) The general factor of intelligence: How general is it? Mahwah, NJ: Erlbaum.

Chapter 13

g: Highly General and Highly Practical

Linda S. Gottfredson
University of Delaware

The general mental ability factor, *g*, is real. Its existence is no longer a serious question among experts on intelligence (Carroll, 1993). Whatever its underlying nature, psychometric *g* is a reliably measured, replicable phenomenon across all age, race, gender, and cultural groups studied so far (Jensen, 1998). Consequently, among intelligence researchers, it has become the most common working definition of "intelligence." A more important question today is: How *generally useful* are higher levels of *g outside* the realm of paper-and-pencil tests and tasks? The term *intelligence* connotes a very general and broadly useful capacity. Is that the label warranted for *g*? Even if it is, might not the label be warranted for other abilities too, leaving *g* as only one among various intelligences?

This chapter addresses these questions. It first outlines criteria for assessing how broadly useful *g* or any other trait is to individuals in "real life." This is *g*'s *sociological* generality as distinct from its *psychometric* generality, the latter referring to its value in explaining the correlations among mental tests themselves. The former is the *range of life tasks* across which higher levels of *g* meaningfully affect performance. Second, the chapter reviews *g*'s utility in one highly studied sphere of life—job performance. The considerable data on *g* in the workplace provide guideposts for understanding the pattern of *g*'s generality in other nonacademic realms. The chapter next uses these guideposts to examine *g*'s generality in two such realms: the specific tasks in daily life, such as driving and health self-management, and cumulative life outcomes such as socioeconomic success and social pathology.

Theories of intelligence differ considerably in their assertions regarding g's generality. Two theories are compared throughout the chapter: g theory, which predicts that g's utility generalizes widely and without regard to a task's manifest content or context, and practical intelligence theory, which postulates that g is useful in "academic" tasks but has relatively little value in practical affairs (where a proposed "practical intelligence" is, instead, said to be essential). As will be shown, g theory is more consonant with the facts. Whether or not a task seems academic offers scant guidance as to whether its performance is enhanced by higher levels of g. In no realm of life is g all that matters, but neither does it seem irrelevant in any. In the vast toolkit of human abilities, none has been found as broadly useful—as general—as g.

CRITERIA FOR GAUGING THE PRACTICAL IMPORTANCE OF AN ABILITY

Mapping the sociological generality of g requires understanding where and why higher levels of g are most and least useful to individuals throughout their lives. It is thus a matter of knowing the *pattern* or topography of g's utility, that is, its depth and breadth of impact across diverse arenas of life. Depth of impact is gauged by a trait's *effect sizes* in individual realms of activity; breadth is gauged by the number of realms in which the trait has meaningful effect, which is its *generality*.

Gauging Effect Sizes

Effect size refers specifically to how big a change in the outcome in question is produced by a given change in the predictor (e.g., Cohen, 1988, p. 22; Jensen, 1980, pp. 305–310). In experimental research, where the predictor can be manipulated, effect size is often calculated in terms of standard deviation units of change in the outcome (reading achievement or cigarettes smoked per day) due to some treatment (reading instruction or smoking cessation program). In nonexperimental psychological research, the possible causal importance of a predictor is typically quantified in terms of correlations between predictors and criteria, including regression coefficients (b and beta) and the multiple correlation (R). Odds ratios are often used in other fields, such as epidemiology (Gerstman, 1998).

Although R squared (proportion of variance explained) is sometimes mistakenly used to measure effect size, R (or its analogs) is the proper measure of a predictor's effect in the real world because it "is directly proportional to the practical value of the [predictor]—whether measured in dollar value of increased output or percentage increase in output" (Schmidt & Hunter, 1998, p. 272). A correlation of .4 (or .2) means that a one standard deviation change in the predictor (say, g) is associated with a .4 (or .2) standard deviation change in the outcome (say, quality of job performance or understanding of a physician's instructions). If the correlation is viewed as the predictor's potential rate of return or leverage for change, a predictor that correlates .4 with the outcome has twice as much leverage as one correlating .2. In the worlds of investing and gambling, these rates would be extraordinary. In the world of psychological intervention, point biserial correlations of this size are respectively considered "large" and "medium" (Lubinski & Humphreys, 1997, Table 2).

Gauging Generality

Generality is the range of human activity across which an ability has meaningful effect sizes. It is greater to the extent that higher levels of g provide an advantage to individuals over a greater variety of task domains, ranges of the g continuum, ages, generations, and cultures. I focus on the first, partly because the skeptics of g's utility have often focused on task characteristics to press their case.

More importantly, the very definition of an ability is rooted in tasks performed. To abbreviate Carroll's (1993, pp. 3–9) meticulously crafted definition, an *ability* is an attribute of individuals revealed by differences in the levels of task difficulty, on a *defined class of tasks*, that individuals perform successfully when conditions for maximal performance are favorable. The broader the class of tasks, the more general the ability is. Another reason for focusing on tasks is that the other four conditions set forth earlier all influence the configuration of tasks people actually undertake. For instance, the young encounter, seek out, and are expected to master different tasks than their elders. Task expectations and preferences likewise differ for the bright versus the dull, for people entering jobs in the information age rather than the industrial age, and for citizens of widely different economies or cultures. Understanding g's generality across different tasks can therefore help explain any variations in its utility across time, place, age, and range of ability.

Predictions of g Theory Versus Practical Intelligence Theory

Theorists of g conceptualize it as a general capacity for processing information of any kind. As such, g undergirds critical thinking skills such as reasoning, thinking abstractly, spotting and solving problems, and quickly and efficiently learning moderately complex material (see chap. 3 by Jensen, this volume). g theory therefore predicts that higher g will enhance performance in all tasks that require information processing. It also predicts, however, that task performance will depend more heavily on differences in *g* (be more "*g* loaded") when the task requires more complex information processing. Accordingly, a task's *g* loading would have little to do with its manifest content, including whether it seems school-like or not. Indeed, it is well known that *g* loadings are low for some manifestly academic tasks (such as spelling and arithmetic computation among adults) but high for others with the same kind of content (reading comprehension, mathematical reasoning).

In contrast, practical intelligence theorists tend to conceptualize *g* as only an "academic" intelligence ("book smarts"), as distinct from "practical" intelligence ("street smarts" or common sense), which they posit is relatively independent of *g* (Sternberg, Wagner, Williams, & Horvath, 1995). Thus narrowing *g* to mere book smarts or "inert" rather than active intelligence, they assert that highly *g*-loaded tests tap only "a tiny and not very important part" of the intellectual spectrum (Sternberg, 1997, p. 11).

The proponents of practical intelligence have clarified the *g*-is-only-book-smarts thesis by defining what they mean by academic versus practical tasks (Sternberg & Wagner, 1993; Sternberg et al., 1995). As shown in Table 13.1, academic tasks are said to call for thought and not action, are imposed rather than chosen, are esoteric, and their answers and means of solution are highly circumscribed. In contrast, both the nature of the problem and the solution of practical tasks are said to be more ambiguous, and their solution (of which there may be several) requires everyday experience and personal interest. The difference between academic and practical is thus a distinction between, on the one hand, the narrow, pedantic, disconnected theoretical and, on the other, the messy, meaningful reality

TABLE 13.1
Sternberg and Wagner's (1993) Definition
of Academic Versus Practical Tasks

"Academic" problems tend to:	*"Practical" problems tend to:*
1. Be formulated by other people	1. Require problem recognition and formulation
2. Be well-defined	2. Be ill-defined
3. Be complete	3. Require information seeking
4. Possess only a single correct answer	4. Possess multiple acceptable solutions
5. Possess only a single method of obtaining the correct answer	5. Allow multiple paths to solution
6. Be disembedded from ordinary experience	6. Be embedded in and require prior everyday experience
7. Be of little or no intrinsic interest	7. Require motivation and personal involvement

in which people actually live. Whereas *g* may be crucial in the former, it is not in the latter, posit the proponents of practical intelligence theory. Their prediction would seem to be that *g*'s criterion validities (its correlations with outcomes) will be higher when tasks are more academic (e.g., are well-defined, disembodied from ordinary experience, and of little intrinsic interest) and smaller when they more practical (e.g., require problem recognition and formulation, information seeking, and personal involvement).

In short, both *g* theory and practical intelligence theory agree that *g*'s impact is moderated by task attributes, but they disagree on which ones. The latter suggests that the effect sizes for *g* rise for more academic tasks, whereas the former suggests that they rise for more complex ones, whether academic or not. Practical intelligence theory suggests that *g* is therefore not very general because academic tasks are confined mostly to school settings. *g* theory suggests, in contrast, that higher *g* has pervasive value because people face complex tasks in many aspects of life; it is not only an academic ability, but also a highly practical one.

g and Job Performance

Outside of education, the most intensely studied sphere of intelligent performance has been job performance. For many decades, teams of military, public, and private sector researchers have spent incalculable person-years documenting the determinants of performance in training and on the job. The century-old field of personnel selection psychology has been devoted to just this effort.

THE NATURE OF RESEARCH ON JOB PERFORMANCE

Initial Reluctance to Entertain *g* Theory

Many personnel psychologists have turned to *g* theory in recent years, but for many decades the field was ruled by the *theory of situational specificity*. This was the belief that there are many independent abilities and that the particular mix of abilities that is relevant to a job—and even to the individual positions within a job classification—depends on the detailed specifics of the position's duties and setting. Intelligence was viewed as only one among many aptitudes affecting performance, and its importance was thought spotty and unpredictable—as was that of all other predictors. By this theory, no trait had general utility. Schmidt and Hunter (1998), among others, showed via meta-analysis that the specificity doctrine was sustained by statistical artifacts owing to most research samples being small and somewhat homogeneous in mental ability. Once those artifacts are taken into account, *g*'s importance is seen to be pervasive and lawfully patterned. The specificity doctrine did not die for lack of enthusiasm, but from the crush of accumulating evidence (e.g., see Humphreys' personal account, 1986).

Personnel psychologists as a group never expected intelligence to be important and many wish that it were not. As *g* has shown ever greater promise for explaining job performance, it has become more subject to concerted efforts to *disconfirm* its functional importance and to find alternatives to mental tests for selecting and promoting workers. The major reason for such efforts, often bordering on the desperate, has been that *g*-loaded employee selection typically screens out proportionately many more Blacks and Hispanics (has "disparate impact"), which makes an employer vulnerable to legal and political attack (Sharf, 1988). Although useful for prompting more interest in how non-*g* traits affect job performance, this effort to negate the apparent functional impact of *g* has, ironically, only further confirmed it (e.g., Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). All but a few personnel psychologists now accept that *g* has special importance for work performance (e.g., see the special issue of *Human Performance* by Viswesvaran & Ones, in press), but that acceptance has been for the most part a grudging concession to empirical evidence. The move to *g* theory therefore cannot be attributed to any so-called "*g*-ocentric" enthusiasm (Sternberg & Wagner, 1993) on the part of personnel psychologists.

The Body of Evidence

Evidence has gone far beyond showing that *g* has a big overall impact on job performance to showing where it has its largest impact, and why. The major discoveries about *g*'s gradients of effect are listed in Table 13.2 and are discussed later. In presenting the evidence for the generalizations in Table 13.2, I have relied on meta-analyses of thousands of small studies and on two very large military projects. All the correlations with *g* in this chapter's Job Performance section have been corrected, except where noted otherwise, for two statistical artifacts that depress correlations below their true level: unreliability and restriction in range of mental ability. Most personnel selection research has measured job performance with supervisor ratings, so, unless specified otherwise, all correlations with performance reported here refer to performance measured subjectively rather than objectively. As we shall see, subjective measures of performance lead to underestimating *g*'s impact on job competence.

***The Measurement of* g.** Only a small proportion of job performance studies have actually correlated job performance with *g* scores, because estimating *g* scores requires administering a broad battery of tests from which the *g* factor can be extracted. However, many kinds of studies have shown, over and again, that *g* is by far the biggest component of all mental tests (Carroll, 1993). So great is their *g* component that mental tests are usually just "flavored" by the special abilities they are meant to measure. Moreover, it is the *g* component of mental tests that usually accounts almost entirely for their predictive value (e.g., Jencks et al., 1979, chap. 4; Jensen, 1998; Ree & Carretta, 1997; Thorndike, 1986). Tests differ in the degree to which they are *g* loaded (that is, in their ability to measure *g*), but the highly *g*-loaded ones can be treated as interchangeable, albeit imperfect, measures of *g*, no matter how they are labeled (verbal, quantitative, spatial, and the like). I shall therefore refer to all mental tests as de facto measures of *g*. One test that will figure prominently in discussions to follow is the Armed Forces Qualifying Test (AFQT), which is derived from the Armed Services Vocational Aptitude Battery (ASVAB). The AFQT is as highly correlated with IQ as IQ tests are with one another (Herrnstein & Murray, 1994, appen. 3), and it has been shown to measure the *g* factor well (Jensen, 1988; Ree, Carretta, & Doub, 1998/1999). Because any single test such as the AFQT can only approximate *g*, such tests *underestimate* *g*'s effects to some extent. Virtually all the estimates to follow therefore understate the impact of *g* for this reason as well.

