6 General mental ability

6.1 Introduction: what is intelligence?

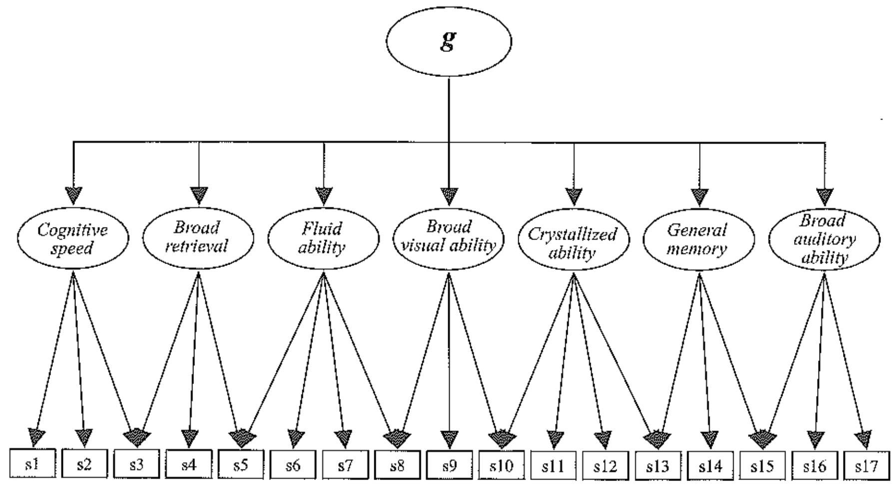
Intelligence is arguably the most important construct in personnel selection, but what is its precise meaning, and that of *IQ*, *g*, *general intelligence* and *cognitive abilities*? The following conceptual differences should be noted:

> IQ (a measure) versus *g* (a construct, the primary latent trait that IQ tests actually measure) versus general intelligence (often used as a synonym for *g*) versus intelligence (a lay word with multiple meanings; an umbrella term in science for a wide range of cognitive abilities). (Gottfredson, 2007, p. 219)

Unfortunately, the above terms are often used interchangeably, producing frequent misunderstandings and unnecessary discussions (that is, discussions based on semantic confusion rather than valid theoretical assertions or empirical evidence). In order to avoid falling into this category we will use each of these terms in the context of Gottfredson's (2007, p. 219) conceptual distinctions. Thus, we will define IQ as intellectual quotient, or the score on a comprehensive battery of tests or global intellectual ability scale (omnibus test) that has a standardised scale with a mean score of 100 and a standard deviation of 15 points (see Section 6.3). On the other hand, the terms '*g*', 'general intelligence' and 'general mental ability' (GMA) will be defined simply in terms of 'the ability to learn' or 'general learning ability' (Hunter, 1986) and used more or less interchangeably. A more detailed definition of GMA would read as follows:

> a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – 'catching on,' 'making sense' of things, or 'figuring out' what to do. (Gottfredson, 1997a, p. 13)

This definition was provided by a panel of fifty-two experts in the field and first appeared on the *Wall Street Journal* in 1994 in response to media and sociological criticisms of the psychological notion of intellectual ability after the publication of *The Bell Curve*. Most experts still agree that it captures the essence of the meaning of *intelligence*, but we will try to avoid using this term from here onwards in order to minimise confusion (with, for example, 'emotional intelligence' or



Note: s1…s17 = specific abilities at the lowest hierarchical level

Figure 6.1 *Graphical representation of the hierarchical structure of cognitive abilities identified by John Carroll (based on Carroll, 1993)*

'social intelligence'[1]), preferring to use *g* or GMA instead. Crucially, *g* and GMA emphasise the generality of human intelligence, setting it apart from the more specific *cognitive abilities* (e.g., spatial, numerical and visual). Whereas there is some debate as to how best to conceptualise specific cognitive abilities (Johnson & Bouchard, 2005), there is robust evidence for the existence of a general *g* factor at the broadest level. This factor accounts for 50–80% of the variance in multiple tests of specific abilities (Deary, 2001), leaving some variance unaccounted for. The unaccounted variance is explained first in terms of broad abilities (e.g., cognitive speed, broad retrieval, general memory, etc.), which, in turn, can be broken down into lower-level factors or aspects of cognitive ability, such as induction, language development and word fluency. This *hierarchical structure* of cognitive abilities (graphically depicted in Figure 6.1) represents the state-of-the-art approach to classifying human aptitudes and places *g* or GMA at the highest hierarchical level; it has been found to explain hundreds of robust datasets (Carretta & Ree, 1996; Carroll, 1993; Jensen, 1998; Spearman, 1904).

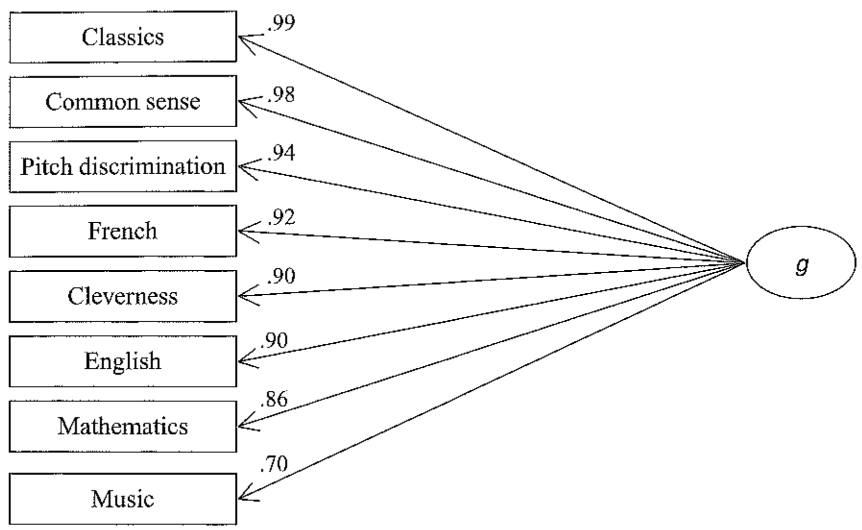
In purely statistical terms, the *g* factor simply indicates that people's scores on various cognitive ability tests are highly intercorrelated, making differences *within* individuals quite negligible in comparison to differences *between* individuals (which is the same as saying that your scores on different tests are much more similar to each other than your overall score is to the overall score of other people). To illustrate this with another example: suppose you know someone who is not just good at football, but also good at tennis, basketball, rugby, sailing and marathon running. You would probably refer to that person as a good sportsman or *sporty* person. Conversely, imagine you know someone who is bad at all possible sports (not just at one); you would probably think of that person as a bad sportsman or *un-sporty* person. While we can probably all think of people who are good at some sports but not all, when it comes to cognitive ability tests, people who do very well at one or two tests but very badly at others are a notable exception to the norm, though it has been shown that specific abilities are more weakly intercorrelated at higher levels of GMA (Legree, Pifer & Grafton, 1996).

