

## Where and Why $g$ Matters: Not a Mystery

Linda S. Gottfredson  
*School of Education*  
*University of Delaware*

$g$  is a highly general capability for processing complex information of any type. This explains its great value in predicting job performance. Complexity is the major distinction among jobs, which explains why  $g$  is more important further up the occupational hierarchy. The predictive validities of  $g$  are moderated by the criteria and other predictors considered in selection research, but the resulting gradients of  $g$ 's effects are systematic. The pattern provides personnel psychologists a road map for how to design better selection batteries. Despite much literature on the meaning and impact of  $g$ , there nonetheless remains an aura of mystery about where and why  $g$  cognitive tests might be useful in selection. The aura of mystery encourages false beliefs and false hopes about how we might reduce disparate impact in employee selection. It is also used to justify new testing techniques whose major effect, witting or not, is to reduce the validity of selection in the service of racial goals.

The general mental ability factor— $g$ —is the best single predictor of job performance. It is probably the best measured and most studied human trait in all of psychology. Much is known about its meaning, distribution, and origins thanks to research across a wide variety of disciplines (Jensen, 1998). Many questions about  $g$  remain unanswered, including its exact nature, but  $g$  is hardly the mystery that some people suggest. The totality—the pattern—of evidence on  $g$  tells us a lot about where and why it is important in the real world. Theoretical obtuseness about  $g$  is too often used to justify so-called technical advances in personnel selection that minimize, for sociopolitical purposes, the use of  $g$  in hiring.

## THE *g* FACTOR AMONG PEOPLE

Our knowledge of the mental skills that are prototypical of *g*, of the aspects of tasks that call forth *g*, and of the factors that increase or decrease its impact on performance together sketch a picture of where and why *g* is useful in daily affairs, including paid work. They show *g*'s predictable gradients of effect. I begin here with the common thread—the *g* factor—that runs through the panoply of people's mental abilities.

### Generality and Stability of the *g* Factor

One of the simplest facts about mental abilities provides one of the most important clues to the nature of *g*. People who do well on one kind of mental test tend to do well on all others. When the scores on a large, diverse battery of mental ability tests are factor analyzed, they yield a large common factor, labeled *g*. Pick any test of mental aptitude or achievement—say, verbal aptitude, spatial visualization, the SAT, a standardized test of academic achievement in 8th grade, or the Block Design or Memory for Sentences subtests of the Stanford–Binet intelligence test—and you will find that it measures mostly *g*. All efforts to build meaningful mental tests that do not measure *g* have failed.

Thus, try as we might to design them otherwise, all our mental tests measure mostly the same thing, no matter how different their manifest content is. This means that *g* must be a highly general ability or property of the mind. It is not bound to any particular kind of task content, such as words, numbers, or shapes. Very different kinds of test content can be used to measure *g* well—or badly.

This dimension of human difference in intellect—the *g* factor—does not seem bound to particular cultures, either, because virtually identical *g* factors have been extracted from test batteries administered to people of different ages, sexes, races, and national groups. In contrast, no general factor emerges from personality inventories, which shows that general factors are not a necessary outcome of factor analysis. (See Jensen, 1998, and Gottfredson, 1997, 2000a, 2002, for fuller discussion and documentation of these and following points on *g*.)

*g*'s high generality is also demonstrated by the predictive validities of mental tests. It is the *g* component of mental tests that accounts almost totally for their predictive validity. Indeed, whole batteries of tests do little better than *g* alone in predicting school and job performance. The more *g*-loaded a test is (the better it correlates with *g*), the better it predicts performance, including school performance, job performance, and income. There are many different abilities, of course, as is confirmed by the same factor analyses that confirm the dominance of the general factor among them. Because *g* is more general in nature than the narrower group factors (such as verbal aptitude, spatial visualization, and memory), it is, not surprisingly, also broader in applicability. The clerical (i.e., non-*g*) component of cler-

ical tests, for instance, enhances performance somewhat in clerical jobs (beyond that afforded by higher *g*), but *g* enhances performance in all domains of work.

The *g* factor shows up in nonpsychometric tests as well, providing more evidence for both its reality and generality. Exceedingly simple reaction time and inspection time tasks, which measure speed of reaction in milliseconds, also yield a strong information processing factor that coincides with psychometric *g*.

In short, the *g* continuum is a reliable, stable phenomenon in human populations. Individual differences along that continuum are also a reliable, stable phenomenon. IQ tests are good measures of individual variation in *g*, and people's IQ scores become quite stable by adolescence. Large changes in IQ from year to year are rare even in childhood, and efforts to link them to particular causes have failed. Indeed, mental tests would not have the pervasive and high predictive validities that they do, and often over long stretches of the life span, if people's rankings in IQ level were unstable.

Theorists have long debated the definition of "intelligence," but that verbal exercise is now moot. *g* has become the working definition of intelligence for most researchers, because it is a stable, replicable phenomenon that—unlike the IQ score—is independent of the "vehicles" (tests) for measuring it. Researchers are far from fully understanding the physiology and genetics of intelligence, but they can be confident that, whatever its nature, they are studying the same phenomenon when they study *g*. That was never the case with IQ scores, which fed the unproductive wrangling to "define intelligence." The task is no longer to define intelligence, but to understand *g*.