Two Especially Important Research Projects. The Army's Project A (Campbell, 1990) and the Joint-Service Job Performance Measurement/Enlistment Standards (JPM) Project (Wigdor & Green, 1991) bear a detailed look because both used the expensive and hence rarely used gold standard in measuring job performance, namely, hands-on job-sample tests, and not just the inexpensive and hence much-used job knowledge tests and supervisor ratings. So, for example, the 4- to 8-hour hands-on tests might include having a naval machinist's mate respond to an alarm signaling loss of pressure in the main engine lube oil pump. These projects also systematically investigated the dimensionality of both performance criteria and their predictors, which had never before been done so systematically. Different criteria are predicted best by different personal traits, so understanding the relative value of *g* hinges on understanding the dimensionality of both performance and its determinants. The two projects are also especially pertinent because they were motivated by a practical intelligence perspective, in particular, by a concern that the military's seemingly *academic* selection tests might not actually predict workers' *practical* performance.

Both of these huge, interrelated projects had their origins in Congressional concern that the military services needed to improve and better validate their procedures for selecting and classifying recruits. At the time, the ASVAB had been validated only against performance in military training, not on the job. The new collective research effort required developing and evaluating multiple ways of assessing performance (including job samples, job simulations, job "walk-through" interviews, job knowledge tests, and ratings by self, peers, and supervisors) for a wide variety of occupational specialties (military police, jet engine mechanic, administrative specialist, rifleman, etc.) in each of the four services; developing a wider array of cognitive and non-cognitive predictors; and then validating all predictors (or subsets thereof) against all the performance criteria available for each service. The validation research was based on relatively large samples of recruits with longitudinal data.

A Crucial Distinction Among Outcome Criteria: Core Technical Versus Noncore "Citizenship" Dimensions of Performance. The JPM project found that its different performance criteria were far from perfectly correlated, even when corrected for unreliability. The median (uncorrected) correlation of hands-on (objectively measured) performance was .57 with job walk-through interviews (where workers describe how they would perform certain tasks), .47 with paper-and-pencil job knowledge tests, .37 with training (school knowledge) scores, and .26 with supervisor ratings (Wigdor & Green, 1991, tables on pp. 151–155; note that the four correlations tend to be for different sets of jobs). The four criteria therefore measure somewhat different aspects of performance, but job knowledge and training grades—the most academic criteria—share more in common with the practical gold standard than do supervisor ratings.

Army Project A systematically investigated the dimensionality of these criteria in 19 entry-level Army jobs via LISREL modeling of 32 criterion scores for 9,430 job incumbents (Campbell, McHenry, & Wise, 1990). The latent structure modeling yielded five factors: (1) core technical task proficiency (job-specific proficiency, such as an armor crewman starting and stopping a tank's engines), (2) general soldiering (proficiency in common duties, such as determining grid coordinates on military maps; Campbell et al., 1990, p. 322), (3) peer support and leadership, effort, and self development, (4) maintaining personal discipline, and (5) physical fitness and military bearing. Correlations among the five dimensions suggest that job performance tends to be divided into technical versus nontechnical dimensions (the first two vs. the last two factors, with the third being intermediate). The latter, noncore kinds of performance were measured mostly by ratings and are often characterized as the "citizenship" or "contextual" aspect of job performance (Organ, 1994).

Precisely because citizenship behaviors tend to be "extra-role, discretionary" behaviors that are *not* part of a job description (essentially, working above and beyond the call of duty), some researchers question whether they ought to be used as criteria in developing selection batteries (Borman & Motowidlo, 1993, pp. 93–94; Schmidt, 1993, p. 505). The noncore dimensions of performance are nonetheless relevant for our purposes here, because they affect people's lives by affecting both their supervisors' ratings and their job satisfaction (Borman & Motowidlo, 1993; Organ & Konovsky, 1989).

A Crucial Distinction Among Predictors: Cognitive Versus "Noncognitive" Predictors of Performance. The military services, like civilian employers, have been criticized in the past for perhaps relying too heavily on cognitive ability measures, especially paper-and-pencil tests, when selecting workers. The concern was that batteries of paper-and-pencil tests, such as the ASVAB, might measure academic abilities not actually relevant to the job while omitting useful noncognitive predictors. This would lead to weak selection batteries as well as to undue emphasis on academic talent. Army Project A addressed these concerns by developing an experimental test battery that measured not only a wide array of personal attributes that the ASVAB does not, but also specific mental skills that are manifestly perceptual–psychomotor rather than academic in content (e.g., choice reaction time and accuracy, target tracking). The trial predictor battery included 65 scales for a wide variety of personality, interest, and other nonacademic traits that an extensive review of the literature had identified as potentially useful predictors. Scores on these scales were combined, together with 4 ASVAB composites, to create a total of 24 predictor composites for 400–600 incumbents in each of nine high-volume Army jobs (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990, Table 2). The predictor composites fell into six categories: general cognitive ability (4 ASVAB composites), spatial ability (1 composite), perceptual–psychomotor abilities (6), temperament–personality (4), vocational interests (6), and job reward preferences (3). The first three categories are highly *g* loaded but the last three are not. There is much other research on the importance of noncognitive traits, some of which is discussed later, but Project A is still the largest and most thorough single study of the relative utility of cognitive and noncognitive traits.

A Crucial Question: Is* g *Causal? The answer is "yes." Ample research, particularly from the military services, shows that performance in training and on the job is correlated with mental ability assessed *before* entering training or the job. Mental ability also predicts job performance controlling for all other factors ever studied, the most important of which are examined shortly. There have also been large-scale quasi-experiments in which the emphasis on *g* in selecting workers was either increased or decreased (Schmidt & Hunter, 2000; see also the results of Project 100,000, Laurence & Ramsberger, 1991). Aggregate performance plummets when *g* is ignored, and it improves substantially in mid- to high-level jobs when *g* is weighted more heavily. The question among personnel researchers now is not whether *g* has a causal role, but instead how much, where, and why.

THE *g*-BASED THEORY OF JOB PERFORMANCE

The breadth and stability of the evidence in personnel selection psychology has led to the causal modeling of job performance (e.g., Hunter, 1983a, 1986; Hunter & Schmidt, 1996; Ree & Carretta, 1997; Ree, Earles, & Teachout, 1995). Figure 13.1 extracts the essence of this modeling. This model helps to explain *g*'s pattern of impact in both work and nonwork realms of life because it is, at heart, a *learning* theory and the need to learn is incessant in modern life (Hunter & Schmidt, 1996). *g* is important because it reflects the ability to learn (cf. Carroll, 1997). By this theory, job *performance* depends chiefly on job-specific *knowledge* that workers have learned either in training or through experience on the job. Differences in both knowledge and performance depend, in turn, on three kinds of differences among workers, summarized here as the "can do" (ability), "will do" (interest), and "have done" (training and experience) components of developed competence. All three precursors are important because they all affect the accumulation of job knowledge: the first affects workers' *rate* of learning from experience; the second, their *effort* to learn when given the opportunity; and the third, their *opportunity* to have learned. The one task attribute that shifts the relative importance of these person-precursors is task complexity. More complex jobs require more learning.

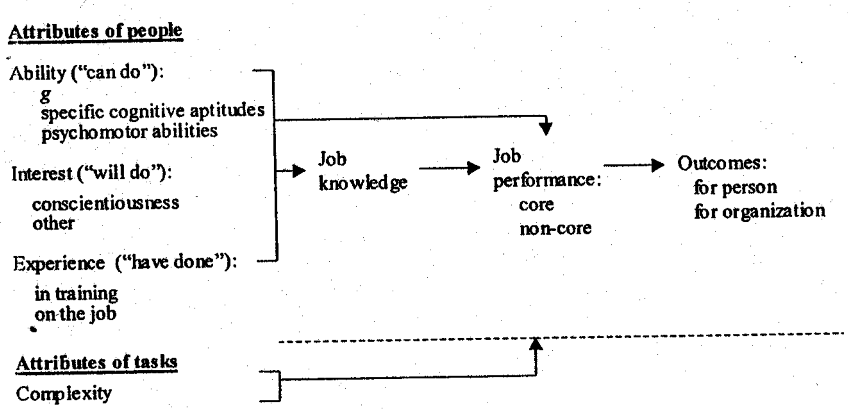


FIG. 13.1. Job performance model based on personnel selection research.

g is the predominant *can do* factor, and it affects job performance primarily via its strong effect on job knowledge. This is g's indirect effect, and it is analogous to *crystallized* intelligence. However, learning is not enough. Past learning never fully prepares people for the challenges they will face on the job (or in life). Many jobs (and life situations) require considerable adaptation and improvisation. Workers must spot and solve new problems that require going beyond merely applying old knowledge. More complex jobs impose more such demands because they are less "automatable." This is *g*'s *direct* effect on performance, and it is analogous to *fluid g*. The indirect effect of *g* on hands-on performance is at least twice as big as its direct effect in the typical job (e.g., Hunter, 1986).

This imbalance of indirect to direct effects is found for both *rated* performance and *actual* hands-on performance. The big difference in *g*'s prediction of the two criteria is that *g* has considerably stronger effects, both direct and indirect, on objectively measured performance than on supervisory ratings. Supervisors' perceptions of performance generally are only moderately responsive to either worker knowledge or worker performance, perhaps because few supervisors have much opportunity to actually observe their subordinates (Borman, White, Pulakos, & Oppler, 1991, p. 870). On the other hand, ratings are much more sensitive to worker personaliity traits that substantially affect employees' apparent "citizenship" but not their performance of core duties.

KEY DISCOVERIES ABOUT *g*'S GENERALITY

With this model as an organizing guide, the 18 major discoveries in Table 13.2 can be used to chart *g*'s degree of generality across different dimensions of work.

Discoveries 1–3a: *g* Has Full Generality Across All Jobs, Performance Criteria, Ability Levels, and Lengths of Experience

The first meta-analyses of job performance showed that *g*'s ability to predict job performance does not vary across time or place for different positions within the same or substantially similar job category (e.g., Hunter & Hunter, 1984, p. 80; Schmidt & Hunter, 1984; Schmidt, Hunter, & Pearlman, 1981; Schmidt, Hunter, Pearlman, & Shane, 1979). That is, criterion validities are not subject to the vagaries of situational specificity, as had long been thought. Since then, repeated meta-analyses have demonstrated even wider generality for *g*, in particular, that *g* is useful in predicting job performance across (1) the full range of jobs in the United States

TABLE 13.2
Major Findings on g's Impact on Job Performance[a]

Utility of g

1. Higher levels of g lead to higher levels of performance in all jobs and along all dimensions of performance. The average correlation of mental tests with overall rated job performance is around .5 (corrected for statistical artifacts).
2. There is no ability threshold above which more g does not enhance performance. The effects of g are linear: successive increments in g lead to successive increments in job performance.
3. (a) The value of higher levels of g does not fade with longer experience on the job. Criterion validities remain high even among highly experienced workers. (b) That they sometimes even appear to rise with experience may be due to the confounding effect of the least experienced groups tending to be more variable in relative level of experience, which obscures the advantages of higher g.
4. g predicts job performance better in more complex jobs. Its (corrected) criterion validities range from about .2 in the simplest jobs to .8 in the most complex.
5. g predicts the core technical dimensions of performance better than it does the non-core "citizenship" dimensions of performance.
6. Perhaps as a consequence, g predicts objectively-measured performance (either job knowledge or job sample performance) better than it does subjectively-measured performance (such as supervisor ratings).

Utility of g *relative to other "can do" components of performance*

7. Specific mental abilities (such as spatial, mechanical, or verbal ability) add very little, beyond g, to the prediction of job performance. g generally accounts for at least 85–95% of a full mental test battery's (cross-validated) ability to predict performance in training or on the job.
8. Specific mental abilities (such as clerical ability) sometimes add usefully to prediction, net of g, but only in certain classes of jobs. They do not have general utility.
9. General psychomotor ability is often useful, but primarily in less complex work. Its predictive validities fall with job complexity while those for g rise.

Utility of g *relative to the "will do" component of job performance*

10. g predicts core performance much better than do "non-cognitive" (less g-loaded) traits, such as vocational interests and different personality traits. The latter add virtually nothing to the prediction of core performance, net of g.
11. g predicts most dimensions of non-core performance (such as personal discipline and soldier bearing) much less well than do "non-cognitive" traits of personality and temperament. When a performance dimension reflects both core and non-core performance (such as leadership), g predicts to about the same modest degree as do non-cognitive (less g-loaded) traits.
12. Different non-cognitive traits appear to usefully supplement g in different jobs, just as specific abilities sometimes add to the prediction of performance in certain classes of jobs. Only one such non-cognitive trait appears to be as generalizable as g: the personality dimension defined by conscientiousness and integrity. Its effect sizes for core performance are substantially smaller than g's, however.

(Continued)

TABLE 13.2
(Continued)

Utility of g *relative to job knowledge*

13. g affects job performance primarily *indirectly* through its effect on job-specific knowledge.
14. g's *direct* effects on job performance increase when jobs are less routinized, training is less complete, and workers retain more discretion.
15. Job-specific knowledge generally predicts job performance as well as does g among experienced workers. However, job knowledge is not generalizable (net of its g component), even among experienced workers. The value of job knowledge is highly job specific; g's value is unrestricted.