The vast amount of accumulated evidence in support of the *g* factor undermines the validity of alternative ability models that de-emphasise the importance of *g* (Sternberg & Wagner, 1993) or question its usefulness (Guilford, 1988). This applies especially to Gardner's theory of 'multiple intelligences' (Gardner, 1983/2003, 1993b), which assumes the independence of cognitive abilities, as well as the inclusion of non-traditional competencies, such as 'spiritual intelligence' and 'body-kinaesthetic intelligence', in the realm of cognitive abilities (for emotional intelligence, see Sections 7.19 to 7.23). Regardless of the popular appeal of this theory and of any progress – so far there has been little (Chamorro-Premuzic, 2007) – that may come in relation to the measurement of novel abilities herewith defined, few psychological findings are as robust as *g*, that is, the idea that different abilities are highly intercorrelated (Carroll, 1993; Jensen, 1998). It is also noteworthy that even though laypeople are seemingly reticent to accept the idea of *g*, ratings of their own cognitive abilities tend to be strongly intercorrelated, even when novel abilities, such as spiritual and existential abilities, are estimated in relation to mathematical and spatial abilities (Chamorro-Premuzic, Ahmetoglu & Furnham, 2008; Chamorro-Premuzic, 2007; Chamorro-Premuzic & Furnham, 2006a; Furnham & Chamorro-Premuzic, 2007).

The theoretical meaning of *g* is not to be debated in detail here, but we examine why it comes to predict job performance and other occupational outcomes in Section 6.4. It should be noted that, for the purposes of this book, what matters most is how good *g* is at predicting (positively) desirable or (negatively) undesirable occupational outcomes. In that sense, whether we agree or not with the idea that *g* is the best marker of human intelligence really is quite irrelevant.

6.2 A brief history of GMA and Raven's Progressive Matrices

The construct of GMA is widely attributed to Spearman's (1904) finding that performance on different mental tests, as well as grades on different academic subjects, are largely intercorrelated, such that performance on these tests can be explained quite accurately in terms of a general psychometric factor, which he called *g*. Theoretical interpretations of this statistical factor have varied quite markedly, though a majority of intelligence researchers endorse the conceptualisation of *g* as a mainly, if not purely, biological, largely unchangeable, culture-free, measure of individual differences in learning and reasoning capacity, best measured through non-verbal tests, such as Raven's Progressive Matrices (see Figure 6.3). However, Spearman's (1904, p. 276) original data showed that



Note: Correlations corrected for measurement error

Figure 6.2 *Some correlates of Spearman's* g *factor (after Spearman, 1904)*

g correlated almost perfectly with knowledge of Greek and Latin (the 'Classics'), which are obviously dependent on previous knowledge and education. Moreover, pitch discrimination, French and 'common sense' correlated more highly with g than mathematics did (see Figure 6.2). Nonetheless, Spearman later shifted his interpretation of g from a more verbal, knowledge and content-based to a more abstract, non-verbal and process-based conceptualisation of g, regarding Raven's Progressive Matrices as the ultimate measure of GMA (Spearman, 1930, 1938). Clearly, Spearman was more interested in the biological aspects of GMA than in the development of adult intellectual abilities (Ackerman, 2008).

The Raven's Progressive Matrices (Raven, Court & Raven, 1998) is a non-verbal, multiple-choice test of abstract reasoning developed in the UK by John C. Raven in 1938 (Raven, 1938). It is widely regarded as a – if not the best single – measure of GMA (Jensen, 1998), though see Carpenter, Just and Schell (1990) and Mackintosh and Bennett (2005) for different views. The test presents test-takers with several series of figures linked by specific implicit rules (see Figure 6.3). Test-takers are required to complete the sequence by identifying the missing figure from a number of possible alternatives given, which requires understanding of the implicit rule by which the given figures are connected. The test has sixty items and is administered in twenty minutes, though different versions exist. The sixty items fall into sets of twelve items and increase in difficulty. Abundant evidence of the test's reliability and validity for a wide range of ages and cultural groups, including clinical and non-clinical groups, has been reported in the manual. The test is used extensively in personnel selection, particularly in the US and the UK (Bertua, Anderson & Salgado, 2005; Jensen, 1998; Raven *et al.*, 1998). Given

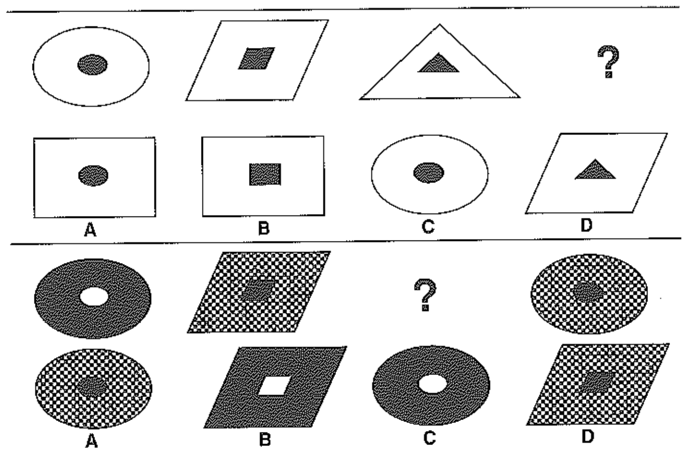


Figure 6.3 *Two examples of Raven-like items*

that this is a non-verbal measure, the Raven's Matrices can be administered in every nation without the need to translate anything more than the instructions. This culture-free element of the test makes it particularly attractive when one wishes to compare the reasoning ability and learning potential of culturally heterogeneous groups of candidates (e.g., applicants from rural China, urban Chile and suburban Germany applying for a postgraduate programme of studies in London or a banking job in New York), which is a common need in globalised job markets and has become increasingly feasible with the growth of the World Wide Web beyond the industrialised world.