### Meaning of *g* as a Construct

Understanding *g* as a construct—its substantive meaning as an ability—is essential for understanding why and where *g* enhances performance of everyday tasks. Some sense of its practical meaning can be gleaned from the overt behaviors and mental skills that are prototypical of *g*—that is, those that best distinguish people with high *g* levels from those with low *g*. Intelligence tests are intended to measure a variety of higher order thinking skills, such as reasoning, abstract thinking, and problem solving, which experts and laypeople alike consider crucial aspects of intelligence. *g* does indeed correlate highly with specific tests of such aptitudes. These higher order skills are context- and content-independent mental skills of high general applicability. The need to reason, learn, and solve problems is ubiquitous and lifelong, so we begin to get an intuitive grasp of why *g* has such pervasive value and is more than mere "book smarts."

We can get closer to the meaning of *g*, however, by looking beyond the close correlates of *g* in the domain of human abilities and instead inspect the nature of the tasks that call it forth. For this, we must analyze data on tasks, not people. Recall that the very definition of an ability is rooted in the tasks that people can per-

form. 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. Superficial inspection of *g*-loaded tests and tasks shows immediately what they are not, but are often mistakenly assumed to be—curriculum or domain dependent. Thus, the distinguishing attributes of *g*-loaded tasks must cut across all content domains.

Comparisons of mental tests and items reveal that the more *g*-loaded ones are more complex, whatever their manifest content. They require more complex processing of information. The hypothetical IQ test items in Figure 1 illustrate the point. Items in the second column are considerably more complex than those in the first column, regardless of item type and regardless of whether they might seem “academic.” To illustrate, the first item in the first row requires only simple computation. In contrast, the second item in that row requires exactly the same computation, but the person must figure out which computation to make. The similarities items in the third row differ in abstractness in the similarities involved. The more difficult block design item uses more blocks and a less regular pattern, and so on.

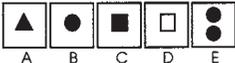
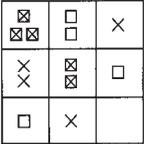
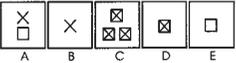
Directions	Simple Item	Complex Item
1. Compute	.50 x 6 = 3.00	Apples cost 50 cents apiece and Susan bought 6. How much did she spend?
2. Define	Table	Ponder
3. State one similarity	Pear — Apple	Seed — Egg
4. Give the next 2 numbers	2, 4, 6, 8, __ __	11, 10, 9, 10, 9, 8, __ __
5. Reproduce pattern with blocks whose sides are 	Use 4 blocks: 	Use 9 blocks: 
6. Complete the pattern	 	 

FIGURE 1 Hypothetical examples of simple versus more complex IQ test items.

Task complexity has been studied systematically in various contexts, some psychometric and some not. Researchers in the fields of information processing, decision making, and goal setting stress the importance of the number, variety, variability, ambiguity, and interrelatedness of information that must be processed to evaluate alternatives and make a decision. Wood (1986), for example, discussed three dimensions of task complexity: component complexity (e.g., number of cues to attend to and integrate, redundancy of demands), coordinative complexity (e.g., timing or sequencing of tasks, length of sequences), and changes in cause–effect chains or means–ends relations. More complex items require more mental manipulation for people to learn something or solve a problem—seeing connections, drawing distinctions, filling in gaps, recalling and applying relevant information, discerning cause and effect relations, interpreting more bits of information, and so forth.

In a detailed analysis of items on the U.S. Department of Education’s National Adult Literacy Survey (NALS), Kirsch and Mosenthal (1990) discovered that the relative difficulty of the items in all three NALS scales (prose, document, quantitative) originated entirely in the same “process complexity”: type of match (literalness), plausibility of distractors (relevance), and type of information (abstractness). The active ingredient in the test items was the complexity, not content, of the information processing they required. Later research (Reder, 1998) showed, not surprisingly, that the three scales represent one general factor and virtually nothing else.

One useful working definition of *g* for understanding everyday competence is therefore the ability to deal with complexity. This definition can be translated into two others that have also been offered to clarify *g*’s real-world applications—the ability to learn moderately complex material quickly and efficiently and the ability to avoid cognitive errors (see the discussion in Gottfredson, 1997). Most globally, then, *g* is the ability to process information. It is not the amount of knowledge per se that people have accumulated. High *g* people tend to possess a lot of knowledge, but its accumulation is a by-product of their ability to understand better and learn faster.

They fare better with many daily tasks for the same reason. Although literacy researchers eschew the concept of intelligence, they have nonetheless confirmed *g*’s importance in highly practical daily affairs. They have concluded, with some surprise, that differences in functional literacy (using maps, menus, order forms, and bank deposit slips; understanding news articles and insurance options; and the like) and health literacy (understanding doctors’ instructions and medicine labels, taking medication correctly, and so on) reflect, at heart, differences in a general ability to process information (Gottfredson, 1997, 2002).

Clearly, there is much yet to be learned about the nature of *g*, especially as a biological construct. We know enough about its manifest nature already, however, to dispel the fog of mystery about why it might be so useful. It is a generic, infinitely

adaptable tool for processing any sort of information, whether on the job or off, in training or after.

## THE COMPLEXITY FACTOR AMONG JOBS

We also know a lot about where high *g* confers its greatest advantages. Its impact is lawful, not ephemeral or unpredictable.

### Analyses of the Skills That Jobs Demand

Just as the skills that people possess have been factor analyzed, so too have the demands that jobs make. Both analyses yield analogous results, hardly a statistically necessary result. Just as there is a general ability factor among individuals, there is a general complexity factor among jobs. (See Gottfredson, 1985, on how the former may cause the latter.) The largest, most consistent distinction among jobs is the complexity of their information processing demands. In some studies, this jobs factor has been labeled “judgment and reasoning” (Arvey, 1986). In sociological research, it is usually labeled “complexity.”