Utility of g *relative to the "have done" (experience) component of job performance*

16. Like job knowledge, the effect sizes of job-specific experience are sometimes high but they are not generalizable.
17. In fact, experience predicts performance less well as all workers become more experienced. In contrast, higher levels of g remain an asset regardless of length of experience.
18. Experience predicts job performance less well as job complexity rises, which is opposite the trend for g. Like general psychomotor ability, experience matters least where g matters most to individuals and their organizations.

[a]See text for citations.

(e.g., Hunter & Hunter, 1984); (2) all levels of experience on the job (e.g., Schmidt, Hunter, Outerbridge, & Goff, 1988); and (3) all broad performance criteria, whether they emphasize *content* (core technical vs. noncore; McHenry et al., 1990) or *method* of assessment (paper-and-pencil tests, hands-on job samples, or ratings; Wigdor & Green, 1991).

The impact of g is also linear, that is, brighter job incumbents always perform better on the average, controlling for other factors (e.g., Coward & Sackett, 1990). There is evidence that additional increments in g are useful even at highest reaches of g and cultural achievement. Simonton (1994, chap. 8) reviews historiographic research indicating that the greatest Western composers, political leaders, and U.S. Presidents were brighter than eminent compatriots of lesser renown. He points out that greatness requires zeal and persistence too, but that additional intelligence enhances performance even at the highest levels of cultural achievement.

No meta-analysis has documented any limits to g's generality in predicting job performance. Effect sizes vary from small to very large, depending on the kind of job and performance criterion considered (as is discussed next), but they are never zero. In fact, average effect sizes are substantial: the average correlation between mental test scores and ratings of overall job performance hovers around .5 in broad collections of jobs (Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Discoveries 3b–6: g's Effect Sizes Vary by Job Complexity, Length of Experience, and Performance Criterion

Although higher levels of *g* are at least somewhat useful in all job circumstances, they are much more useful in some than others. This variability behaves lawfully, creating predictable gradients of effect size across the topography of work.

Job Complexity. The best known variation in predictive validities for *g* is that higher levels of *g* are a bigger advantage in more complex jobs. When Hunter (1983b; Hunter & Hunter, 1984) classified civilian jobs into five broad levels of complexity, average criterion validities for mental tests ranged from .23 in the simplest work (feeding/offbearing work) to .58 in the most complex (synthesizing/coordinating work). Almost 63% of the workforce held middle-complexity jobs. These jobs had an average criterion validity of .51 and included skilled blue collar occupations as well as midlevel white collar occupations. When narrower job families and more objective performance criteria are considered, validities range more widely—from about .2 in the simplest jobs to almost .8 in the most complex.

This complexity-related gradient of effect sizes is especially important because research in both sociology and psychology has shown that the major distinction among occupations in the U.S. economy is the complexity of their duties (e.g., Miller, Treiman, Cain, & Roos, 1980). The complexity dimension among jobs is highly correlated with their prestige or general social desirability. It is also moderately highly correlated to a job's criticality to the employer (Gottfredson, 1997, Table 7) and the dollar value of differences in worker performance in a job (Hunter, Schmidt, & Judiesch, 1990). In short, *g* tends to give the "biggest bang for the buck" in the jobs most highly valued by workers and their employers.

Job analyses indicate that *g* is more important in more complex jobs for reasons that *g* theory would predict: The key to a job's complexity is how much information processing the work demands. As seen in Table 13.3, these information-processing skills are prototypical of *g*. Compiling and combining information, advising, reasoning, planning, analyzing, and decision making all correlated at least .8 with job complexity level in a wide array of civilian jobs (Gottfredson, 1997, Table 7). This result mirrors Arvey's (1986, p. 418) earlier finding, also shown in Table 13.3, that the strongest correlates of complexity across jobs in the petrochemical industry (his "Reasoning and Judgment" factor) were requirements for dealing with unexpected situations (.75), learning and recalling job-related information (.71), reasoning and making judgments (.69), identifying problem situations quickly (.69), and reacting swiftly when unexpected problems occur

TABLE 13.3
Selected Correlates of Job Complexity

Task requirements	*Correlation (uncorrected)*
With Job "Complexity" factor: PAQ Job Analysis Data for 276 Broad Census occupations[a]	
Compiling information (importance of)	.90
Combining information (importance)	.88
Advising (importance)	.86
Writing (importance)	.86
Reasoning (level of)	.83
Planning/scheduling (amount)	.83
Analyzing (importance)	.83
Decision making (level)	.82
Negotiating (importance)	.79
Persuading (importance)	.79
Oral information (extent of use)	.68
Coding/decoding (importance)	.68
Instructing (importance)	.67
With "Judgment and Reasoning" factor: Analysis of 140 Jobs in Petrochemical Industry[b]	
Deal with unexpected situations	.75
Able to learn and recall job-related information	.71
Able to reason and make judgements	.69
Able to identify problem situations quickly	.69
React swiftly when unexpected problems occur	.67
Able to apply common sense to solve problems	.66
Able to learn new procedures quickly	.66
Alert and quick to understand things	.55
Able to compare information from two or more sources to reach a conclusion	.49

[a]From Gottfredson (1997, pp. 100–101), with permission of Elsevier Science. PAQ = Position Analysis Questionnaire.
[b]From Arvey (1986, p. 418), with permission of Academic Press.

(.67). These specific task requirements seem to defy classification as consistently academic, which practical intelligence theory would seem to require.

Another finding also seems to reflect the fact that greater complexity yields greater effect sizes for *g*. Namely, both experience and ability level predicted performance better in civilian than in military jobs of roughly comparable (moderate) complexity (Hunter, 1983a; Schmidt, Hunter, & Outerbridge, 1986). For example, the average correlations of mental ability with work sample performance and supervisor ratings were, respectively, .75 and .47 in 10 civilian jobs but .53 and .24 in 4 Army jobs (Hunter, 1983a). This civilian–military difference in effect sizes is thought to result from the military's more intense training and its greater insistence on following standard operating procedures. Both would reduce the g loadedness of military work, the first by reducing how much recruits must learn on their own after they start the job and the second by reducing their discretion in performing the job, that is, their opportunity to use their own judgment in deciding which problems to tackle and how (Schmidt et al., 1986, p. 433).

Experience. The skeptics of *g* might logically predict that its value would fade, especially in nonacademic jobs, as workers confront problems in the messy real world for which they were not specifically trained (e.g., Table 13.3's "dealing with unexpected situations" and "decision making"). Decaying criterion validities on the job, if confirmed, would be a matter of great concern in personnel selection. Hunter, Schmidt, and their colleagues (Hunter & Schmidt, 1996; Schmidt, Hunter, Outerbridge, & Goff, 1988) examined whether lengthier experience would, in fact, wash out the advantages of higher *g* for later job performance. In a meta-analysis of civilian work, they found that criterion validities for *g* actually *increased* with length of experience. The criterion validities for *g* ranged from .35 among incumbents with 0 to 3 years of job-specific experience to .59 for workers with an average of over 12 years of experience (Hunter & Schmidt, 1996). Hunter and Schmidt suggest that this increase is probably artifactual, however, at least for midlevel jobs, because the trend across experience categories disappears when differences in experience within the categories are controlled. A study of four Army jobs found that, when such differences in experience were controlled, *g*'s correlations with hands-on performance held steady at about .4 throughout the 5 years of experience for which the study had data (Schmidt et al., 1988, Table 6).

Effect sizes for experience, and their consequent ability to disguise the role of *g*, seem to increase when incumbents vary more widely in *relative* experience. When experience is not controlled, effect sizes for *g* increase as relative experience becomes more similar, as happens, for instance, when all incumbents have 10 to 12 rather than 0 to 2 years of experience. An additional year of experience makes a much bigger difference when the average tenure on a job is 1 year rather than 11.

What this means in pragmatic terms is that even lengthy experience does not compensate for below-average IQ. Differences in experience can hide but never nullify the value of higher *g*. Moreover, the advantages of higher *g* remain substantial in all but perhaps the simplest jobs. In 23 military jobs of moderate complexity, low-ability recruits (CAT IV, which is AFQT percentiles 10–30) with over 36 months of experience still performed notably worse than bright men (CAT I–II, which is AFQT percentiles 65–99) with only 0 to 12 months of experience (Wigdor & Green, 1991, pp. 163–164). When jobs are simpler than these, low-*g* workers may eventually catch up with sufficient practice, but when jobs are more complex, such workers are apt to be left hopelessly behind as their more able peers continue to advance in mastery.

Core Versus Noncore Dimensions of Performance (Army Project A Data). Most of the criterion validities discussed so far come from civilian studies, and they are based almost entirely on supervisor ratings of job performance. However, both Army Project A and the JPM research program developed a wide range of performance criteria, as described earlier. They found that *g*'s effect sizes were two to four times as large for some criteria as for others in these moderately complex military jobs. (Keep in mind that the effect sizes for military jobs may underestimate those for civilian jobs.)

Table 13.4 summarizes Army Project A's pertinent findings for nine jobs, whose average validities were all essentially the same (McHenry et al., 1990). Column 1 shows that general cognitive ability, as measured by the ASVAB, is a strong predictor of *core* task performance, whether the tasks be job-specific (core technical proficiency) or common across the variety of Army jobs (general soldiering proficiency). The predictive validities of general cognitive ability were .63 and .65, respectively, for these two proficiency criteria. Moreover, the *predictive* component of *g*, so measured, was also captured well by mental test composites that are *not* paper-and-pencil or even verbal in format, in this case, the set of computerized perceptual–psychomotor tests, which had concurrent validities, respectively, of .59 and .65 for the same two criteria (see column 3). Although these two sets of cognitive composites, the ASVAB and the experimental, are quite different in content and format, their predictive component must be essentially identical—and probably consists almost entirely of *g*—because the experimental cognitive battery adds only .01 to .04, net of the ASVAB, to the prediction of performance, no matter which of the five performance criteria are considered (compare columns 1 and 6). That small incremental validity is due almost entirely to the spatial composite in the experimental cognitive battery (compare columns 5 and 6).

The results for the noncore dimensions of performance are very different. The *g* factor, as represented in the ASVAB composites, is not nearly as strong a predictor of the *noncore, nontechnical* aspects of performance as it is the core technical aspects, whether job specific or general: the predictive validities are, respectively, .31, .16, and .20 for effort–leadership, personal discipline, and fitness–bearing (column 1). As the study authors suggest (McHenry et al., 1990, p. 352), *g*'s higher validity for effort–leadership than for the other two criteria may be due to the effort–leadership measure including ratings of core task performance. To summarize, Project A's results suggest that *g* has a small effect (about .2) on job-related self-control (the last two criteria), a moderate effect on being an effective team

TABLE 13.4
The Abilities of Different Cognitive and Noncognitive Ability Composites to Predict 5 Dimensions of Job Performance: Army Project A Data for 9 Mid-Level Jobs

	Predictor Sets[a]							
	ASVAB	Project A's New Composites			ASVAB Plus New Composites			All
Dimension of Job Performance	General Cognitive Ability Composites (K = 4) (1)	New Spatial Composite (K = 1) (2)	7 New Cognitive Composites (Includes spatial) (K = 7) (3)	13 New Temperament/ Interest/Reward Preferences (TIR) Composites (K = 13) (4)	New Spatial Composite plus 4 ASVAB Composites (K = 5) (5)	7 New Cognitive Composites Plus 4 ASVAB Composites (K = 11) (6)	13 New Temperament-Interest-Rewards (TIR) Composites Plus 4 ASVAB Composites (K = 17) (7)	(K = 24) (8)
Core technical proficiency	.63	.56	.59	.44	.65	.65	.65	.67
General soldiering	.65	.63	.65	.44	.68	.69	.67	.70
Effort and leadership	.31	.25	.27	.38	.32	.32	.43	.44
Personal discipline	.16	.12	.13	.35	.17	.17	.37	.37
Physical fitness and military bearing	.20	.10	.14	.38	.22	.23	.41	.42

Source: From McHenry et al. (1990, Tables 4-7). Reprinted with permission of *Personnel Psychology*, copyright (1990).
[a]Multiple R's were corrected for range restriction and adjusted for shrinkage. K = number of composites.

leader or team member (.3), and a very strong effect on the core technical performances for which workers bear individual responsibility (.6).

Objective Versus Subjective Measures of Performance (JPM Project Data). Predictive validities from the JPM project flesh out the picture of the relation of *g* to different performance criteria. The Project A core proficiency criteria (job-specific and general) were based on a combination of job knowledge and hands-on proficiency tests. In contrast, the JPM project kept these two types of criteria separate when examining the predictive validity of *g*. It thus provides a more direct test of the skeptics' hypothesis that *academic* ability tests (such as the ASVAB) do not predict *practical* (hands-on) performance well, only other paper-and-pencil performances (such as job knowledge tests). The JPM reports that the median (uncorrected) correlation of hands-on (job-sample) performance with the AFQT was .38 for the 23 jobs studied across the four services (Wigdor & Green, 1991, p. 161). Predictive validity was as high in the Army and Marine jobs, few of which could be construed as academic, as it was in the Air Force and Navy jobs.

AFQT (uncorrected) predictions of hands-on performance in the four Marine jobs reinforce the point that the supposedly academic AFQT actually predicts performance on practical as well as academic tasks: rifleman (.55), machinegunner (.66), mortarman (.38), and assaultman (.46; Wigdor & Green, 1991, p. 161). These jobs are hardly the picture of esoteric academic work. To illustrate that the criterion tasks actually chosen for measurement were not mostly school-like, the two tasks that were most highly correlated with overall hands-on performance among rifleman were "land navigation" and "live rifle fire" (Wigdor & Green, 1991, p. 148).