6.3 A brief history of IQ and the Wonderlic Personnel Test

At the same time as Spearman's investigations on the g factor in the UK, a different approach to the measurement of cognitive abilities was being pioneered in France by Binet and Simon (Binet & Simon, 1905/1973). Unlike Spearman, Binet and Simon were interested in the development of intellectual abilities and 'sum of acquired knowledge' (p. 40), so that their battery of mental tests was directly designed to address *change* in cognitive abilities as a function of chronological age. Binet and Simon's original battery of tests comprised thirty tasks of increasing complexity, ranging from simple items, such as assessing whether children managed to shake hands with the test administrator, to more complex, intellectual problems, such as testing children's ability to memorise drawings and random sequences of digits, to come up with rhymes and to answer questions such as 'My neighbour has been receiving strange visitors. He has received in turn a doctor, a lawyer, and then a priest. What is taking place?' Thus Binet and Simon's approach very much highlighted the practical or common-sense aspects of cognitive ability (as noted in Section 6.2, this interpretation was congruent with Spearman's early findings). Indeed, their definition of cognitive ability stated that:

> It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which is of the utmost importance for practical life. This faculty is judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. A person may be a moron or an imbecile if he is lacking in judgment; but with good judgment he can never be either. Indeed the rest of the intellectual faculties seem of little importance in comparison with judgment. (Binet & Simon, 1916/1973, pp. 42–3)

Children's IQ scores are calculated by dividing their mental age (i.e., the average chronological age of the children who solve problems of similar complexity or level of difficulty) by their chronological age, and then multiplying that by 100. Thus an IQ of 120 may indicate that a 10-year-old child has the mental ability of a typical 12-year-old, i.e., 12/10* 100 = 120, which would make that child more than 1 standard deviation smarter than most children. This distribution means that 68 per cent of individuals score between 85 and 115 points, and this distribution applies to adults as well as children.

However, as the effects of age on most cognitive abilities vanish by the age of 15, the IQ formula cannot be directly applied to adults. Early attempts to do so, when IQ tests were first implemented in the context of personnel selection, created interpretational problems. Most notably, when an American adult adaptation of Binet and Simon's scale (the Army Alpha test) was administered to 1.7 million conscripts, the reported average age for the sample was 13 years (Ackerman, 2008). However, adult IQ norms were implemented for several scales soon thereafter.

The most widely used adult IQ test (for 16 to 89-year-olds) is the Wechsler Adult Intelligence Scale (WAIS), currently in its third edition (Wechsler, 2002).[2] The first version of the test was designed for the US army seventy years ago by David Wechsler (a student of Spearman's) to measure 'the global capacity of a person to act purposefully, to think rationally, and to deal effectively with his/her environment' (Wechsler, 1939, p. 229). However, the WAIS (which includes fourteen scales and measures the four broad components of verbal comprehension, perceptual organisation, working memory and processing speed) is used primarily in clinical settings, not least because of its lengthy administration time, 75 minutes on average. In research and even in most clinical settings, subscales of this test are employed more often than the full battery. In occupational settings, particularly for personnel selection, shorter IQ tests or omnibus GMA tests are always preferred.

The most widely used IQ test in personnel selection is probably the Wonderlic Personnel Test (Wonderlic, 1992) (Murphy, 1984; Schmidt & Hunter, 2004), an omnibus GMA test developed by Eldon F. Wonderlic in 1937 and used by the US Navy for the selection and training of pilots during World War II. The main advantage of this test is that it is administered in only twelve minutes, meaning it is less intimidating for job applicants and more efficient for employers than long

Table 6.1 *Wonderlic Personnel Test: sample items (adapted from Wonderlic, 1992)*

1. When rope is selling at 20 cents a foot, how many feet can you buy for 60 cents?
2. Assume the first two statements are true. Is the final one: True, False, Uncertain?
   1. The girl is a hockey player.
   2. All hockey players wear hats.
   3. The girl wears a hat.
3. Paper sells for 14 cents per pad. What will four pads cost?
4. How many of the five pairs of items listed below are exact duplicates?
   Ackerman, P. L., Ackerman, P. L., Gottfredson, L. S., Gottfredson, M. R., Thorndike, E. L., Thorndike, P. L., Vernon, P. E., Vernon, P. A., Bach, J. S., Bach, J. A.
5. PREVENT PRESENT – Do these words
   a. Have similar meanings
   b. Have contradictory meanings
   c. Mean neither the same nor opposite?

- **Answers: 1** = *3* / **2** = *True* / **3** = *56 cents* / **4** = *1 (Ackerman, P. L.)* / **5** = c

IQ tests. Despite its short administration time, the Wonderlic produces overall IQ scores that correlate at $r = .93$ with the WAIS and in 80–90 per cent of cases are within 10 points difference from scores derived from the WAIS (Dodrill, 1981). Moreover, the test–retest reliability of the Wonderlic ($r = .94$ in five years) approaches that of the WAIS ($r = .96$) (Dodrill, 1983).

Modified sample items of the Wonderlic are shown in Table 6.1. Even a quick glance at these items suggests a marked difference from the Raven-like items shown in Figure 6.3, namely that Wonderlic (but not Raven) items are mostly verbal. Indeed, the distinction between verbal and non-verbal ability tests is at the heart of a fundamental theoretical distinction frequently applied to the realm of cognitive abilities, that is, the difference between *fluid* and *crystallised* abilities (Cattell, 1943). Fluid intelligence, usually notated as *gf*, refers to individual differences in abstract reasoning and is therefore equivalent to Spearman's late interpretation of *g* and measured by tests such as Raven's Progressive Matrices (see Section 6.2). Crystallised intelligence, usually notated as *gc*, refers to individual differences in acquired knowledge and is therefore measured by tests that require information retrieval (e.g., vocabulary, general knowledge and learned rules).

Although it has been recently questioned whether *gf* and *gc* provide an accurate grouping of cognitive abilities (Johnson & Bouchard, 2005), and whether one can truly measure *gc* without eliminating or 'decanting' variance caused by *gf* (Reeve, Meyer & Bonaccio, 2006), this distinction has important implications for personnel selection because it acknowledges that certain tests are more useful for measuring what people have already learned and already know, whereas other tests are more useful for providing an indicator of what people will be able to learn. For instance, being a physicist or a mathematician will probably require high levels of *gf*, whilst being a historian or a lawyer will probably require high levels of *gc* (McGrew & Flanagan, 1998).