Table 1 reveals the meaning of the job complexity factor by listing its strongest correlates. The results in Table 1 are from a principal components analysis of 64% of the broad occupational categories (and 86% of jobs) in the 1970 census. That analysis used all job analysis data then available that could be linked to the census titles. All those job analysis attributes are listed in Table 1 so that it is clear which ones do and do not correlate with job complexity. Table 1 lists them according to whether they correlate most highly with the complexity factor rather than some other factor. (None of these items was used in actually deriving the factors. See Gottfredson, 1997, for the items used in the principal components analysis.) The data come primarily from the Position Analysis Questionnaire (PAQ), but also from the 1970 U.S. Census, ratings in *Dictionary of Occupational Titles*, and several smaller bodies of occupational data (labeled here as the Temme and Holland data). All the attributes listed in Table 1 are from the PAQ, unless otherwise noted.

Almost all of the many items pertaining to information processing correlate most highly with the complexity factor. These items represent requirements for perceiving, retrieving, manipulating, and transmitting information. Those that are generally viewed as higher level processing skills, such as compiling and combining information (.90, .88), reasoning (.86), and analyzing (.83), have the highest correlations with the complexity factor. Somewhat lower level processes, such as memory (.40) and transcribing (.51), have lower but still substantial correlations. Only the highly visual information processing activities (e.g., seeing, vigilance with machines) fail to correlate most with the complexity factor. They correlate, instead, with factors reflecting use of objects (“things”) and machines, independ-

TABLE 1  
Job Attributes That Correlate Most With the Job Complexity Factor

<i>Correlate Most With "Complexity" Factor</i>	<i>r</i>	<i>Correlate Most With Another Factor</i>	<i>r</i>	<i>The Other Factor</i>
Processing information (perceiving, retrieving, manipulating, transmitting)				
Compiling information, importance of	0.90	Seeing (DOT)	0.66	Work with complex things
Combining information, importance of	0.88	Information from events, extent of use	0.58	Vigilance with machines
Language, level of (DOT)	0.88	Vigilance—changing events, importance of	0.57	Vigilance with machines
Reasoning, level of (DOT)	0.86	Pictorial materials, extent of use	0.44	Work with complex things
Writing, importance of	0.86	Apply measurable, verifiable criteria (DOT)	0.43	Work with complex things
Intelligence (DOT)	0.84	Vigilance—infrequent events, importance of	0.41	Vigilance with machines
Written information, extent of use	0.84	Patterns, extent of use	0.41	Work with complex things
Analyzing information, importance of	0.83	Interpret others' feelings, ideas, facts (DOT)	0.22	Catering to people
Math, level of (DOT)	0.79			
Math, level of	0.70			
Quantitative information, extent of use	0.68			
Coding and decoding, importance of	0.68			
Oral information, extent of use	0.68			
Talking (DOT)	0.68			
Behavioral information, extent of use	0.59			
Apply sensory and judgmental criteria (DOT)	0.55			
Attention to detail, importance of	0.54			
Transcribing, importance of	0.51			
Short-term memory, importance of	0.40			
Recognize and identify, importance of	0.36			
Practical problem solving				
Advising, importance of	0.86	Supervising nonemployees, importance of	0.64	Catering to people
Planning and scheduling, amount of	0.83	Catering and serving, importance of	0.61	Catering to people
Decision making, level of	0.82	Entertaining, importance of	0.59	Catering to people
Negotiating, importance of	0.79	Non-job-required social contact, opportunity	0.25	Catering to people
Persuading, importance of	0.79			

*(continued)*

TABLE 1 (Continued)

<i>Correlate Most With "Complexity" Factor</i>	<i>r</i>	<i>Correlate Most With Another Factor</i>	<i>r</i>	<i>The Other Factor</i>
Staff functions, importance of	0.79			
Coordinate without line authority, importance of	0.74			
Public speaking, importance of	0.68			
Instructing, importance of	0.67			
Direction, control, and planning (DOT)	0.59			
Dealing with people (DOT)	0.59			
Influencing (DOT)	0.42			
Level of responsibility and respect				
Prestige (Temme)	0.82	Responsibility for materials, degree of	0.48	Vigilance with machines
General responsibility, degree of	0.76	Responsibility for safety, degree of	0.47	Vigilance with machines
Criticality of position, degree of	0.71			
Job structure				
Self-direction (Temme)	0.88	Complexity of dealings with things (DOT)	0.77	Work with complex things
Complexity of dealings with data (DOT)	0.83	Follow set procedures, importance of	0.54	Operating machines
Work under distractions, importance of	0.78	Meet set limits, tolerances, standards (DOT)	0.53	Work with complex things
Frustrating situations, importance of	0.77	Specified work pace, importance of	0.44	Operating machines
Interpersonal conflict, importance of	0.76	Cycled activities, importance of	0.42	Operating machines
Strained contacts, importance of	0.69	Perform under stress and risk (DOT)	0.27	Vigilance with machines
Complexity of dealings with people (DOT)	0.68			
Personal contact required, extent of	0.66			
Personal sacrifice, importance of	0.65			
Civic obligations, importance of	0.64			
Time pressure, importance of	0.55			
Precision, importance of	0.53			
Variety and change (DOT)	0.41			
Repetitive activities, importance of	-0.49			
Supervision, level of	-0.73			
Repetitive or continuous (DOT)	-0.74			
Structure, amount of	-0.79			