Hunter (1986, p. 352) summarized data showing, further, that the predictive validity of *g* is higher for objectively measured than subjectively measured job performance, regardless of whether the objective measures are paper-and-pencil (academic) or hands-on (practical). *g* correlated .80 and .75, respectively, with job knowledge and work sample performance but only .47 with supervisory ratings in 10 civilian jobs, and it correlated .63 and .53 with the two objective measures but only .24 with ratings in four military jobs. As noted earlier, skeptics of *g* had long predicted the opposite pattern, based on the mistaken assumption that supervisors are unduly impressed by intelligence.

Discoveries 7–9: Generality of *g* is High Relative to Other Ability ("Can Do") Factors

Assessing *g*'s practical value requires knowing how large its effects typically are *relative* to other personal traits that might also create differences in task performance. Meta-analyses have been consistent in showing, not only that *g* has very general value for job performance, but also that narrower mental abilities (such as verbal, spatial, and mechanical aptitude) do *not* have general value after removing their large *g* component. Examples would include the clerical, mechanical, and electrical composites of the ASVAB. Tests of specific aptitudes seldom predict performance as well as does *g*, and they generally predict only to the extent that they measure *g* (Hunter, 1983c, 1983d; Schmidt, Ones, & Hunter, 1992). They seldom add more than .01 to .03 to the prediction of job performance beyond *g*, no matter how performance is measured, as is illustrated by spatial ability in Table 13.4 (compare columns 1 and 5) for Army Project A data. Such weak incremental validity is a consistent finding from research for the other military services (Ree, Earles, & Teachout, 1994) and civilian jobs (Schmidt et al., 1992, pp. 646–647). The finding should not be surprising in view of the moderately high correlations among all mental abilities (their "positive manifold"), which means that once a mental test's *g* component is removed, it retains little with which to predict anything. The only meta-analytically-derived exception to the .03 ceiling so far has been for speeded clerical tests in clerical work. They are among the least *g*-loaded mental tests, which gives them greater opportunity to add to criterion validity beyond what *g* contributes, but even here the addition is small (e.g., from .64 to .68 in Hunter, 1985, p. 15). The typical finding is that an aptitude composite that is tailored for one family of jobs (say, mechanical) predicts performance about equally well in all others (say, clerical or general technical; e.g., Hunter, 1985, 1986, p. 357).

In fact, the *g* factor always carries the freight of prediction in any full battery of mental tests. Thorndike's (1986) systematic analysis of the issue is particularly illuminating. Criterion validities for entire aptitude batteries, such as the U.S. Employment Service's General Aptitude Test Battery (GATB), are often higher than those for the *g* factor alone, although *g* always accounts for the lion's share of a battery's validity. Thorndike's special contribution was to calculate how superior (or inferior) a battery is to *g* alone after cross-validating the battery's prediction equations in new samples (in order to eliminate the capitalization on chance that occurs in deriving a prediction equation, which capitalization increases with the number of tests in a battery). The apparent superiority of batteries whose prediction equations are tailored to specific school subjects, sexes, or jobs is much reduced or disappears with cross-validation. In two large samples, *g* yielded 85% to 95% of the criterion validity of the cross-validated aptitude batteries for predicting grades in high school and military training. In small samples of incumbents from various jobs, a single *g* factor predicted job performance *better* than did the cross-validated GATB prediction equations developed for each job.

The point is not that *g* is the only mental ability that matters. It is not. Rather, the point is that no other well-studied mental ability has average effect sizes that are more than a tiny fraction of *g*'s (net of their large *g* component), and none has such general effects across the world of work. Narrower aptitudes, such as verbal, spatial, and quantitative ability, may make special contributions to core job performance in some jobs, net of *g*, but—as with clerical speed—they would contribute only in limited domains of activity. As argued earlier, generality is gauged by the variety of tasks in which an aptitude enhances performance. Special aptitudes have quite circumscribed generality, net of their *g* component.

A largely *non*-mental ability—general psychomotor ability (which includes eye–hand coordination and manual dexterity)—is the only ability that meta-analyses have shown to be general and also have effect sizes that sometimes exceed those of *g* (Hunter & Hunter, 1984). As with *g*, its effect sizes vary systematically with job complexity—but in the *opposite* direction: Criterion validities of psychomotor ability *fall* from .48 to .21 as those for *g* rise from .23 to .58 across Hunter's five levels of increasing job complexity (Hunter & Hunter, 1984, p. 83). In other words, the general psychomotor factor tends to provide the biggest competitive advantages in the lowest level, least attractive jobs.

Discoveries 10–12: Generality of *g* is High Relative to Interest ("Will Do") Factors

Personnel researchers have devoted keen attention lately to personality traits and vocational interests because they are correlated little or not at all with either *g* or race and therefore hold out the greatest hope of improving the prediction of performance while simultaneously reducing disparate impact. Meta-analyses for vocational interests reveal very low validities for predicting supervisor ratings—.10 (Schmidt & Hunter, 1998, Table 1). Army Project A's vocational interest composite predicted core performance much better (.35), but the authors note that their interest composite behaved more like a test of cognitive ability than like one of temperament and personality (McHenry et al., 1990, Table 4 & pp. 351–352).

Of the "big five" personality factors (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience), only conscientiousness and its variants seem to have general validity across the world of work. However, that validity is substantial: .31 for tests of conscientiousness and .41 for tests of integrity. Meta-analyses show that these tests add, respectively, .09 and .14 to the prediction of supervisor ratings, beyond *g*, to yield multiple Rs of .60 and .65 (Schmidt & Hunter, 1998, Table 1). Conscientiousness and integrity capture both the willingness to expend the effort to learn and work harder (which enhances core knowledge and performance) and the citizenship behaviors that impress supervisors, regardless of a worker's core performance.

Other personality traits have been found useful in predicting performance in particular jobs or job families, but they appear not to have general utility (Hogan, 1991, p. 898). Their value is more local, that is, more specific to the job family in question, such as sales or management. In short, conscientiousness–integrity has broad importance, but no noncognitive trait rivals *g* in *both* generality and effect size.

Although restricted primarily to midlevel jobs, the military research provides a more systematic confrontation between cognitive and noncognitive predictors of performance. (Putatively *non*-cognitive factors such as Army Project A's are actually only *less* cognitive—less *g* loaded—than are mental tests. The *g* loadings of so-called noncognitive tests are seldom ascertained, however.) The research is more systematic because it includes a wide array of both criteria and predictors, something which civilian research seldom if ever does. As described earlier, Army Project A developed three sets of noncognitive predictor composites: temperament/personality (T), vocational interests (I), and job reward preferences (R). All are measured by paper-and-pencil, multiple-choice inventories. The TIR composites represent dispositions or motivations to perform different tasks. They thus reflect primarily the willingness ("will do") rather than the ability ("can do") to perform a job or task well. As shown in Table 13.4, the TIR composites were much weaker than *g* in predicting core performance, whether job specific or general (compare column 4 to 1). They added virtually nothing, whether singly or in combination, to the explanation of core performance beyond that afforded by general cognitive ability (column 7 vs. 1). On the other hand, together they greatly outperformed *g* in predicting discipline (.35 vs. .16, column 4 vs. 1) and fitness/bearing (.38 vs. .20), indicating that performing well along these lines may be primarily a matter of motivation rather than ability. General mental ability added very little (.02–.03) to their prediction (column 7 vs. 4). In contrast, performance on the effort/leadership criterion was about equally responsive to both will do and can do factors, although slightly more to the former (.38) than to the latter (.31, column 4 vs. 1). This is the only criterion out of the five on which *both* the *can do* and *will do* traits were necessary for distinguishing better from worse performers.

Note that the criterion validities for the TIR composites, which range from .35 to .44 for the five performance criteria (column 4 of Table 13.4), are comparable to those mentioned earlier for conscientiousness (.31) and integrity (.41). In fact, much of the predictive value of the TIR composites

lay in their measurement of dependability (McHenry et al., 1990, pp. 344 and 349). The Army and civilian results thus seem quite consistent regarding the relative utility of *g* versus noncognitive traits.

In summary, the noncognitive traits add little to the prediction of core technical performance but may be superior predictors of noncore discretionary behaviors. It would seem, then, that *g*'s effects tend to be high at the core of job performance but fade toward the periphery, whereas the opposite is true for noncognitive traits.

Discoveries 13–15: Generality of *g* is High Relative to Job Knowledge

As discussed earlier, *g* is a good predictor of job knowledge; correlations are generally around .5 to .8 (Hunter, 1986, p. 352; Schmidt & Hunter, 1998, p. 265). Job knowledge, in turn, is the chief precursor of hands-on performance, with correlations somewhat higher. As noted, *g* influences performance primarily indirectly via job knowledge, although its (smaller) direct effects seem to increase when jobs are more complex and less routinized, training is less complete, and incumbents retain more discretion.

Job knowledge is sometimes viewed as a performance criterion, one which is intermediate between *g* and hands-on performance or supervisor ratings. However, it can also be viewed as a competitor to *g* in predicting these criteria. It should come as no surprise that it can outpredict *g* among experienced workers, because job-specific knowledge—*expertise*—is a function of both ability and experience (e.g., Hunter, 1986, p. 352). The correlations of job knowledge and *g* with hands-on performance are, respectively, .80 versus .75 (10 civilian jobs) and .70 versus .53 (4 Army jobs); and with supervisor ratings they are .56 versus .47 (civilian) and .37 versus .24 (Army).

However, job knowledge is not general, because it is always specific to a job or occupation. Although we can use one test of *g* to predict performance in all jobs, there must be as many job knowledge tests as there are jobs or job families, because all jobs are by definition comprised of different core duties with different content to be learned. A knowledge test that does not cover required core knowledge is not "content valid." Knowledge tests are thus a general *strategy* for assessing job competence, but their *content* must always be specific to a job or job family. Moreover, they are suitable only for persons who are already trained or experienced. By necessity, they are highly local—unlike *g*, which crosses all boundaries of content. To the extent that job knowledge is general at all, it is because it measures *g*, the facility with which people have learned that knowledge.

The *g*-based theory of job performance would predict a moderately strong correlation between job knowledge and hands-on performance in mid- to high-level jobs, because complexity of information processing is the most distinctive feature of those jobs and because *g* is the ability to process information. *g* is useful not only in learning jobs, but also in solving the new problems they continually present, especially when the jobs are highly complex. Recall, however, that the massive military studies were prompted by the opposite expectation—that simply *knowing*, especially when assessed by paper-and-pencil methods (*academic* knowledge), might not be much related to actually *doing* (*practical* action). The research has now shown a strong connection between the knowing and doing: *g* and *g*-based knowledge are highly practical. *g* theory was correct.

Discoveries 16–18: Generality of *g* is High Relative to Experience ("Have Done") Factors

Experience functions somewhat like *g* in enhancing job performance: It leads to more learning and thus more job knowledge, which in turn increases both hands-on performance and supervisor ratings. However, it is a weak competitor to *g*, even when its effect sizes are comparable. Like job knowledge, it is content valid only when it is job specific. It is not general. There are two other major differences between experience and *g*.

First, higher *g* always leads to better average performance, as was noted earlier, but lengthier experience does not. Specifically, further analyses of the four midlevel Army jobs found that average absolute levels of performance stop rising by the time incumbents have been on a job about 5 years, meaning that differences in experience lose much of their predictive validity after 5 years, at least for midlevel military jobs. In contrast, *g*'s validity remains strong (Schmidt et al., 1986, p. 436; see also McDaniel, Schmidt, & Hunter, 1988). In a meta-analysis of civilian jobs, *g*'s correlation with supervisory ratings rose, as noted earlier, from .35 for workers with 0 to 3 years of experience to .59 for workers with more than 12 years of experience, but the validities for experience *dropped* from .49 to .15 (Hunter & Schmidt, 1996, Table 5). Only in the least experienced group (0–3 years) did experience outpredict *g*; this is the average level of experience at which *relative* disparities in experience are typically at their maximum.

The second difference between experience and *g* is that the criterion validities for experience are higher in *less* complex jobs (as is the case for psychomotor ability too), which is opposite the pattern for *g*: they average .39 in low complexity jobs and .28 in high complexity jobs (McDaniel et al., 1988, Table 2). The reason for this inversion is probably that workers in low-level jobs receive little formal training and therefore must learn their jobs mostly through experience once on the job (McDaniel et al., 1988, p. 330).

Job-specific experience may be essential to good performance in most jobs, but differences in experience are generally useful in distinguishing better from worse workers only in the first few months or years on the job. In contrast, differences in *g always* matter. Their impact can, however, be concealed by differences in experience, especially in groups including people with almost no job-specific experience.

SUMMARY FOR JOB PERFORMANCE

Ample evidence has shown that *g* predicts core job performance widely and well, overall. The personality dimension of conscientiousness–integrity may possibly rival *g* in generality, but not in average effect size. Psychomotor ability may have both generality and occasionally big effect sizes, but it matters least where *g* matters most—in complex jobs. These are precisely the prestigious jobs for which individuals compete most avidly and whose good performance is most critical to the organization. Job knowledge and experience are important, but they are highly specific. They are general only to the extent that they tap *g* or the opportunity to use it in order to learn a job's essentials.