The Wonderlic Personnel Test has been reported to be correlated equally with both *gf* and *gc*, thus measuring both abstract reasoning capability and knowledge possessed (Bell, Matthews, Lassiter & Leveret, 2002). In that sense, the two conceptual differences between IQ and GMA are that IQ tests have standardized scores and also measure *gc*.

6.4 Like no other: the predictive power of GMA

To say that GMA predicts occupational outcomes, such as job or training performance, is as much a truism as an understatement, and is really beyond debate (Murphy, 2002; Schmidt, 2002). Indeed, there is so much evidence for the validity of GMA in the prediction of job and training performance that an entire book could be written simply describing these findings. There are several great and relatively compact sources of reference, the most robust and widely cited (Hunter & Hunter, 1984; Judge, Higgins, Thoresen & Barrick, 1999; Schmidt, 2002; Schmidt & Hunter, 1998, 2004) are discussed in this and the forthcoming sections.[3]

The predictive power of GMA at work is rivalled by no other psychological trait (Ree & Earles, 1992). That said, GMA should not be used as single predictor of job performance as some traits, notably Conscientiousness and Integrity (discussed in Section 7.4), have incremental validity over and above GMA, explaining additional variance in occupational outcomes of interest (Bobko, Roth & Potosky, 1999; Schmidt & Hunter, 1998). The validity of GMA at work has been documented quite systematically since the end of World War I (Harrell & Harrell, 1945; Yerkes, 1921), first in military and then in civil occupations. Yet – as shown in Figure 6.4, based on Dany & Torchy (1994) – most employers do not test for aptitude when selecting their workforce.

The first seminal reviews on the subject (Ghiselli, 1966, 1973) concluded that GMA, as well as spatial, mechanical and perceptual abilities, were the best aptitude predictors of training and job performance, and that for every family of professions (e.g., managerial, clerical, sales and drivers) there is a cognitive ability test that is at least moderately related to these outcomes.

As shown in Table 6.2, average GMA levels tend to increase with occupational level, that is, with the prestige of the job (Jensen, 1980). Furthermore, higher-level jobs tend to have substantially higher levels of inbound GMA, indicating that it is far more unlikely to find low-IQ scorers in high-level professions than it is to find high-IQ scorers in low-level professions. This is probably because laziness can drive bright people to simple jobs, but hard work can hardly push dim people to

Table 6.2 *GMA across civilian jobs in US Army; simplified adaptation of original source (Harrell & Harrell, 1945)*

Jobs	*N*	*M*	SD	Range
Accountants	172	128.1	11.7	94–157
Lawyers	94	127.6	10.9	96–157
Repairman	96	115.8	13.1	76–149
Cashier	111	115.8	11.9	80–145
Tractor driver	354	99.5	19.1	42–147
General painter	440	98.3	18.7	38–147
Farmer	700	92.7	21.8	24–147
Farmhand	817	91.4	20.7	24–141

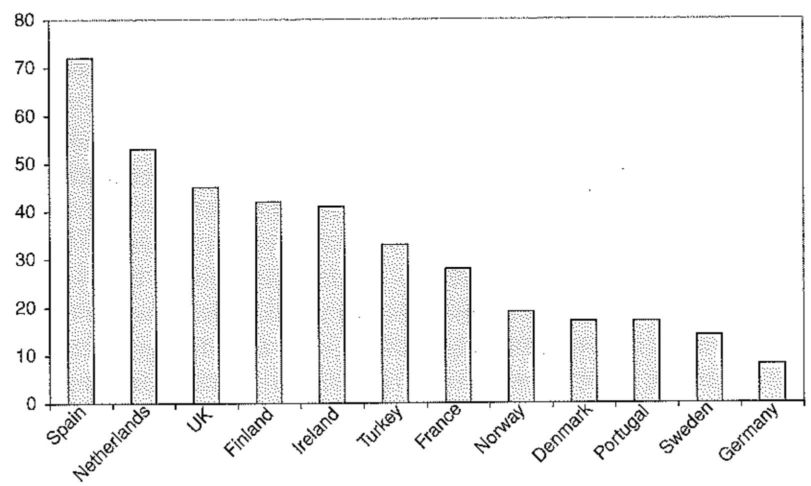


Figure 6.4 *Percentage of employers using aptitude tests in Western Europe (Price Waterhouse Cranfield survey, 1994)*

complex jobs (note, of course, that most but not all complex jobs pay higher wages than simple ones). In addition, the standard deviations tend to decrease as people move up to higher-level professions, showing that these jobs tend to have not only people with higher but also more homogeneous GMA (see also Figure 6.5, taken from Chamorro-Premuzic, 2007, and based on Gottfredson, 2004).

Although the notion that better jobs attract brighter people is not really counterintuitive (think of how many times you surprised yourself when people who did not seem very bright told you they have very well-paid jobs), the strength of the association between GMA and job success is robust. Ratings of occupational level, which are extremely reliable (even when only two or three individuals are asked to independently rank different occupations, they provide very homogeneous rankings), correlate as highly as r = .95 with average IQ scores of people in those jobs (Jensen, 1998). Even at the individual (rather than aggregate) level,