Focus of work and interests required				
Interest in data versus things (DOT)	0.73	“Conventional” field of work (Holland)	0.51	Coordination without sight
Interest in creative versus routine work (DOT)	0.63	“Social” field of work (Holland)	0.45	Catering to people
Interest in social welfare versus machines (DOT)	0.55	Interest in science vs. business (DOT)	0.42	Work with complex things
Interest in producing versus esteem (DOT)	-0.48	“Investigative” field of work (Holland)	0.37	Work with complex things
“Realistic” field of work (Holland)	-0.74	“Enterprising” field of work (Holland)	0.33	Selling
		“Artistic” field of work (Holland)	0.20	Work with complex things
Education and experience required				
Education, level of curriculum	0.88			
General education development level (DOT)	0.86			
Update job knowledge, importance of	0.85			
Specific vocational preparation (DOT)	0.76			
Experience, length of	0.62			
Training, length of	0.51			
Physical requirements				
Wet, humid (DOT)	-0.37	Outside versus inside location (DOT)	0.48	Vigilance with machines
Hazardous conditions (DOT)	-0.39	Climbing (DOT)	0.42	Controlled manual work
Fumes, odors, dust, gases (DOT)	-0.45			
Stooping (DOT)	-0.48			
Noise, vibration (DOT)	-0.53			
Physical exertion, level of	-0.56			
Reaching (DOT)	-0.66			
Other correlates				
Salary, yes or no	0.70	Commission, yes or no	0.53	Selling
Percentage government workers, men (census)	0.45	Tips, yes or no	0.50	Selling
Percentage government workers, women (census)	0.45	Licensing and certification	0.42	Catering to people
Percentage black, women (census)	-0.48	Median age, men (census)	0.31	Vigilance with machines
Percentage black, men (census)	-0.53	Mean hours, men (census)	0.31	Controlled manual
Wage, yes or no	-0.66	Median age, women (census)	-0.28	Coordination without sight
		Mean hours, women (census)	-0.34	Catering to people
		Percentage women (census)	-0.37	Controlled manual

*Note.* Source of data: Gottfredson (1997). DOT = Dictionary of Occupational Titles; Temme = Temme’s ratings of occupational prestige and self-direction; Holland = Holland’s vocational personality type codes for occupations (see Gottfredson, 1994, for description and use of these scales).

ent of the job's overall complexity. The extent of use of most forms of information (behavioral, oral, written, quantitative) is also strongly correlated with overall job complexity (.59–.84) but no other factor. The primary exception, once again, is visual (use of patterns and pictorial materials).

Many job duties can be described as general kinds of problem solving—for instance, advising, planning, negotiating, instructing, and coordinating employees without line authority. As Table 1 shows, they are also consistently and substantially correlated with job complexity (.74–.86). In contrast, the requirements for amusing, entertaining, and pleasing people mostly distinguish among jobs at the same complexity level, for they help to define the independent factor of “catering to people.”

Complex dealings with data (.83) and people (.68) are more typical of highly complex than simple jobs, as might be expected. Complex dealings with things (material objects) help to define a separate and independent factor: “work with complex things” (which distinguishes the work of engineers and physicians, e.g., from that of lawyers and professors). Constant change in duties or the data to be processed (“variety and change,” .41) also increase a job's complexity. As the data show, the more repetitive (–.49, –.74), tightly structured (–.79), and highly supervised (–.73) a job is, the less complex it is. Complexity does not rule out the need for tight adherence to procedure, a set work pace, cycled activities, or other particular forms of structure required in some moderately complex domains of work. As can be seen in Table 1, these attributes typify work that is high on the “operating machines” (and vehicles) factor of work.

That the overall complexity of a job might be enhanced by the greater complexity of its component parts is no surprise. However, Table 1 reveals a less well-appreciated point—namely, that job complexity also depends on the configuration of tasks, not just on the sum of their individual demands. Any configuration of tasks or circumstances that strains one's information processing abilities puts a premium on higher *g*. Consider dual-processing and multitasking, for instance, which tax people's ability to perform tasks simultaneously that they have no trouble doing sequentially. The data in Table 1 suggest that information processing may also be strained by the pressures imposed by deadlines (.55), frustration (.77), and interpersonal conflict (.76), and the need to work in situations where distractions (.78) compete for limited cognitive resources. Certain personality traits would aid performance in these situations, but higher *g* would also allow for more effective handling of these competing stresses.

The importance of performing well tends to rise with job complexity, because both the criticality of the position for the organization (.71) and the general responsibility it entails (.76) correlate strongly with job complexity. Responsibility for materials and safety are more domain specific, however, because they correlate most with the “vigilance with machines” factor.

Education and training are highly *g*-loaded activities, as virtually everyone recognizes. Table 1 shows, however, that more complex jobs tend not only to require

higher levels of education (.88), but also lengthier specific vocational training (.76) and experience (.62). The data on experience are especially important in this context, because experience signals knowledge picked up on the job. It reflects a form of self-instruction, which becomes less effective the lower one's *g* level. Consistent with this interpretation, the importance of "updating job knowledge" correlates very highly (.85) with job complexity.

More complex jobs tend to require more education and pay better, which in turn garners them greater social regard. Hence, the job complexity factor closely tracks the prestige hierarchy among occupations (.82), another dimension of work that sociologists documented decades ago.