Effect sizes for *g* vary systematically according to the complexity of tasks, but not whether they seem academic (e.g., well defined, have only one right answer) or practical (e.g., require diagnosing and solving ill-defined problems). Indeed, the so-called academic attributes in Table 13.1 do not apply even to many school subjects, from history to science, which often involve ill-defined problems with more than one solution or means to it. What at least five of the seven supposedly academic attributes represent are actually strategies for creating *test items* that will be reliable (e.g., items have a single correct answer) and unbiased (e.g., tasks are disembedded from personal experience and require no outside knowledge), and thus more valid. Practical intelligence theorists have confused *how* tests measure traits well with *what* traits they measure.

There is no body of evidence even remotely comparable to that for *g* that there exists any other highly general "intelligence," if by intelligence we mean a primarily mental ability. Gardner, for instance, has not yet even measured his eight "multiple intelligences" (Gardner, 1997), let alone shown them independent of *g* or able to predict anything. Research on *practical intelligence*, as described by Sternberg and his colleagues (Sternberg et al., 2000), has been limited to looking at the concurrent validity of "tacit knowledge" in school and work settings. Their five studies (8 samples) relating tacit knowledge to job performance in the civilian sector (they also have one study on three samples of Army officers) have focused on only four narrow and mostly high-*g* occupations (academic psychologists, business managers, bank managers, and life insurance salespeople) and relied on small samples (average $N = 55$) and mostly careerist rather than core technical performance criteria (e.g., reputation, salary; see Gottfredson, in press-a, for a detailed critique). In any case, that research program has provided far too few correlations for a meta-analysis of job performance and therefore cannot support its claim that practical intelligence "is at least as good a predictor of future success as is the academic form of intelligence [*g*]" (Sternberg et al., 2000, pp. xi–xii; also Sternberg, 2000).

Moreover, that program has provided no data at all for performance in everyday activities outside school and work. In contrast, as we see next, large bodies of evidence show that high *g* provides tremendous advantages throughout the breadth and length of life, and that low *g* constitutes a very practical, very pervasive disadvantage.

g and Life Performance

No other personal trait has been shown to correlate with so many valued outcomes as has *g*. The outcomes include altruism, breadth of interests, educational attainment, emotional sensitivity, leadership, moral reasoning, motor skills, musical abilities, occupational success, social skills, and much more (Brand, 1987). *g* also correlates, negatively, with a wide range of problem behaviors, including accident-proneness, delinquency, dogmatism, racial prejudice, smoking, and truancy. The correlations range from strong to weak, but seem pervasive. (Recall that correlations in this section will be *uncorrected* unless otherwise noted.) Although *g*'s generality would thus seem to be quite broad, there is yet little systematic mapping of *g*'s gradients of effect across social life. The job performance model provides a start, because the individual duties in a job are analogous to individual tasks in life. In addition, jobs themselves (limited but somewhat fluid subsets of duties) are analogous to people's lives (the reasonably circumscribed ebb and flow of activities and challenges over a person's lifetime).

RISKS IN MANAGING INDIVIDUAL LIFE TASKS

Many work activities are life tasks as well: managing people and money, selecting products and paying bills, preventing and responding to accidents, driving, teaching, and the like. Many of these activities require the complex information processing skills in Table 13.3, including advising, reasoning, planning, analyzing information, negotiating, persuading, coordinating, and instructing—not to mention "dealing with unexpected situations" and "applying common sense to solve problems." Because higher levels of *g* enhance performance of jobs in which these are key duties, they can be expected to enhance performance of analogous tasks in daily life. There is indeed evidence, for instance, that higher *g* is advantageous in driving. A longitudinal study of Australian servicemen, none of them in the retarded range, found that IQ correlated with rate of death by automobile accident, even after controlling for other characteristics (O'Toole, 1990). The auto fatality rate for men of IQ 85 to 100 (92.2 per 100,000) was double that for men of IQ 110 to 115 (51.5 per 100,000). The rate for men of IQ 80 to 85 was three times as high (146.7 per 100,000).

The discoveries about *g*'s impact on job performance (Table 13.2) provide a roadmap for collecting, classifying, and interpreting data on *g*'s gradients of effect in everyday affairs. They suggest, most importantly, that *g* will be useful wherever information processing is required, and that its impact will be highest when tasks are complex or novel and lower when they are simple, routine, repetitious, much practiced, and supervised. For instance, they predict that higher levels of *g* will be a bigger advantage (controlling for experience) when driving an unfamiliar route during rush hour or bad weather than when driving a familiar route when traffic is light and the weather good. The research thereby also suggests that knowing more about the distribution of task demands and opportunities across daily life can speed our understanding of *g*'s generality in practical affairs.

A Matrix of Life Tasks

The two important dimensions of tasks discussed earlier, and shown here in Fig. 13.2, are their complexity and whether they entail mostly technical matters rather than citizenship. The latter distinction seems to correspond roughly to the difference between *instrumental* and *socioemotional* tasks. The job performance validity data suggested that *g*'s impact is highest for complex technical tasks and that it drops gradually as tasks become simpler or more socioemotional, in which case they depend increasingly on "will do" (personality) factors such as conscientiousness. It is not clear, however, that the advantages of greater mental competence ever fall to zero either on the job or in the myriad details of managing one's way in life.

Research on individuals' adaptation and adjustment illustrate how this matrix can illuminate *g*'s effects in daily life as well as paid employment. As described next, *adaptation* probably falls more toward the instrumental side of the matrix than does *adjustment*, and its correlation with *g* is correspondingly higher. The latter, personal adjustment, is measured in many ways, but refers to "a complex of behaviors involving such features as emotional stability, freedom from neurotic symptoms, responsibility, getting along with people, social participation, realistic self-confidence, absence

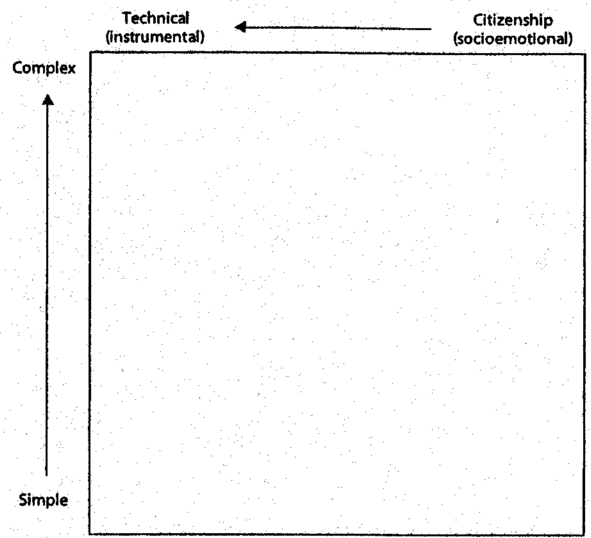


FIG. 13.2. The matrix of life tasks.

of socially disruptive and self-defeating behavior, . . . [and much more, including] displaying a capacity for self-discipline and planful and sustained goal-directed effort" (Jensen, 1980, pp. 357–358). This seems similar to the "citizenship" dimension of job performance, which is very much dependent on "will do" (personality) traits. However, adjustment still correlates .4 to .6 with IQ, even when IQ is measured in childhood and adjustment in adulthood (Jensen, 1980, p. 357).

Adaptive behavior is similar to adjustment, but emphasizes instrumental competence more heavily relative to emotional adjustment than does the concept of adjustment. Typical measures focus, for instance, on "everyday skills such as handling money, personal care and hygiene, telling time, domestic skills, ability to go shopping alone, and the like" (Jensen, 1980, p. 358). Most adaptive behavior measures have been developed in the context of diagnosing mental retardation, but they correlate .6 to .7 with IQ in normal children. In short, socioemotional outcomes may depend less on *g* than do more strictly instrumental ones, but both seem to be enhanced by higher *g*. There is also a sizeable literature on competence in late adulthood (Diehl, 1998), when mental powers fade, which also seems to reflect this instrumental–socioemotional gradient in effect sizes for *g*. Studies show that "everyday problem-solving" late in life correlates from .3 (Cornelius & Caspi, 1987) to .8 with fluid *g* (Willis & Schaie, 1986), with the latter studies using measures that seem to more closely represent the instrumental side of the life-task matrix.

Other socioemotional tasks and behaviors may be simpler (not require much information processing) than those captured in research on adjustment, and therefore depend little on *g* but a lot on personality traits such as sociability and dependability. These would be located to the lower right of the matrix. So, although the interpersonal task of persuading people may be moderately to highly *g* loaded (see Table 13.3), being a pleasant and dependable coworker or companion may not. And just as the latter can increase one's ratings as a worker, being a "good citizen" can gain one regard and resources elsewhere in life regardless of one's mental competence. Indeed, they may be especially important for retarded individuals, whose successful adjustment often depends on good will and regular help from others (Edgerton, 1981; Koegel & Edgerton, 1984).

Performance of life tasks throughout this matrix would often depend, of course, on more than "can do" and "will do" factors." For instance, "have done" factors (previous training and experience) can also affect performance, the prior evidence suggesting that their relative advantages are greatest when tasks are simple and one's peers have little or no such experience. Help from family, friends, and organizations matter, too, as do other aspects of people's circumstances (Gottfredson, in press-b), but they are beyond the scope of this chapter.

Functional Literacy

The literature on functional literacy is especially illuminating because it has examined the relation of everyday competence to various task characteristics, finding the academic–practical distinction irrelevant but complexity crucial. The study of literacy in recent decades has been driven by a growing concern that a high proportion of Americans is unable to meet many of the daily demands of modern life that most of us take for granted. *Functional literacy* refers, most broadly, to using the written word effectively (literacy) for everyday (functional) purposes. Literacy research focuses on practical tasks, and ignores merely *academic* reading, which does not affect understanding of or effective adaptation to one's options and circumstances in life. By focusing on reading, which is a solitary technical activity, functional literacy research targets tasks primarily to the left (technical) side of the matrix in Fig. 13.2.

There are three independent research literatures on functional literacy, referred to as work literacy (Sticht, 1975), adult literacy (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993), and health literacy (National Work Group on Literacy and Health, 1998). They involve reading and understanding prose, documents, forms, and other written materials, respectively, at work, in everyday activities, and in the health system. All three fields of research began with the same belief in situational specificity, in particular, that there are as many types of literacy as there are major types of written material. All three have therefore taken pains to develop scales whose items simulate tasks in the relevant settings: for example, their tests have workers read the tables of contents of their technical manuals, citizens read bus schedules and menus, and patients read the labels on vials of prescription medicine. Like personnel psychologists, then, literacy researchers started out skeptical of, even hostile to, anything like *g* theory. Also like personnel researchers, their findings have nonetheless turned out to support *g* theory. They do so by revealing that literacy is the ability to process information (*g*) and that it has enormous practical importance.

Although largely independent in personnel, aims, and methods, all three literatures tell the same story. All initially sought to measure multiple dimensions of literacy, but their evidence reveals that literacy is unidimensional. For instance, the three scales (Prose, Document, and Quantitative) comprising the U.S. Department of Education's National Adult Literacy Survey (NALS) correlate over .9 *before* correction for unreliability despite the developers' effort to create three independent scales. As a result, "major [NALS] survey results are nearly identical for each of the three scales . . . with findings appearing to be reported essentially in triplicate, as it were" (Reder, 1998, pp. 39, 44).

Work literacy and health literacy researchers have both investigated whether delivering information orally rather than in written form might enhance comprehension, but they have found that "poor readers" comprehend information no better when they listen to it instead. Army research on the matter was quite extensive (Sticht, 1975, chap. 7). The NALS adult literacy researchers also did a detailed task analysis showing that the difficulty of NALS items does not depend on their readability per se (Kirsch, Jungeblut, & Mosenthal, 1994). As has been shown many times (e.g., Jensen, 1980, pp. 325–326), "reading ability" for native speakers is far more than decoding skill. Rather, it is comprehension. Speech and written symbols are just different vehicles for transmitting information.

All three literatures also eventually identified complexity as the crucial distinction among literacy tasks. The NALS researchers did an excellent analysis in which they detailed the task characteristics that account for the relative difficulty of NALS items. Described as "process complexity," the attributes include the abstractness of information, its embeddedness in lengthy or irrelevant information, and the difficulty of the inferences it requires (Kirsch & Mosenthal, 1990). These results are summarized in Table 13.5, which illustrates both the items and their information-processing demands at each of the five NALS literacy levels.