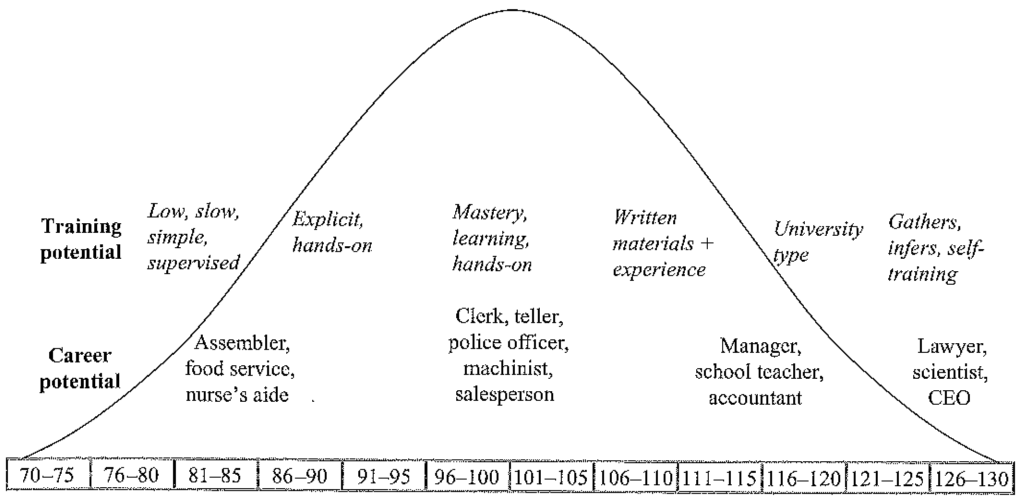


Figure 6.5 *Occupational consequences of IQ (based on Gottfredson, 2004)*

the correlation between occupational level and GMA has been reported to be as high as r = .72, meaning that about half of the variance in occupational level can be explained by GMA alone (Schmidt & Hunter, 2004).

Studies examining job complexity yield very similar results. Indeed, it has been noted that occupational prestige is a direct function of both job complexity and GMA (Gottfredson, 1997b). Complex jobs are characterised by a cognitive diversity and a wide range of requirements, such as gathering, relating, analysing and combining information through spoken and written means under situations of high responsibility and low supervision and substantial psychological pressure. Conversely, low prestige and low-complexity situations – where GMA plays a much more limited role – are characterised by repetitive, systematic, monotonous and physical rather than intellectual tasks; they are also more highly supervised and demand less responsibility from the employee (see Gottfredson, 1997a, pp. 100–5) (see also Figure 6.5, taken from Chamorro-Premuzic, 2007, and based on Gottfredson, 2004).

In a colossal quantitative review and meta-analysis of 425 studies on GMA and job performance across different levels of complexity (Hunter, 1980; Hunter & Hunter, 1984), typically referred to as 'validity studies', GMA was found to correlate significantly with performance at *all* levels of job complexity, though it is clear that the more complex the job, the more important GMA is; hence the common assertion that the relationship between GMA and job performance is *moderated by job complexity*. Indeed, it has been noted that job complexity is one of the few moderators of the effects of GMA on job performance (Murphy, 2002, p. 175).

Table 6.3 summarises the main results from Hunter and Hunter's (1984) meta-analysis. The authors operationalised job performance in terms of supervisory ratings, which are known to be less reliable than objective measures of performance,

Table 6.3 *GMA correlates of job and training performance across various job complexity levels (adapted from Hunter, 1980; Hunter & Hunter, 1984)*

Job complexity	% workforce	Job performance 425 studies	Training performance 90 studies
1 *most complex*	14.7	.58	.59
2	2.5	.56	.65
3	62.7	.51	.57
4	17.7	.40	.54
5 *least complex*	2.4	.23	not reported
Overall *N*		32,124	6,496

and training performance in terms of the degree of learning during training. Correlations corrected for error in the outcome measures but not in the predictor. Thus at the construct level the authors estimated that correlations would be approximately 8.5 per cent higher.

Hunter and Hunter coded different occupations and their complexity level according to the Dictionary of Occupational Titles.[4] Examples of complex jobs (level 1) were professional, scientific and management jobs, where GMA correlated r = .58 with job performance and r = .59 with training performance. Examples of low-complexity jobs (level 5) were feeding/off-bearing (that is, placing or removing materials from machines that are automatic or operated by others), where GMA correlated r = .23 with job performance. On the other hand, Hunter (1986) tested the predictive power of GMA against objective performance measures and found a predictive validity of r = .75.

Subsequent meta-analyses in the US were by and large congruent with Hunter's findings (see Table 6.4 for a summary) and highlight the strength of the effects of both GMA and specific abilities on both job and training performance across different occupations. A special issue of the journal *Personnel Psychology* (edited by J. P. Campbell, 1990) was devoted entirely to the discussion of the 'Project A', a huge US-based personnel selection project that cost $25 million and was 'probably the largest and most expensive selection research project in history' (Schmidt *et al.*, 1992, p. 632). Findings showed that military job performance, as operationalised via both technical and general soldiering proficiency, was strongly related to GMA, and that GMA predicted these outcomes better than any other psychological trait. Athough non-ability traits, such as personality, were better predictors of supervisory ratings of effort, leadership and discipline (McHenry, Hough, Toquam, Hanson & Ashworth, 1990), they provided only limited incremental validity over GMA in predicting job performance. Thus project A largely replicated previous findings on the validity of GMA (Schmidt *et al.*, 1992).

More recently, meta-analyses were also conducted in the UK and Europe (see Sections 6.5 and 6.6).

Table 6.4 *Other meta-analyses on the validity of GMA since 1980*

Reference	Outcomes	Validities
(Pearlman, Schmidt & Hunter, 1980)	Clerical occupations (job and training)	Moderate to large (generalise across different samples and jobs)
(Schmidt, Gast-Rosenberg & Hunter, 1980)	Computer programmers (job proficiency and training)	Very large (.73 for job proficiency and .91 for training)
(Hirsh, Northrop & Schmidt, 1986)	Law enforcement jobs (job proficiency and training)	For job proficiency modest (but only spatial-mechanical ability generalised), for training large
(Nathan & Alexander, 1988)	Supervisory ratings and rankings, work samples, production quality and quantity	Moderate to high (.44 for supervisory ratings to .60 for work sample criteria); lower validities for objective job performance but they were not corrected for unreliability
(Hartigan & Wigdor, 1989)	Reanalysed Hunter and Hunter's (1984) data plus new datasets	Substantial (.80 for job performance, .81[5] for training) Job complexity moderates validity of GMA and abilities
(Levine, Spector, Menon, Narayanan & Cannon-Bowers, 1996)	Crafts jobs in utility industry (e.g., mechanical utility, telephone technical and electrical assembly) (job performance)	Moderate for job performance
(Vinchur, Schippman, Switzer & Roth, 1998)	Sales jobs (job performance)	Moderate (.40 for GMA)