The other attributes that correlate most highly with complexity, as well as those that do not, support the conclusion that the job complexity factor rests on distinctions among jobs in their information processing demands, generally without regard to the type of information being processed. Of the six Holland fields of work, only one—Realistic—correlates best (and negatively) with the complexity factor (–.74). Such work, which emphasizes manipulating concrete things rather than people or abstract processes, comprises the vast bulk of low-level jobs in the American economy. The nature of these jobs comports with the data on vocational interests associated with the complexity factor. Complex work is associated with interests in creative rather than routine work (.63), with data (.73), and with social welfare (.55), respectively, rather than things and machines, and with social esteem rather than having tangible products (.48). This characterization of low-level, frequently Realistic work is also consistent with the data on physical requirements: All the physically unpleasant conditions of work (working in wet, hazardous, noisy, or highly polluted conditions) are most characteristic of the simplest, lowest-level jobs (–.37 to –.45). In contrast, the skill and activity demands associated with the other factors of work are consistently specific to particular functional domains (fields) of work—for example, selling with "enterprising" work and coordination without sight (such as typing) with "conventional" (mostly clerical) work. So, too, are various other circumstances of work, such as how workers are paid (salary, wages, tips, commissions), which tend to distinguish jobs that require selling from those that do not, whatever their complexity level.

As we saw, the job analysis items that correlate most highly with overall job complexity use the very language of information processing, such as compiling and combining information. Some of the most highly correlated mental demands, such as reasoning and analyzing, are known as prototypical manifestations of intelligence in action. The other dimensions of difference among jobs rarely involve such language. Instead, they generally relate to the material in different domains of work activity, how (not how much) such activity is remunerated, and the vocational interests they satisfy. They are noncognitive by contrast.

The information processing requirements that distinguish complex jobs from simple ones are therefore essentially the same as the task requirements that distin-

guish highly  $g$ -loaded mental tests, such as IQ tests, from less  $g$ -loaded ones, such as tests of short-term memory. In short, jobs are like (unstandardized) mental tests. They differ systematically in  $g$ -loading, depending on the complexity of their information processing demands. Because we know the relative complexity of different occupations, we can predict where job performance (when well measured) will be most sensitive to differences in workers'  $g$  levels. This allows us to predict major trends in the predictive validity of  $g$  across the full landscape of work in modern life. One prediction, which has already been borne out, is that mental tests predict job performance best in the most complex jobs.

The important point is that the predictive validities of  $g$  behave lawfully. They vary, but they vary systematically and for reasons that are beginning to be well understood. Over 2 decades of meta-analyses have shown that they are not sensitive to small variations in job duties and circumstance, after controlling for sampling error and other statistical artifacts. Complex jobs will always put a premium on higher  $g$ . Their performance will always be notably enhanced by higher  $g$ , all else equal. Higher  $g$  will also enhance performance in simple jobs, but to a much smaller degree.

This lawfulness can, in turn, be used to evaluate the credibility of claims in personnel selection research concerning the importance, or lack thereof, of mental ability in jobs of at least moderate complexity, such as police work. If a mental test fails to predict performance in a job of at least moderate complexity (which includes most jobs), we cannot jump to the conclusion that differences in mental ability are unimportant on that job. Instead, we must suspect either that the test does not measure  $g$  well or that the job performance criterion does not measure the most crucial aspects of job performance. The law-like relation between job complexity and the value of  $g$  demands such doubt. Credulous acceptance of the null result requires ignoring the vast web of well-known evidence on  $g$ , much of it emanating from industrial–organizational (I/O) psychology itself.

### RELATIVE IMPORTANCE OF $g$ FOR JOB PERFORMANCE

The I/O literature has been especially useful in documenting the value of other predictors, such as personality traits and job experience, in forecasting various dimensions of performance. It thus illuminates the ways in which  $g$ 's predictive validities can be moderated by the performance criteria and other predictors considered. These relations, too, are lawful. They must be understood to appreciate where, and to what degree, higher levels of  $g$  actually have functional value on the job. I/O research has shown, for instance, how  $g$ 's absolute and relative levels of predictive validity both vary according to the kind of performance criterion used. A failure to

understand these gradients of effect sustains the mistaken view that *g*'s impact on performance is capricious or highly specific across different settings and samples.

The Appendix outlines the topography of *g*—that is, its gradients of effect relative to other predictors. It summarizes much evidence on the prediction of job performance, which is discussed more fully elsewhere (Gottfredson, 2002). This summary is organized around two distinctions, one among performance criteria and one among predictors, that are absolutely essential for understanding the topography of *g* and other precursors of performance. First, job performance criteria differ in whether they measure mostly the core technical aspects of job performance rather than a job's often discretionary "contextual" (citizenship) aspects. Second, predictors can be classified as "can do" (ability), "will do" (motivation), or "have done" (experience) factors.

The Appendix repeats some of the points already made, specifically that (a) *g* has pervasive value but its value varies by the complexity of the task at hand, and (b) specific mental abilities have little incremental validity net of *g*, and then only in limited domains of activity. The summary points to other important regularities. As shown in the Appendix, personality traits generally have more incremental validity than do specific abilities, because "will do" traits are correlated little or not at all with *g*, the dominant "can do" trait, and thus have greater opportunity to add to prediction. These noncognitive traits do, however, tend to show the same high domain specificity that specific abilities do. The exception is the personality factor representing conscientiousness and integrity, which substantially enhances performance in all kinds of work, although generally not as much as does *g*.

An especially important aspect of *g*'s topography is that the functional value of *g* increases, both in absolute and relative terms, as performance criteria focus more on the core technical aspects of performance rather than on worker citizenship (helping coworkers, representing the profession well, and so on). The reverse is generally true for the noncognitive "will do" predictors, such as temperaments and interests: They predict the noncore elements best. Another important regularity is that, although the predictive validities of *g* rise with job complexity, the opposite is true for two other major predictors of performance—length of experience and psychomotor abilities. The latter's predictive validities are sometimes high, but they tend to be highest in the simplest work.