Work and health literacy researchers have not performed any formal task analyses of their scales, but they have focused on making everyday materials more readable for poor readers by reducing their complexity

TABLE 13.5
Sample Items and Information-Processing Demands at Five Levels of NALS Literacy

Proficiency Level	*Sample Items*[a]	*Information-Processing Demands*[b]
0		
>	69 Sign your name (D)	
>		*Level 1* (NALS $\leq$225) tasks require identifying or matching single pieces of information or performing a single, simple, specified arithmetic operation (like addition) in contexts where there is little or no distracting information. (Includes about 14% of white and 38% of black adults aged 16 and over.[c])
	191 Total a bank deposit entry (Q)	
	224 Underline sentence explaining action stated in short article (P)	
225		
	232 Locate intersection on a street map (D)	
		Level 2 (NALS 226–275) tasks introduce distractors, more varied information, and the need for low-level inferences or to integrate two or more pieces of information. Information tends to be easily identifiable, despite the presence of distractors, and numeric operations are easily determined from the format of the material provided (say, an order form). (Includes about 25% of white and 37% of black adults.)
	250 Locate two features of information in sports article (P)	
	270 Calculate total costs of purchase from an order form (Q)	
275		
	280 Write a brief letter explaining error made on a credit card bill (P)	
		Level 3 (NALS 276–325) tasks require integrating multiple pieces of information from one or more documents, which themselves may be complex and contain much irrelevant information. However, the matches to be made between information and text tend to be literal or synonymous, and correct information is not located near incorrect information. (Includes about 36% of white and 21% of black adults.)
	308 Using calculator, determine the discount from an oil bill if paid within 10 days (Q)	

(Continued)

TABLE 13.5
(Continued)

Proficiency Level	*Sample items*[a]	*Information-Processing Demands*[b]
	323 Enter information given into an automobile maintenance record form (D)	
325		
	328 State in writing an argument made in lengthy newspaper article (P)	
	348 Use bus schedule to determine appropriate bus for given set of conditions (D)	*Level 4* (NALS 326–375) tasks require more inferences, multiple-features matches, integration and synthesis of information from complex passages or documents, and use of multiple sequential operations. (Includes about 21% of white and 4% of black adults.)
	368 Using eligibility pamphlet, calculate the yearly amount a couple would receive for basic supplemental security income (Q)	
375		
	387 Using table comparing credit cards, identify the two categories used and write two differences between them (D)	
	410 Summarize from text two ways lawyers may challenge prospective jurors (P)	*Level 5* (NALS 376–500) tasks require the application of specialized background knowledge, disembedding the features of a problem from text, and drawing high-level inferences from highly complex text with multiple distractors. (Includes about 4% of white and less than 0.5% of black adults.)
> > 500	421 Using calculator, determine the actual cost of carpet to cover a room (Q)	

[a]*Source:* Brown, Prisuta, Jacobs, & Campbell (1996, p. 10). P = prose scale, D = documents scale, Q = quantitative scale.
[b]*Source:* Brown et al. (1996, p. 11).
[c]*Source:* Kirsch, Jungleblut, Jenkins, & Kolstad (1993, Table 1.1A). Percentages are for Prose Scale.

(Doak, Doak, & Root, 1996; Sticht, 1975, chap. 7). Their methods, separated by decades and specific purpose, are strikingly similar and include limiting information to the bare essentials and then providing it in a concise, standardized manner that breaks down all complex material into small, carefully sequenced, concretely illustrated chunks that tell people exactly what to do. Health literacy specialists suggest that materials be written at the fifth-grade level, although they concede that this is beyond the capabilities of a substantial proportion of urban and elderly patients, even when they average 10 years of schooling.

Discovering that neither task content nor modality affect literacy demands but that complexity is key, analysts turned to the language of learning, problem solving, and critical thinking (although without actually mentioning *intelligence* or *g*) to describe the meaning of literacy. NALS analysts describe adult literacy in terms such as "problem solving," "complex information processing," and "verbal comprehension and reasoning, or the ability to understand, analyze, interpret, and evaluate written information and apply fundamental principles and concepts" (Baldwin, Kirsch, Rock, & Yamamoto, 1995, p. xv; Venezky, Kaestle, & Sum, 1987, pp. 25, 28). Health literacy researchers now suggest that health literacy is the "ability to acquire new information and complete complex cognitive tasks," and that low literacy reflects "limited problem-solving abilities" (Baker, Parker, Williams, & Clark, 1998, pp. 795–797).

The three fields have shown that their seemingly different literacies all mimic *g*. Two of the fields have specifically correlated literacy with tests known to be highly *g*-loaded. Work literacy, as intensively studied several decades ago by the Army, is measured well by the AFQT, its (uncorrected) correlations with job-specific work literacy being .6 to .8 (Sticht, 1975, pp. 46–48, 75). Although the data are less extensive for measures of health literacy, the various health literacy scales also correlate .7 to .9 with each other and with tests of known high *g* loading (Davis, Michielutte, Askov, Williams, & Weiss, 1998), such as the Wide Range Achievement Test (WRAT). "Literacy" appears to be a surrogate measure of *g*.

Impact of Adult Literacy. The adult literacy scales reveal the stark meaning of low *g* in many daily activities, because the scales are criterion-referenced. Table 13.5 samples the tasks that people at each of the five NALS literacy levels are able to perform on a routine basis (with 80% proficiency). The longer list from which they are drawn (Brown, Prisuta, Jacobs, & Campbell, 1996, p. 10) shows that they represent skills needed to carry out routine transactions with banks, social welfare agencies, restaurants, the post office, and credit card agencies; to understand contrasting views on public issues (such as parental involvement in schools); and to comprehend the public events of the day (such as stories about sports and fuel efficiency) and one's personal options (welfare benefits, discount for early payment of bills, and so on). As shown in the table, almost 40% of White adults and 75% of Black adults are able to function routinely at no higher than Level 2, which limits them to making only low-level inferences and integrating only a few pieces of clearly identifiable information—a clear disadvantage in the Information Age. In fact, NALS researchers note that individuals at Levels 1 or 2 "are not likely to be able to perform the range of complex literacy tasks that the National Education Goals Panel considers important for competing successfully in a global economy and exercising fully the rights and responsibilities of citizenship" (Baldwin et al., 1995, p. 16).

The socioeconomic correlates of NALS literacy level suggest that low literacy does, in fact, greatly affect individuals' overall life chances. Table 13.6 shows that the likelihood of White adults working only part-time, not even looking for work, using food stamps, and living in poverty rises steadily as literacy falls. For instance, less than 4% of adults with Level 5 literacy live in poverty, but more than 10 times that proportion with Level 1 literacy do. The odds ratios in the table show that the *relative risks* posed by low literacy are greater for some outcomes (living in poverty as an adult) than others (working part-time), but they are substantial for all (see Gottfredson, in press-b, for further explanation and analysis of these odds ratios). Epidemiologists consider any factor that doubles risk relative to a comparison group to be a "moderately strong" risk factor and one that quadruples risk to be a "strong" risk factor (Gerstman, 1998, pp. 127–128). The comparison group here is Level 3, which probably averages around IQ 105.

TABLE 13.6
Economic Outcomes at Different Levels of NALS Literacy: Whites Aged 16 and Over (% and Odds Ratios)

		Prose Literacy Level				
Outcome		*1* (≤ 225)	*2* (226–275)	*3* (276–325)	*4* (326–375)	*5* (376–500)
Employed only part-time	%	70	57	46	36	28
	OR	2.7	1.6	1.0	0.7	0.5
Out of labor force	%	52	35	25	17	11
	OR	3.2	1.6	1.0	0.6	0.4
Uses food stamps	%	17	13	6	3	1
	OR	3.2	2.3	1.0	0.5	0.2
Lives in poverty	%	43	23	12	8	4
	OR	5.5	2.2	1.0	0.6	0.3
Employed *not* as professional or manager	%	95	88	77	54	30
	OR	5.6	2.2	1.0	0.4	0.1

Source of percentages: Kirsch, Jungleblut, Jenkins, & Kolstad (1993, Figures 2.5, 2.6, 2.7, 2.9, & 2.10).

Everyday Effects of Health Literacy. A recent overview of health literacy in *Patient Care* (Davis, Meldrum, Tippy, Weiss, & Williams, 1996, p. 94) concluded that "[m]edication errors and adverse drug reactions may be due in no small measure to the patient's inability to read and follow written and oral instructions. Poor compliance with medical recommendations, long the bane of well-intentioned physicians, may not be so much a matter of willful disobedience as one of failure to understand the clinician's instructions and expectations." Such risks owing to low *g* are illustrated in Table 13.7, which samples items on the major test of health literacy, the Test of Functional Health Literacy in Adults (TOFHLA; Williams et al., 1995). In this case, the "adequate" literacy group is of roughly average IQ, so it is the comparison group for gauging the relative risks of low

TABLE 13.7
Percentage and Relative Risk (Odds Ratios) of Patients Incorrectly Answering Test Items on the TOFHLA, by Level of Health Literacy

		Literacy Level		
Test Item		*Inadequate*	*Marginal*	*Adequate*
Numeracy items				
How to take medication on an empty stomach	%	65.3	52.1	23.9
	OR	6.0	3.2	1.0
How to take medication four times a day	%	23.6	9.4	4.5
	OR	6.6	2.2	1.0
How many times a prescription can be refilled	%	42.0	24.7	9.6
	OR	6.8	3.1	1.0
How to determine financial eligibility	%	74.3	49.0	31.5
	OR	9.0	3.0	1.0
When next appointment is scheduled	%	39.6	12.7	4.7
	OR	13.5	3.0	1.0
How many pills of a prescription should be taken	%	69.9	33.7	13.0
	OR	15.6	3.4	1.0
Prose Cloze passages				
Instructions for preparing for upper gastro-intestinal tract radiographic procedure	%	57.2	11.9	3.6
	OR	36.2	3.7	1.0
Rights and Responsibilities section of Medicaid application	%	81.1	31.0	7.3
	OR	54.3	5.7	1.0
Standard informed consent document	%	95.1	72.1	21.8
	OR	70.5	9.4	1.0

From Williams, Parker, Baker, Parikh, Pitkin, Coates, & Nurss (1995, Table 3), *Journal of the American Medical Association, 274*, pp. 1677–1682. Copyrighted 1995, American Medical Association.

g. The data are for the walk-in and emergency care center patients of two large urban hospitals. Of the 2,659 patients, 26% did not understand information about when a next appointment is scheduled, 42% did not understand directions for taking medicine on an empty stomach, and 60% could not understand a standard informed consent document (data not shown in table). Table 13.7 shows how patients with "inadequate" health literacy were from 2 to 15 times more likely to fail the items than were individuals with "adequate" literacy. Even the flattest risk gradients in Table 13.7 reveal that low health literacy is more than a "very strong" risk factor (odds ratios of 6.0–6.8) for being unable to follow even the simplest medical instructions.

Table 13.8 provides even more disturbing data on the risks posed by low *g*, because it illustrates that low-ability patients being treated for serious chronic illnesses, such as diabetes and hypertension, often do not understand the most basic facts about their disease and how to manage it. Being in long-term treatment, these patients presumably have been instructed in how to care for themselves and are motivated to do so. That the odds ratios for many of these questions are more favorable than those for the easier TOFHLA items (fall below 6.0 for patients with "inadequate" literacy) probably results from "have done" factors, including training and practice, and possibly higher than normal motivation ("will do" factors). They are nonetheless still high. It is shocking, for instance, that about half of the diabetic patients with inadequate literacy do not know how to recognize the daily symptoms of their disease that require quick management (shakiness, thirst, and the like).

TABLE 13.8
Percentage and Relative Risk (Odds Ratios) of Patients Incorrectly Answering Selected Questions about their Chronic Disease, by Level of Health Literacy

		Literacy Level		
Patient does not know that		*Inadequate*	*Marginal*	*Adequate*
Diabetes				
If you feel thirsty, tired, and weak, it usually	%	40.0	30.8	25.5
means your blood glucose level is high	OR	2.0	1.3	1.0
When you exercise, your blood glucose level	%	60.0	53.8	35.3
goes down	OR	2.7	2.1	1.0
If you suddenly get sweaty, nervous, and	%	62.0	46.1	27.4
shaky, you should eat some form of sugar	OR	4.3	2.3	1.0
Normal blood glucose level is between	%	42.0	23.1	11.8
3.8–7.7 mmol/L (70–140 mg/dL)	OR	5.4	2.2	1.0
If you feel shaky, sweaty, and hungry, it usu-	%	50.0	15.4	5.9
ally means your blood glucose level is low	OR	15.9	2.9	1.0
Hypertension				
Canned vegetables are high in salt	%	36.7	24.0	19.2
	OR	2.4	1.1	1.0
Exercise lowers blood pressure	%	59.7	56.0	32.0
	OR	3.1	2.7	1.0
Blood pressure of 130/80 mm Hg is normal	%	58.2	32.0	28.8
	OR	3.4	1.2	1.0
Losing weight lowers blood pressure	%	33.2	16.0	8.3
	OR	5.5	2.1	1.0
Blood pressure of 160/100 mm Hg is high	%	44.9	30.0	8.3
	OR	9.0	4.7	1.0

From Williams, Baker, Parker, & Nurss (1998, Tables 2 and 3), *Archives of Internal Medicine, 158*, pp. 166–172. Copyrighted 1995, American Medical Association.

Research relating literacy to broad health outcomes suggests that the sorts of risks identified earlier do indeed cumulate in some manner to damage health. For example, low-level readers (reading grade level 0–3) who were enrolled in basic education classes had sickness profiles similar to people with serious chronic illnesses (Weiss, Hart, McGee, & D'Estelle, 1992). In a study of Medicaid patients, the worst readers (Grades 0–2) had annual health care costs of $12,974 compared to the average of $2,969 for the whole sample (Weiss, Blanchard, McGee et al., 1994). A third study that prospectively followed 958 urban hospital patients for 2 years found that patients with *inadequate* TOFHLA literacy were twice as likely (31.5%) to be admitted to the hospital during the next 2 years as were patients with *adequate* literacy (14.9%; Baker et al., 1998). Controlling for all demographic factors yielded a risk ratio for hospitalization of 1.7, which is a moderately strong effect.