6.5 GMA in the UK

Although measures of GMA are used more widely for personnel selection in the UK than the US (Salgado & Anderson, 2002), until recently there was only scattered evidence for the validity of GMA in the prediction of job-related outcomes in UK studies, with most studies traditionally focusing on US data (Anderson, Born & Cunningham-Snell, 2001). However, a comprehensive meta-analysis of UK validity studies (283 independent samples and 75,311 people) has recently provided robust evidence for the power of GMA and specific abilities in the prediction of job performance and training across a wide range of occupations (Bertua, Anderson & Salgado, 2005). Table 6.5 reports the key findings. As shown, the validity of GMA in UK samples is comparable to that reported in US studies (see again Table 6.4). In addition, the UK results highlight the importance of specific cognitive abilities, particularly perceptual and numerical abilities in relation to both job and training performance, and verbal abilities in relation to training performance. Although the relevance of these specific abilities in UK studies overshadows the validity of specific abilities reported in the US by Hunter (1983b) (see also Section 6.8 below), three important issues should be noted. First, Bertua *et al.* did not test for incremental validity, which means it is uncertain whether the validities reported for specific abilities would have held if GMA had been controlled for. Second (a related issue), each of the specific abilities tested in the UK meta-analysis, especially those that showed the highest validities for both job and training performance, is *g*-saturated, meaning their correlations with job and training performance are at least in part a function of GMA. Third, the UK meta-analysis included many more recent studies than those reported in the above US studies.

Table 6.5 *Ability validities for job and training performance in the UK (based on Bertua* et al., *2005)*

	Job performance		Training performance	
Measure	*N* (60 studies)	Corrected *r*	*N* (223 studies)	Corrected *r*
GMA	2,469	.48	17,982	.50
Verbal	3,464	.39	12,679	.49
Numerical	3,410	.42	15,925	.54
Perceptual	1,968	.50	13,134	.50
Spatial	1,951	.35	15,591	.42
Total	13,262	.42	75,311	.49

Despite these minor methodological differences, one can only conclude that the UK studies on GMA and job performance and training mirror the findings from the US. This is in line with the reported overlap in choices of test for measuring GMA. Indeed, several US-designed tests are commonly used in the UK (e.g., the Differential Aptitude Test and the Minnesota Clerical Test) and vice versa (see Section 6.2 on the Raven's Progressive Matrices).

The UK data also replicated the positive association between the predictive power of GMA and the complexity levels of the job. Figure 6.6 depicts the corrected validities for GMA across various occupations in predicting job (white bars) and training (black bars) performance. As can be seen, professions that are more intellectually demanding, such as professional, engineering and managerial, require more GMA than those less cognitively complex professions (e.g., operator and driver). Interestingly, for complex jobs, GMA is a better predictor of job performance than of job training, whereas for simpler jobs, training seems to require more GMA than performance. This pattern of results is

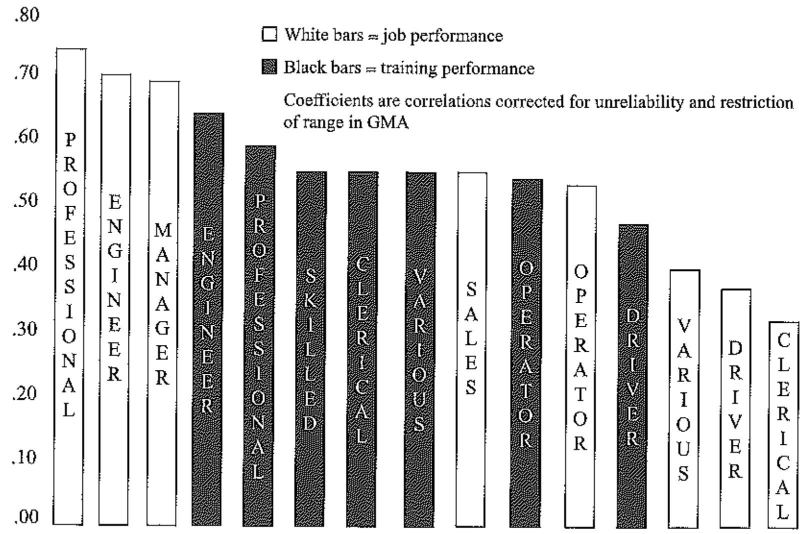


Figure 6.6 *Validity of GMA across occupations in the UK (based on Bertua* et al., *2005)*

consistent with the idea that GMA is especially important for explaining individual differences in learning, which, in low-complexity professions, may not be that important after training, but, in high-complexity professions, may be needed especially in the job. However, it is clear (as much in the UK as in the US data) that GMA matters in every job and for both training and performance. Thus the authors concluded that:

> Selection practitioners and HR professionals in UK organisations should be encouraged to use psychometrically developed cognitive ability tests regardless if job type, hierarchical seniority, potential future changes in job role composition, or whether the tests are principally for general or specific abilities. Moreover, these findings highlight the importance of research-based practice in selection psychology and provide unequivocal evidence for the continued and expanded use of GMA tests for employee selection in UK organisations. (Bertua *et al.*, 2005, p. 403)

6.6 GMA in Europe

Studies in the European Community (EC) echo the pattern of results from US and UK studies. A relatively recent examination of sixteen EC-member nations revealed that GMA tests are used in personnel selection more frequently in European than in US organisations, even though there is a larger percentage of small- and medium-size organisations in the EC than in the US (Salgado & Anderson, 2002). Another difference between the EC and the US is that EC nations are very homogeneous (i.e., there is little variability within countries in the EC) with regard to their testing policies and practices, which vary hugely across the US (Viswesvaran & Ones, 2002). But how comparable are the US validity studies of GMA to those of the EC?

A major meta-analysis (Salgado, Anderson, Moscoso, Bertua & De Fruyt, 2003) examined the results from ten EC nations (N range from 946 to 16,065). Salgado *et al.* reported operational validities (corrected correlation coefficients) for GMA of .54 for training and .62 for job performance. Specific cognitive abilities were also found to account for a substantial amount of variance of job and training performance. For training performance, corrected correlations ranged from .25 for perceptual ability to .48 for numerical ability. For job performance, corrected correlations ranging from .35 in the case of verbal ability to .56 in the case of memory ability. Although some of these validities approached the validity of GMA, as in the UK meta-analysis (see Section 6.5), the authors did not examine the incremental validity of specific abilities over and above GMA. However, verbal, spatial, numerical, perceptual and memory abilities had lower validities than GMA. Thus, based on the clear similarity between their own and US findings, Salgado *et al.* concluded that '[T]here is international validity generalisation for GMA and specific cognitive abilities for predicting job performance and training success. In order words, the criterion validity of cognitive measures generalises across different conceptualisations of job performance and training, differences in unemployment rates, differences in tests used, and differences in cultural values, demographics, and languages' (p. 592).