Another regularity is that "have done" factors sometimes rival *g* in predicting complex performance, but they are highly job specific. Take job experience—long experience as a carpenter does not enhance performance as a bank teller. The same is true of job sample or tacit knowledge tests, which assess workers' developed competence in a particular job: Potential bank tellers cannot be screened with a sample of carpentry work. In any case, these "have done" predictors can be used to select only among experienced applicants. Measures of *g* (or personality) pose no such constraints. *g* is generalizable, but experience is not.

As for  $g$ , there are also consistent gradients of effect for job experience. The value of longer experience relative to one's peers fades with time on the job, but the advantages of higher  $g$  do not. Experience is therefore not a substitute for  $g$ . After controlling for differences in experience,  $g$ 's validities are revealed to be stable and substantial over many years of experience. Large relative differences in experience among workers with low absolute levels of experience can obscure the advantages of higher  $g$ . The reason is that a little experience provides a big advantage when other workers still have little or none. The advantage is only temporary, however. As all workers gain experience, the brighter ones will glean relatively more from their experience and, as research shows, soon surpass the performance of more experienced but less able peers. Research that ignores large relative differences in experience fuels mistaken conceptions about  $g$ . Such research is often cited to support the view that everyday competence depends more on a separate "practical intelligence" than on  $g$ —for example, that we need to posit a practical intelligence to explain why inexperienced college students cannot pack boxes in a factory as efficiently as do experienced workers who have little education (e.g., see Sternberg, Wagner, Williams, & Horvath, 1995).

The foregoing gradients of  $g$ 's impact, when appreciated, can be used to guide personnel selection practice. They confirm that selection batteries should select for more than  $g$ , if the goal is to maximize aggregate performance, but that  $g$  should be a progressively more important part of the mix for increasingly complex jobs (unless applicants have somehow already been winnowed by  $g$ ). Many kinds of mental tests will work well for screening people yet to be trained, *if* the tests are highly  $g$ -loaded. Their validity derives from their ability to assess the operation of critical thinking skills, either on the spot ("fluid"  $g$ ) or in past endeavors ("crystallized"  $g$ ). Their validity does not depend on their manifest content or "fidelity"—that is, whether they "look like" the job. Face validity is useful for gaining acceptance of a test, but it has no relation to the test's ability to measure key cognitive skills. Cognitive tests that look like the job can measure  $g$  well (as do tests of mathematical reasoning) or poorly (as do tests of arithmetic computation).

Tests of noncognitive traits are useful supplements to  $g$ -loaded tests in a selection battery, but they cannot substitute for tests of  $g$ . The reason is that noncognitive traits cannot substitute for the information-processing skills that  $g$  provides. Noncognitive traits also cannot be considered as useful as  $g$  even when they have the same predictive validity (say, .3) against a multidimensional criterion (say, supervisor ratings), because they predict different aspects of job performance. The former predict primarily citizenship and the latter primarily core performance. You get what you select for, and the wise organization will never forego selection for core performance.

There are circumstances where one might want to trade away some  $g$  to gain higher levels of experience. The magnitude of the appropriate trade-off, if any, would depend on the sensitivity of job performance to higher levels of  $g$  (the com-

plexity of the work), the importance of short-term performance relative to long-term performance (probable tenure), and the feasibility and cost of training brighter recruits rather than hiring more experienced ones (more complex jobs require longer, more complex training). In short, understanding the gradients of effect outlined in the Appendix can help practitioners systematically improve—or knowingly degrade—their selection procedures.

## THE FLIGHT FROM *g*

Sociopolitical goals for racial parity in hiring and the strong legal pressure to attain it, regardless of large racial disparities in *g*, invite a facade of mystery and doubt about *g*'s functional impact on performance, because the facade releases practitioners from the constraints of evidence in defending untenable selection practices. The facade promotes the false belief that the impact of *g* is small, unpredictable, or ill-understood. It thereby encourages the false hope that cognitive tests, if properly formed and used, need not routinely have much, if any, disparate impact—or even that they could be eliminated altogether. Practitioners can reduce disparate impact in ways that flout the evidence on *g*, but they, and their clients, cannot escape the relentless reality of *g*. To see why, it is useful to review the most troublesome racial gap in *g*—that between Blacks and Whites. Like *g*, its effects in selection are highly predictable.

### The Predictable Impact of Racial Disparities in *g*

The roughly one standard deviation IQ difference between American Blacks and Whites (about 15 points) is well known. It is not due to bias in mental tests (Jensen, 1980; Neisser et al., 1996), but reflects disparities in the information-processing capabilities that *g* embodies (Jensen, 1998). Figure 2 shows the IQ bell curves for the two populations against the backdrop of the job complexity continuum. The point to be made with them—specifically, that patterns of disparate impact are predictable from group differences in *g*—applies to other racial-ethnic comparisons as well. The IQ bell curves for Hispanic and Native American groups in the United States are generally centered about midway between those for Blacks and Whites. The disparate impact of mental tests is therefore predictably smaller for them than for Blacks when *g* matters in selection. The bell curves for other groups (Asian Americans and Jewish Americans) cluster above those for Whites, so their members can usually be expected to be overrepresented when selection is *g* loaded. The higher the groups' IQ bell curves, the greater their overrepresentation relative to their proportion in the general population. It is the Black-White gap, however, that drives the flight from *g* in selection and thus merits closest attention.