Health literacy researchers worry that the disadvantages of low literacy are rising because the explosive growth in new treatments and technologies has created "tremendous learning demands." "For example, a patient who was lucky enough to survive an acute myocardial infarction in the 1960s was typically discharged with only a pat on the back and wishes for good luck. In the 1990s, such a patient is likely to be discharged on a regimen of aspirin, a beta-blocker, an angiotensin-converting enzyme inhibitor, and possibly a low-salt and low-cholesterol diet and medications to control hypertension, diabetes, and hypercholesterolemia. A patient's ability to learn this regimen and follow it correctly will determine a trajectory toward recovery or a downward path to recurrent myocardial infarction, disability, and death" (Baker et al., 1998, p. 791).

CUMULATIVE RISKS OVER THE LIFE COURSE

The literacy research suggests how differences in performance on discrete everyday tasks in self-management can cumulate to produce serious long-term disadvantages for living a long and good life. Just as the typical IQ item is not by itself a good measure of *g* (the typical item correlates only weakly with IQ), no one daily episode of competence may reflect mostly *g* at work. However, as long as items *consistently* tap *g* but not any other one thing, as apparently do the NALS items, their bits of *g* loading pile up while their "specificities" cancel each other out. With sufficient items, the result is a test that measures virtually nothing but *g*.

The issue here, then, is the consistency with which *g* runs through life's many daily activities, whether simple or complex and whether technical or socioemotional, and thereby consistently tilts the odds of success in favor of the more able. As is apparent from the example of IQ test items, *g* doesn't need to tilt the odds very much in any particular instant to have a dramatic impact as those instants accumulate. Even small effect sizes, when their effects add up, can produce big gains over the long run—as do the small odds favoring the house in gambling (Gordon, Lewis, & Quigley, 1988). A big cumulative impact for *g* does not rule out the possibility, of course, that other consistent attributes of people (conscientiousness) and their surrounds (family assistance) also cumulate in the same manner—as would be the case, for example, when both ability and effort influence the many individual grades that cumulate to produce a high school grade point average (GPA). As with job performance, one must always assume that other things matter too.

The following pages look at evidence on how *g* tilts the odds of success in different domains of life as well as how those odds can themselves multiply across domains. I focus on social outcomes that emerge from long, cumulative histories of behavior and specifically on the nexus of good outcomes (higher education, occupation, and income) and the nexus of bad outcomes (unemployment, crime, welfare use, and illegitimacy) that so concern policy researchers. Only the barest summary can be provided because each major outcome represents a huge literature in itself.

The Nexus of Good Outcomes

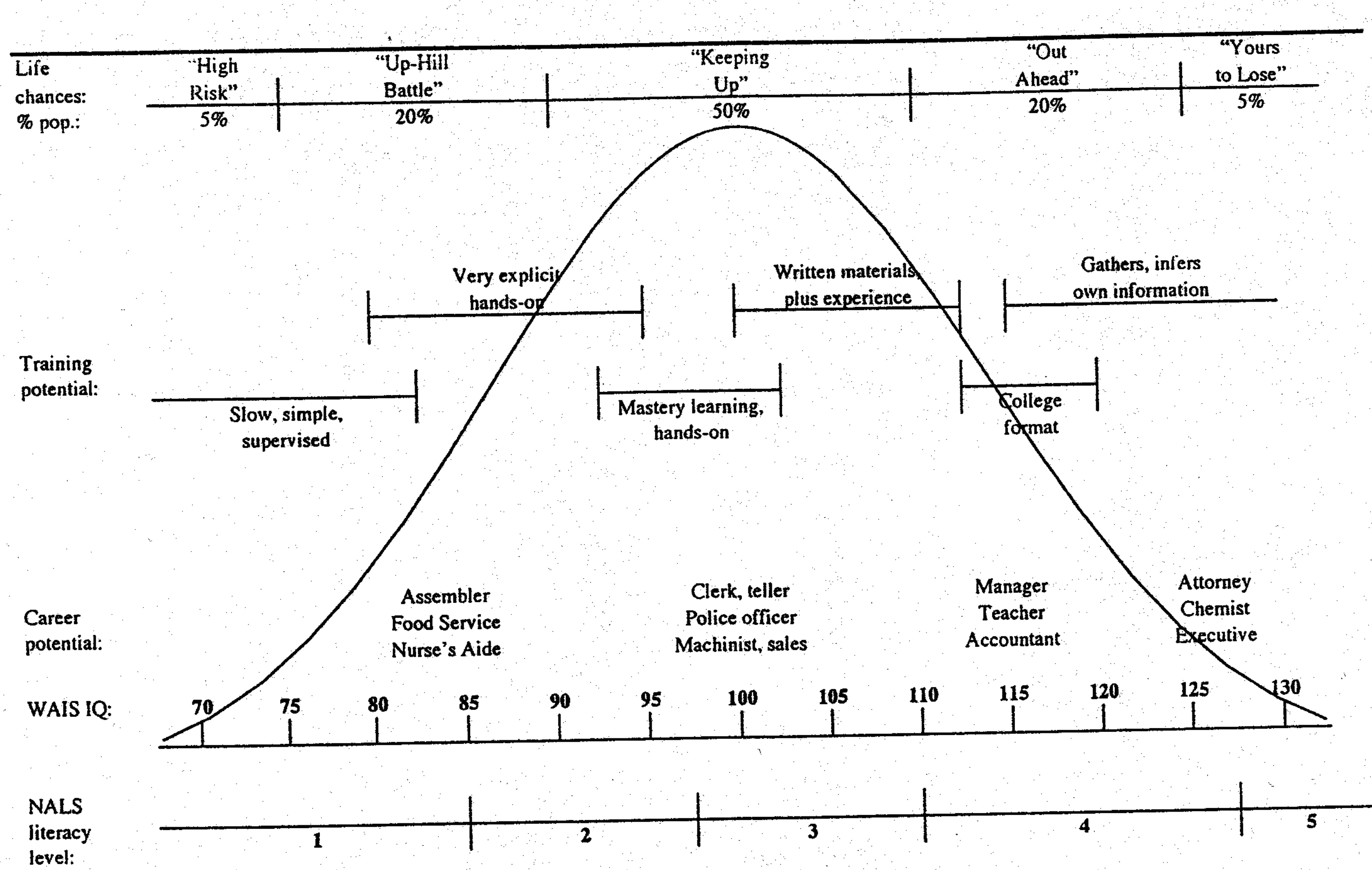
Correlations of IQ with socioeconomic outcomes vary in size depending on the outcome in question, but they are consistent and substantial: years of education (generally .5–.6), occupational status (.4–.5), and earnings, where the correlations *rise* with age (.2–.4; see especially the reanalysis of 10 large samples by Jencks et al., 1979, chap. 4). The predictions are the same whether IQ is measured in Grades 3 to 6, high school, or adulthood (Jencks et al., 1979, pp. 96–99). Moreover, they are underestimates, because they come from single tests of uncertain *g* loading (Jencks et al., 1979, p. 91). Various specific aptitude and achievement tests (both academic and nonacademic) also predict education, occupation, and earnings, but essentially only to the extent that they also measure *g* (Jencks et al., 1979, pp. 87–96; Gottfredson & Crouse, 1986). In short, *g* is what drives a test's predictions of socioeconomic success, and the predictions are substantial when *g* is reasonably well measured, even in childhood.

A large body of work in sociology and economics (e.g., Behrman, Hrubec, Taubman, & Wales, 1980; Jencks et al., 1972; Jencks et al., 1979; Sewell & Hauser, 1975; Taubman, 1977) has painted what might be called an augmented-chain-reaction portrait of *g*'s effects. IQ has a very large effect on years of education obtained; education in turn has a strong effect on occupational level attained; which in turn has a modest effect on income. That is the chain reaction. IQ's effects are carried through at each stage of the process, indirectly. This *indirect* effect of IQ, via education or occupation, diminishes down the line as other non-*g* forces come into play to determine level of occupation and earnings. The other forces include structural ones, such as employers relying on cheap but fallible signals of talent (educational credentials) to hire workers (Gottfredson, 1985) and using set salary schedules to pay them, regardless of their actual job performance.

However, IQ also has *direct* effects at each stage; that is, it helps to predict occupational status *net* of education and earnings *net* of occupation. Via these direct effects, *g* has a "modest influence" through age 25 in boosting young adults up the occupational ladder (Jencks et al., 1979, p. 220) and a "substantively important" one through at least middle age for upping their relative earnings (p. 119). Thus, *g* tilts the odds of success far down the chain of outcomes by tilting the odds of success at each stage along the way, but it also gives an independent boost at each stage of the process. For example, *g*'s independent boost toward higher earnings may be partly via the better job performance that higher *g* produces. *g* is like a persistent tailwind—or headwind, as the case may be.

Criterion-referenced data yield a portrait of life chances that differ dramatically for people of different IQ levels. They are summarized in the upper part of Fig. 13.3 for five segments of the IQ continuum: IQ 75 and below ("high risk" zone), IQ 76 to 90 ("up-hill battle"), IQ 91 to 110 ("keeping up"), IQ 111 to 125 ("out ahead"), and IQ above 125 ("yours to lose"). Occupational opportunities are virtually unlimited, when only *g* is considered, for people with IQs above 125 (95th percentile), but opportunities become quite restricted and unfavorable for persons of IQ 76 to 90 (5th to 25th percentile). No jobs routinely recruit individuals below IQ 80 (10th percentile), the military is forbidden to do so, and it currently accepts no one below about IQ 85.

A major reason for these differences in employment opportunities is that, as shown in the figure, trainability falls quickly at lower levels of *g*. As was reaffirmed by the literacy research, instruction must be drastically simplified for low-*g* people and stripped of anything abstract or "theoretical."



Ever incarcerated (%, white men)	7	7	3	1	0
Chronic welfare recipient (%, white mothers)	31	17	8	2	0
Had illegitimate child (%, white women)	32	17	8	4	2
High school dropout (%, whites)	55	35	6	0.4	0

FIG. 13.3. Life chances in different ranges of the IQ bell curve. *Note.* Adapted from Gottfredson (1997, Figure 3) by permission of Elsevier Publishing.

Whereas the learning of low-*g* individuals must be carefully guided and highly supervised, high-*g* individuals can essentially train themselves when given access to the requisite information.

The Nexus of Bad Outcomes

Just as high *g* is associated with a nexus of good socioeconomic outcomes, so too is low *g* associated with a nexus of social pathologies. Juvenile delinquency, for instance, correlates about –.25 with IQ (Gordon, 1986) and –.3 with poor school deportment (frequency of disciplinary actions; Roberts & Baird, 1972; see also Herrnstein & Murray, 1994, chap. 11; Moffitt, Gabrielli, Mednick, Schulsinger, 1981; Moffitt & Silva, 1988). Although those correlations might seem low, they are equal to or higher than that between social class background and delinquency (Gordon, 1997), which social commentators generally consider important. In any case, small correlations can represent big shifts in odds of success and failure across the IQ continuum.

This is shown in the lower part of Fig. 13.3, which gives the *g*-related risks of incarceration and several other social pathologies that are typically treated as dichotomous (either–or) outcomes. The rows show the percentage of young White adults in each broad IQ range who exhibit each of the bad outcomes. The data, which are from the National Longitudinal Survey of Youth (NLSY; Herrnstein & Murray, 1994), reveal that the prevalence of these serious problems typically doubles at *each* successively lower range of IQ. This serial doubling was also seen for poverty and use of food stamps in the NALS research (Table 13.6), which included a large sample of white adults from a much wider age range (Kirsch et al., 1993).

Table 13.9 provides data for yet other outcomes from the NLSY research, particularly for welfare use and attachment to the labor force. The figure and two tables tell the same story: lower levels of cognitive skill are associated with substantially higher rates of social pathology. Looking down the columns in all three arrays of data suggests how risks *compound* across outcomes. Compared to the average person, low-IQ people have many times the risk of not just one bad outcome, but many, and they are at correspondingly little risk of good outcomes, such as employment in managerial or professional work. Exactly the reverse is true for high-*g* persons. The lower one's *g* level is, the higher one's risk of tripping at least one of the landmines littering the fields of life.