Salgado *et al.* also noted a few differences between their own and US results. At the level of specific abilities, perceptual ability had lower validities predicting training performance in the EC than in the US. At the GMA level, validities were larger in the EC for job performance, but larger in the US for training performance. The authors explained the former in terms of the higher-complexity jobs of the EC jobs examined (compared to the US studies, which included more lower-complexity jobs), and the latter in terms of the fact that US studies tended to measure training performance objectively, whereas EC studies tended to assess it subjectively (i.e., supervisory ratings) (for additional explanations see p. 593 of their article). In all, though, there are clearly more similarities than differences among EC, US and UK studies, and other international studies can be expected to yield very similar validities for GMA and specific abilities. It should, however, be noted that the US, UK and EC studies all looked at GMA in the context of industrialised and globalised economies. Given that the validity of GMA is moderated by job complexity, and since industrialised economies are characterised by a higher proportion of intellectually demanding jobs, one would expect somewhat lower predictive validities for GMA in undeveloped and developing nations, but meta-analytic evidence for these nations is yet to be provided (Lievens, 2007).

In another meta-analysis the same group of researchers examined the validity of GMA across different occupations in the EC (Salgado, Anderson, Moscoso,

Table 6.6 *Ability validities for job and training performance in the EC (Salgado* et al.*, 2003a)*

Measure	Job performance		Training performance	
	N (*93 studies*)	Corrected *r*	*N* (*97 studies*)	Corrected *r*
GMA	9,554	.62	16,065	.54
Verbal	4,781	.35	11,123	.44
Numerical	5,241	.52	10,860	.48
Spatial-mechanical	3,750	.51	15,834	.40
Perceptual	3,798	.52	3,935	.25
Memory	946	.56	3,323	.34

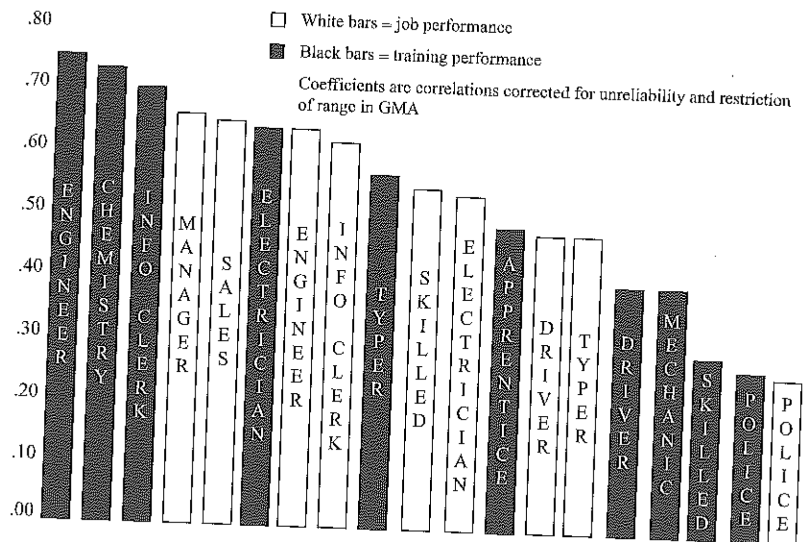


Figure 6.7 *Validity of GMA across occupations in the EC (based on Salgado* et al.*, 2003)*

Bertua, De Fruyt & Rolland, 2003b). Results are graphically depicted in Figure 6.7. As shown, and similarly to Figure 6.6, validities for GMA ranged across occupations for both job and training performance, with GMA having moderate to substantial validities for job performance on all occupations except the police, where it had only modest validity. As regards training, validities showed an even wider range, with the weakest being police force jobs. These results replicate the findings of US meta-analysis (Hirsh, Northrop & Schmidt, 1986; Pearlman, Schmidt & Hunter, 1980), even with regard to the lower validity of police jobs. Thus the authors concluded that '[P]revious meta-analytic findings for the North American continent and now our findings for the European continent unequivocally support the use of tests of GMA for personnel selection regardless of job complexity or job occupation being selected for' (Salgado *et al.*, 2003b, p. 1077).

6.7 Longitudinal evidence

Whilst powerful, the above reviewed studies provide no longitudinal evidence for the predictive power of GMA, making interpretation of causational paths largely speculative. However, there is equally impressive evidence for the longitudinal validity of GMA in the prediction of job performance.

One of the first compelling demonstrations of the longitudinal predictive power of GMA in the context of job outcomes was published by Austin and Hanisch (1990). The authors analysed data from a large sample of the Project Talent dataset and found that GMA measures obtained during high school years predicted occupational attainment eleven years later. Two subsequent studies by Wilk and Sackett provided additional support for the longitudinal importance of GMA, showing that, over a period of five years, people with higher GMA tend to move upwards in the job level scale, whereas lower GMA people tend to move downwards (Wilk, Desmarais & Sackett, 1995). Moreover, congruency between GMA levels and typical GMA levels found for each job also predict people's likelihood of moving upwards or downwards (Wilk & Sackett, 1996).

Quite astonishingly, the longitudinal associations tend to hold even when socioeconomic status (SES) is taken into account. Thus, within the same family (where SES levels are the same for every member) family members with higher GMA tend to have significantly better jobs, and earn more, than their lower GMA relatives (Murray, 1998). In fact, at 1993 figures, and controlling for SES, people with average levels of GMA (IQ = 100) earned almost $20,000 less than siblings who were 20 IQ points brighter, and almost $10,000 more than siblings who scored 20 IQ points lower. These findings are consistent with earlier longitudinal data showing that, when GMA levels go up from one generation to the next, SES levels go up accordingly (Mascie-Taylor & Gibson, 1978). In a similar vein, Jencks *et al.* (1972) reported that sons' occupation and income levels correlated more significantly with their own IQ than with the occupation or educational level of their fathers. Recent studies have replicated these findings. For example, Deary and colleagues found that childhood GMA accounted for 23.2 per cent and parental social class for 17.6 per cent of the total variance in social status attainment in mid life (Deary *et al.*, 2005).