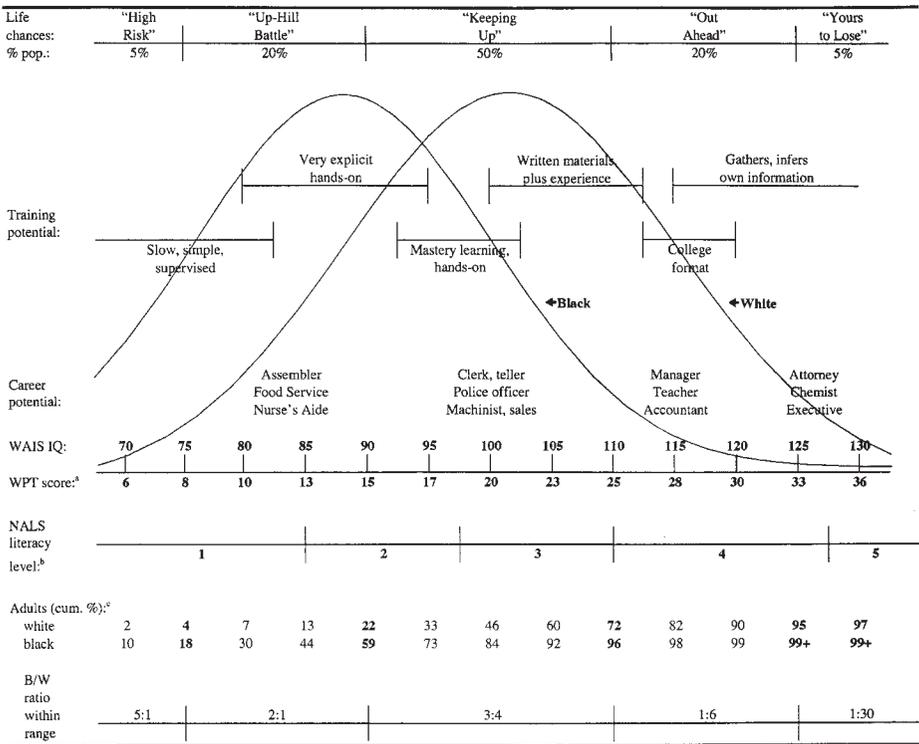


FIGURE 2 Adapted from Figure 3 in Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, 24, 79–132, with permission from Elsevier Science. <sup>a</sup>WPT = Wonderlic Personnel Test. <sup>b</sup>NALS = National Adult Literacy Survey. See Gottfredson (1997) for translation of NALS scores into IQ equivalents. <sup>c</sup>WAIS = Wechsler Adult Intelligence Scale. <sup>d</sup>See Gottfredson (1997) for calculation of percentiles.

The bell curves in Figure 2 are for representative samples of American Blacks and Whites. Racial disparities can differ somewhat from one setting to another for a host of reasons, so that the Black–White differences will sometimes be larger or smaller than those shown here. However, Figure 2 illuminates the big picture—namely, both populations in the context of the American economy. Specifically, it shows the two bell curves against the backdrop of the job complexity factor, which is arrayed along the “normal” range of the IQ continuum (from the threshold for borderline mental retardation to that for giftedness). Common occupations are arrayed along this continuum according to the IQ ranges from which they draw most of their incumbents. Those ranges therefore define the IQ ranges that make a person competitive for such work. Typical

modes of training that are possible (at the higher ranges of IQ) or required (at the lower ranges) at different IQ levels are also shown.

The cumulative percentages of American Blacks and Whites at each IQ level are shown at the bottom of Figure 2. The ratios in the last row represent the proportion of all Blacks to the proportion of all Whites within five different broad ranges of IQ. Blacks are overrepresented (5:1) in the lowest range (below IQ 75, labeled here as the “high risk” zone) and extremely underrepresented (1:30) in the highest (above IQ 125, the range where “success is yours to lose”). These ratios represent racial differences in the per capita availability of applicants who will be competitive for different levels of work, and they portend a clear trend in disparate impact. Under race-neutral hiring, disparate impact will generally be high enough to fail the 80% rule (which triggers the presumption of racial discrimination under federal guidelines) in hiring for all but the simplest jobs.

When Black and White applicants are drawn from the same IQ ranges, disparate impact will therefore be the rule, not the exception, even in jobs of modest complexity. It will get progressively worse at successively higher levels of education, training, and employment, and it will be extremely high in the most desirable jobs. Cognitive tests cannot meet the 80% rule with these two populations until the threshold for consideration falls to about IQ 77 to 78 (Gottfredson, 2000b). This low estimate is consistent with other research showing that mental tests have to be virtually eliminated from test batteries to satisfy the 80% rule under typical conditions (Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). The estimate also falls below the minimum mental standard (about IQ 80) that federal law sets for inducting recruits into the military.

To take some more specific examples, about 22% of Whites and 59% of Blacks have IQs below 90, which makes considerably fewer Blacks competitive for mid-level jobs, such as firefighting, the skilled trades, and many clerical jobs. The average IQ of incumbents in such jobs is nearer IQ 100, one standard deviation above the Black average of roughly IQ 85. IQ 80 seems to be the threshold for competitiveness in even the lowest level jobs, and four times as many Blacks (30%) as Whites (7%) fall below that threshold. Looking toward the other tail of the IQ distribution, IQ 125 is about average for professionals (e.g., lawyers, physicians, engineers, professors) and high-level executives. The Black–White ratio of availability is only 1:30 at this level. Disparate impact, and therefore political and legal tension, is thus particularly acute in the most complex, most socially desirable jobs.

Actual employment ratios are not as extreme as the per capita availability ratios shown here (other factors matter in hiring), but they follow the same systematic decline up the job complexity continuum. There is considerable IQ variability among incumbents in any occupation, of course, the standard deviation among incumbents generally averaging about 8 IQ points. The average Black–White difference is twice that large, however, which guarantees that Blacks will often cluster at the

lowest levels of performance when workers are hired randomly by  $g$  or with race-norming.