Table 13.9 lists the outcomes in roughly ascending order according to how strongly *g* affects the *relative* risks of these outcomes for people of very low IQ. The odds ratios (ORs) range from less than 2.0 for men being out of the labor force (not looking for work) or unemployed, to around 5.0 to 6.0 for illegitimacy, poverty, and welfare use, to 19.0 for dropping out of

TABLE 13.9
Relative Risk of Bad Outcomes Associated With Lower IQ: Prevalence (%) and Odds Ratios (OR) for Young White Adults

		IQ Level				
Outcome		*≤ 75*	*76–90*	*91–110*	*111–125*	*> 125*
Bell Curve data: General population[a]						
Out of labor force 1+ mo/yr (men)	%	22	19	15	14	10
	OR	1.6	1.3	1.0	0.9	0.6
Unemployed 1+ mo/yr (men)	%	12	10	7	7	2
	OR	1.8	1.5	1.0	1.0	0.3
Ever incarcerated (men)	%	[7][b]	7	3	1	0
	OR	[2.4]	2.4	1.0	0.3	0.1[c]
Chronic welfare recipient (women)	%	31	17	8	2	0
	OR	5.2	2.4	1.0	0.2	0.05[d]
Had illegitimate children (women)	%	32	17	8	4	2
	OR	5.4	2.4	1.0	0.5	0.2
Lives in poverty as an adult	%	30	16	6	3	2
	OR	6.7	3.0	1.0	0.5	0.3
Went on welfare after 1st child (women)	%	55	21	12	4	1
	OR	9.0	2.0	1.0	0.3	0.1
High school dropout	%	55	35	6	0.4	0
	OR	19.0	8.4	1.0	0.1	0
Bell Curve data: Sibling pairs[e]						
Not working in professional job	%	100	99	98	92	77
	OR	hi[f]	2.0	1.0	0.2	0.1
Not a college graduate	%	100	97	81	50	18
	OR	hi[f]	7.6	1.0	0.2	0.1

[a]Source of percentages: Herrnstein & Murray (1994, respectively, pp. 158, 163, 247, 194, 180, 132, 194, & 146).
[b]See text for explanation.
[c]Assuming that % rounded to zero from 0.4, which yields odds of .004 and an odds ratio of .13.
[d]Assuming that % rounded to zero from 0.4, which yields odds of .004 and an odds ratio of .046.
[e]Source of percentages: Murray (1997b).
[f]OR can not be calculated because the odds of 100:0 (its numerator) cannot be calculated.

high school. ORs for people of *above* average IQ are approximately the mirror image of those for people of below-average IQ. Just as a below-average IQ creates a disadvantage relative to the average person, an above-average IQ confers a relative advantage.

g is clearly at the center of the nexus of both good and bad outcomes, but social scientists generally resist attributing it causal power. Sibling studies provide strong evidence, however, that *g* has a big causal influence and social class a comparatively weak one in determining socioeconomic outcomes. Biological siblings differ two thirds as much in IQ, on the average, as do random strangers (12 vs. 17 IQ points). Despite siblings growing up in the very same households, their differences in IQ portend differences in life outcomes that are almost as large as those observed in the general population (Jencks et al., 1979, chap. 4; Murray, 1997a, 1997b; Olneck, 1977, pp. 137–138). Even in intact, nonpoor families, siblings of above average intelligence are much more likely to have a college degree, work in a professional job, and have high earnings than are their average-IQ siblings, who in turn do much better than their low-IQ siblings (Murray, 1997b). Table 13.9 provides the sibling data for holding a college degree and a high-level job. The story is the same for bad outcomes: lower-IQ siblings have much higher rates of bearing illegitimate children, living in poverty, and the like (Murray, 1997a).

TASK, PERSON, AND STRUCTURAL FACTORS THAT TILT *g*'S GRADIENTS OF RELATIVE RISK

The *g* factor relates, often strongly, to a wide range of important outcomes in life. However, *g*'s effect sizes vary across different dimensions of personal life just as they do across tasks in the workplace. The correlational data suggest, for instance, that some good outcomes, such as level of education and occupation, depend more on *g* level than others, such as income. The odds ratios likewise indicate that some bad outcomes, such as chronic welfare dependence, are more sensitive to differences in *g* than are others, such as unemployment. The relative risk of misunderstanding standard hospital documents is greater than the relative risk of misunderstanding medicine labels and the daily signals of whether one is maintaining good control of a chronic illness.

Understanding the practical impact of *g* requires understanding the pattern of these different gradients of risk across the landscape of people's lives. The job performance model discussed earlier (Fig. 13.1) provides some guides. Relative risks can be pushed up or down by both the nature of the *tasks* and the range of variation in the *people* studied. Among task attributes, complexity may be the most crucial. More cognitively demanding tasks (consent forms vs. prescription labels) always tilt the odds of competent behavior in favor of bright people, all else equal. A task's complexity can be pushed downward over time for the individuals performing it, however, when people get much experience with (that is, learn) the task at hand, whether it be a job or a chronic illness. Complexity can also be pushed upward by, among other things, diverting people's cognitive resources away from the task (multitasking) or by granting them more freedom—freedom to use their "judgment"—in performing the task as they wish. As shown earlier, the validity of *g* for predicting job performance was higher in civilian jobs than in comparably complex military jobs, where people have to "work by the book." Social norms and mores can also be seen as pressure to work by the book. They vary over time and place (the acceptability of bearing illegitimate children), so we might expect to see corresponding shifts in relative risk for low *g* people when supervision and social pressure wax and wane.

Gradients of relative risk can also steepen or flatten for reasons having nothing to do with task complexity, as the job performance model indicates. Risk gradients can tilt when the people being studied differ in other traits that affect performance on a particular task: special abilities (other "can do" factors), personality and interests ("will do" factors), and task-specific experience ("have done" factors). When they are task-relevant, these other person factors will cause *g*-related shifts in relative risks, however, only when they are correlated with *g* itself. Relative odds will become steeper when those other factors are *positively* correlated with *g* (bright people are more motivated or experienced at a task than are dull people). Conversely, relative risks will be leveled somewhat if these factors are *negatively* correlated with *g* (dull people are more motivated or have more experience at the task). In certain extreme conditions, such as when bright people have no experience but dull people a lot, the relative risks can be reversed.

We have to go beyond the job performance model for the institutional and contextual factors that moderate *g*'s influence on outcomes such as income or welfare use. Personnel researchers try to eliminate such factors in their studies in order to better understand task performance itself, but such contextual factors can be expected to moderate the impact of *g* in many arenas of life. Perhaps most important among them is the degree to which different life outcomes are even sensitive to a person's *performance* rather than to institutional dictates. To take an example, a salesman's income may be quite sensitive to his sales performance, but a teacher's may depend exclusively on years of tenure and education, regardless of performance in the classroom. The more responsive institutions are to differences in performance, whether that be raising salary for a job well done or increasing welfare for more children borne, the more likely it is that differences in *g* will steepen the gradients of relative risk (for income, welfare dependence). As discussed earlier, the more freedom society gives individuals to make unconstrained choices (to get more education, to bear children out of wedlock), the bigger the impact any differences in *g* will have. Personal freedom increases *g*'s already vast generality.

REFERENCES

Arvey, R. D. (1986). General ability in employment: A discussion. *Journal of Vocational Behavior, 29*(3), 415–420.

Baker, D. W., Parker, R. M., Williams, M. V., & Clark, W. S. (1998). Health literacy and the risk of hospital admission. *Journal of General Internal Medicine, 13*, 791–798.

Baldwin, J., Kirsch, I. S., Rock, D., & Yamamoto, K. (1995). *The literacy proficiencies of GED examinees: Results from the GED-NALS comparison study*. Washington, DC: American Council on Education and Educational Testing.

Behrman, J. R., Hrubec, Z., Taubman, P., & Wales, T. (1980). *Socioeconomic success: A study of the effects of genetic endowments, family environment, and schooling*. New York: North-Holland.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco: Jossey-Bass.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76*(6), 863–872.

Brand, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 251–265). New York: Falmer Press.

Brown, H., Prisuta, R., Jacobs, B., & Campbell, A. (1996). *Literacy of older adults in America: Results from the National Adult Literacy Survey*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Campbell, J. P. (Ed.). (1990). Project A: The U. S. Army Selection and Classification Project. *Personnel Psychology, 43*(2) [special issue].

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*(2), 313–333.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24*(1), 25–52.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cornelius, S. W., & Caspi, A. (1987). Everyday problem solving in adulthood and old age. *Psychology and Aging, 2*(2), 144–153.

Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75*, 295–300.

Davis, T. C., Meldrum, H., Tippy, P. K. P, Weiss, B. D., & Williams, M. V. (1996). How poor literacy leads to poor health care. *Patient Care, 30*(16), 94–127.

Davis, T. C., Michielutte, R., Askov, E. N., Williams, M. V., & Weiss, B. D. (1998). Practical assessment of adult literacy in health care. *Health Education & Behavior, 25*(5), 613–624.

Diehl, M. (1998). Everyday competence in later life: Current status and future directions. *The Gerontologist, 38*(4), 422–433.

Doak, C. C., Doak, L. G., & Root, J. H. (1996). *Teaching patients with low literacy skills* (2nd ed.). Philadelphia: J. B. Lippincott.

Edgerton, R. B. (1981). Another look at culture and mental retardation. In M. J. Begab, H. D. Haywood, & H. L. Garber (Eds.), *Psychosocial influences in retarded performance. Vol. I: Issues and theories in development* (pp. 309–323). Baltimore: University Park Press.

Gardner, H. (1997). A multiplicity of intelligences. *Scientific American Presents, 9*(4), 18–23.

Gerstman, B. B. (1998). *Epidemiology kept simple: An introduction to classic and modern epidemiology*. New York: Wiley.

Gordon, R. A. (1986). Scientific justification and the race-IQ-delinquency model. In T. F. Hartnagel & R. A. Silverman (Eds.), *Critique and explanation: Essays in honor of Gwynne Nettler* (pp. 91–131). New Brunswick, NJ: Transaction Books.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence, 24*(1), 203–320.

Gordon, R. A., Lewis, M. A., & Quigley, A. M. (1988). Can we count on muddling through the *g* crisis in employment? *Journal of Vocational Behavior, 33*, 424–451.

Gottfredson, L. S. (1985). Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In A. C. Kerckhoff (Ed.), *Research in sociology of education and socialization: Vol. 5* (pp. 119–165). Greenwich, CT: JAI Press.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132.

Gottfredson, L. S. (in press-a). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*.

Gottfredson, L. S. (in press-b). *g*, jobs, and life. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. Elmsford, NY: Pergamon Press.

Gottfredson, L. S., & Crouse, J. (1986). The validity versus utility of mental tests: Example of the SAT. *Journal of Vocational Behavior, 29*(3), 363–378.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology: Volume 2* (2nd ed., pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.

Humphreys, L. G. (1986). Commentary. *Journal of Vocational Behavior, 29*(3), 421–437.

Hunter, J. E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hunter, J. E. (1983b). *Overview of validity generalization for the U.S. Employment Service*. (USES Test Research Report No. 43). Washington, DC: U. S. Department of Labor, Employment and Training Administration, Division of Counseling and Test Development.

Hunter, J. E. (1983c). *The dimensionality of the General Aptitude Test Battery (GATB) and the dominance of the general factors over specific factors in the prediction of job performance for USES* (Test Research Report No. 44). Washington, DC: U.S. Department of Labor, U.S. Employment Service.

Hunter, J. E. (1983d). *The prediction of job performance in the military using ability composites: The dominance of general cognitive ability over specific aptitudes* (DOD Contract No. F41689-83-C-0025). Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1985). *Differential validity across jobs in the military*. Unpublished report, Department of Psychology, Michigan State University.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340–362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72–98.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2*(3/4), 447–472.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*, 28–42.

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G., McClelland, K., Mueser, P., Olneck, M., Schwartz, J., Ward, S., & Williams, J. (1979). *Who gets ahead? The determinants of economic success in America*. New York: Basic Books.

Sternberg, R. J., & Wagner, R. K. (1993). The *g*-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2*(1), 1–5.
Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50*, 912–926.
Sticht, T. (Ed.). (1975). *Reading for working: A functional literacy anthology*. Alexandria, VA: Human Resources Research Organization.
Taubman, P. (Ed.). (1977). *Kinometrics: Determinants of socioeconomic success within and between families*. New York: North-Holland.
Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29*(3), 332–339.
Venezky, R. L., Kaestle, C. F., & Sum, A. M. (1987). *The subtle danger: Reflections on the literacy abilities of America's young adults*. Princeton, NJ: Educational Testing Service.
Viswesvaran, C., & Ones, D. S. (Eds.). (in press). Role of general mental ability in industrial, work, and organizational (IWO) psychology [Special issue]. *Human Performance*.
Weiss, B. D., Blanchard, J. S., McGee, D. L., Hart, G., Warren, B., Burgoon, M., & Smith, K. J. (1994). Illiteracy among Medicaid recipients and its relationship to health care costs. *Journal of Health Care for the Poor and Underserved, 5*, 99–111.
Weiss, B. D., Hart, G., & McGee, D., & D'Estelle, S. (1992). Health status of illiterate adults: Relation between literacy and health status among persons with low literacy skills. *Journal of the Board of Family Practice, 5*, 257–264.
Weiss, B. D., Reed, R. L., & Kligman, E. W. (1995). Literacy skills and communication methods of low-income older persons. *Patient Education and Counseling, 25*, 109–119.
Wigdor, A. K., & Green, B. F., Jr. (Eds.). (1991). *Performance for the workplace. Volume 1*. Washington, DC: National Academy Press.
Williams, M. V., Baker, D. W., Parker, R. M., & Nurss, J. R. (1998). Relationship of functional health literacy to patients' knowledge of their chronic disease. *Archives of Internal Medicine, 158*, 166–172.
Williams, M. V., Parker, R. M., Baker, D. W., Parikh, N. S., Pitkin, K., Coates, W. C., & Nurss, J. R. (1995). Inadequate functional health literacy among patients at two public hospitals. *Journal of the American Medical Association, 274*(21), 1677–1682.
Willis, S. L., & Schaie, K. W. (1986). Practical intelligence in later adulthood. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 236–268). New York: Cambridge University Press.