### Minimizing Selection for $g$ to Minimize Disparate Impact

The hope in personnel selection for a long time was that personnel psychologists could reduce disparate impact by developing technically better cognitive tests. If anything, improvements only worsened the disparate impact challenge because they resulted in more accurate measurement of  $g$ . Because good measurement has not provided the solution, it now tends to be treated as part of the problem, hence the popularity of test score banding in some quarters (it treats all scores within a specified range as equal), which reduces the reliability of measurement. Hence, also, the turning away from proven mental tests in major selection projects in favor of unproven “innovative” cognitive tests that substitute fidelity for validity and outside help for standardized conditions in taking the test. The suggestions that noncognitive tests can substitute for cognitive ones, or contextual performance for core performance, also promise to reduce the role of  $g$  in selecting workers. Such changes will do nothing, of course, to nullify the impact of lower  $g$  levels once workers are actually on the job.

One suggestion during the “Millennial Debate on  $g$ ” at the Society for Industrial/Organizational Psychology convention in 2000 was that the value of good worker performance itself has been overemphasized, that we have overstated its utility. Such suggestions reflect the impact-driven claim, growing even in I/O circles, that a racially-balanced workforce is at least as important as a competent one; or that racial parity might even be a prerequisite to productivity. Going further along this line of argument, one panelist warned that Blacks simply will not put up with disparate impact, implying that balance should be our primary concern. No one at the debate argued that  $g$  was unimportant. Nonetheless, the cumulative message from its doubters, implicit but persistent, was that (without the option to race-norm) progressive practice requires cutting back on the use of  $g$  in selection.

Some of the arguments for doing so were implicit appeals to discredited theories. For instance, the claim that we ought to be more reluctant to use mental tests because Blacks suffer from stereotype threat when taking tests amounts to a claim that highly cognitive tests are biased against Blacks. We already know this claim to be false. The typical cognitive test has been exonerated of bias against low-scoring minorities. Indeed, personnel psychologists know that mental tests overpredict performance when they are used in a race-neutral manner. Another untenable claim, still offered frequently and flush with promise, is that we can create equally valid cognitive tests with considerably reduced disparate impact. Any claim to have succeeded is suspect. “Innovative” formats, item types, and scoring procedures for tests have all been offered with fanfare in recent years, but to the extent that they reduce disparate impact, we must suspect that they have degraded selection for mental skills. The same is true for any impact-driven switch in perfor-

mance criteria. The vexing fact, which no tinkering with measurement can eliminate, is that Blacks and Whites differ most, on the average, on the most important predictor of job performance.

Some panelists also retreated into the unsubstantiated claim that there are multiple forms of intelligence, independent of *g*, that could predict job performance with less disparate impact. However, even the strongest body of evidence—that for so-called practical intelligence and its associated triarchic theory of intelligence (Sternberg et al., 2000)—provides only scant and contradictory bits of evidence for such a claim. Coming from a mere six studies (four of which remain unpublished) of five occupations, those data provide no support whatsoever (see Gottfredson, in press; also Brody, in press) for Sternberg et al.’s (2000, p. xi) assertion that “practical intelligence is a construct that is distinct from general intelligence and ... is at least as good a predictor of future success as is the academic form of intelligence [*g*]”.

Reducing disparate impact is a worthy goal to which probably all selection professionals subscribe. What is troubling are the new means being promulgated: minimizing or eliminating the best overall predictor of job performance. They amount to a call for reducing test validity and thereby violating personnel psychology’s primary testing standard. Reducing the role of *g* in selection may be legally and politically expedient in the short term, but it delays more effective responses to the huge racial gaps in job-relevant skills, abilities, and knowledges.

## REFERENCES

- Arvey, R. D. (1986). General ability in employment: A discussion. *Journal of Vocational Behavior*, 29, 415–420.
- Brody, N. (in press). Construct validation of the Sternberg Triarchic Abilities Test. (STAT): Comment and reanalysis. *Intelligence*, 30.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- 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.
- Gottfredson, L. S. (1994). *The role of intelligence and education in the division of labor*. Report No. 355. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.
- Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, 24, 79–132.
- Gottfredson, L. S. (2000a). Intelligence. In E. F. Borgatta & R. J. V. Montgomery (Eds.), *Encyclopedia of sociology, revised edition* (pp. 1359–1386). New York: Macmillan.
- Gottfredson, L. S. (2000b). Skills gaps, not mental tests, make racial proportionality impossible. *Psychology, Public Policy, and Law*, 6, 129–143.
- Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R.J. Sternberg & E. L. Grigorenko (Eds.), *The general intelligence factor: How general is it?* (pp. 331–380) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Gottfredson, L. S., Sternberg J. & Grigorenko E. L. (Eds.), *The general intelligence factor: How general is it?* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Gottfredson, L. S. (in press). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*, 30.
- Jenson, A.R. (1980). *Bias in mental testing*. New York; Free Press.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kirsh I. S., & Mosenthal P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25, 5-30.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.
- Reder, S. (1998). Dimensionality and construct validity of the NALS assessment. In M. C. Smith (Ed.), *Literacy for the twenty-first century* (pp. 37–57). Westport, CT: Praeger.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719–730.
- Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.
- Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist*, 50, 912–926.
- Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37, 60–82.

## APPENDIX 1

### 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” dimension of performance.

---

<sup>a</sup>See Gottfredson (2002) for fuller discussion and citations.

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 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 (effort and 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 trait of conscientiousness/integrity. Its effect sizes for core performance are substantially smaller than  $g$ ’s, however.

### Utility of $g$ Relative to the 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.