The most compelling evidence for the longitudinal validity of GMA in the prediction of occupational level and income was provided by a study spanning back almost four decades (Judge *et al.*, 1999). The authors reported correlations between GMA at age 12 and occupational level ($r = .51$) and income ($r = .53$) almost forty years later. Moreover, a reanalysis of these data (which also included the Big Five personality traits) estimated that the predictive power of GMA was almost 60 per cent higher than that of Conscientiousness (the trait that came second) (Schmidt & Hunter, 2004). Thus even very early measures of GMA – which can be expected to change quite a bit after the age of 12 (Ackerman, 2008, pp. 461–8) – seem to account for over 25 per cent of the variance in occupational status and income over almost two generations later, in a way that no other psychological traits can match.

6.8 Why, then, is IQ so unpopular?

Given the vast amount of compelling evidence for the predictive validity of GMA in the workplace, one has to wonder why IQ has remained such an unpopular and politically contested concept for many decades (Cronbach, 1975; Murphy, 2002). Although the reasons are multiple, let us at least examine some of the most probable causes for the 'bad image' of IQ inside and outside academic and occupational settings. Let us start with the academic counterclaims to the validity and importance of GMA.

A common academic counterclaim (though, given the above-reviewed evidence, the extent to which such claims may be considered academic is rightfully questionable) is that the predictive power of GMA is largely dependent on the specificity of the outcome measure or context. In simple terms, this argument refers to the idea (admittedly intuitive) that no two jobs are the same, and that IQ scores will inevitably be associated with some jobs more than others (just as, say, biodata items may be predictive of certain jobs but not others, as seen in Chapter 3). The alert reader will have already noted that the data reported above is in stark conflict with this assumption (for instance, Tables 6.3 and 6.4 refer to a wide range of studies looking at various types of jobs). Thus the 'situational specificity' argument – as this idea is called – is hard to accept unless one is unaware of the wealth of data mentioned above (see also Hunter & Hirsh, 1987). A more viable way of putting this argument forward, however, may be to argue that *specific* cognitive abilities rather than GMA are necessary to predict job outcomes accurately, or, at least, that specific abilities would be more appropriate predictors of specific job types (if, say, one were to rely on measures of spatial ability measures to predict how quickly an employee would learn a spatial task, e.g., driving, arranging furniture or reading maps). This point is still very intuitive, and respects even the original conceptualisation of GMA (Spearman, 1904), which left some room for specific abilities; but what do the data tell us?

When both specific and general abilities are taken into account, any variance in specific abilities unaccounted for by the general factor *g* (GMA) seems to be unrelated to job outcomes (Hunter, 1986; Jensen, 1986; Thorndike, 1986). This was the finding of another large-scale study by Hunter, looking at 20,256 US marines' training performance (Hunter, 1983b). As shown in Figure 6.8, when the latent factor of GMA is considered, training performance is not predicted

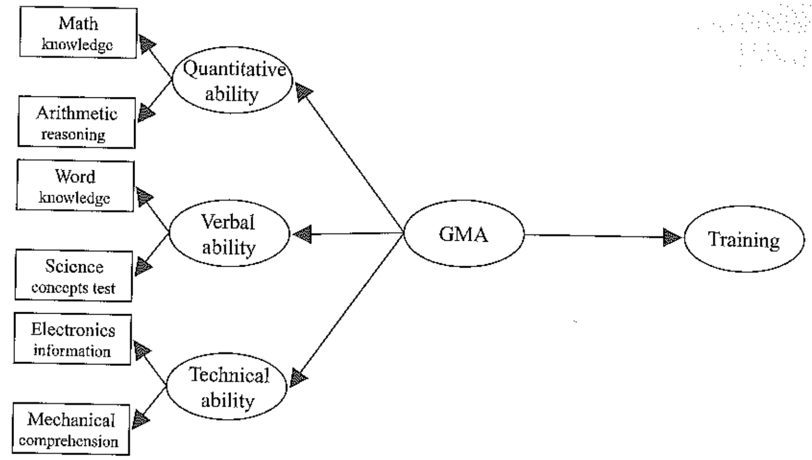


Figure 6.8 *Training performance is predicted by GMA rather than specific abilities (adapted from Hunter, 1983)*

by any specific abilities, i.e., there are no paths from quantitative, verbal or technical aptitudes to training performance. These results were also replicated by later investigations by Ree and colleagues, who reported minimal incremental validity for specific abilities over and above GMA (Ree & Earles, 1991a, 1991b, 1992; Ree, Earles & Teachout, 1994).

With regard to non-academic counterclaims or reasons why laypeople question the validity of IQ tests, perhaps the biggest reason is that laypeople (including the average employee) underestimate the real impact of individual differences underlying performance, masked by everyday behaviours at work. Thus people are quick to judge fellow workers on the basis of how sociable, well or badly behaved, they are at work. The fact of the matter is, though, that social behaviours have little to do with individuals' performance or actual productivity. As noted by Hunter and Schmidt:

> Many people who do not study work scientifically fail to distinguish between poor citizenship behaviour and poor performance. These are not highly correlated. There are people who are poor citizens but who can perform the work well. These poor citizens who can perform well when they perform are the workers that laypeople think of when they say 'Intelligence just doesn't matter that much at work'. However, there are many workers who are excellent citizens but who cannot do the job very well. These poor performers are usually tolerated by supervisors because they obviously try hard. These individuals also tend to be ignored by lay theories. (1996, p. 449)

Thus laypeople have a natural tendency to judge their co-workers in terms of their social rather than performance-related behaviour. Even supervisory ratings of employees' performance are 'contaminated' by social behaviours (Orr, Sackett & Mercer, 1996), despite the fact that such behaviours bear little relationship to productivity. This is true not only for desirable but also undesirable social behaviours, such that unreliable, disorganised and argumentative employees are likely to be punished by both co-workers and supervisors even though they may be more productive than their reliable, organised and agreeable counterparts, though it has been recently shown that GMA is also associated with a number of counterproductive behaviours at work (see Section 6.11).

6.9 The inconvenient effects of the 'triple *g*': general, genetic and group differences

It is only when we jointly examine some of the causes and consequences of GMA that we can fully understand the reasons for the low popularity and political delicacy of the notion of IQ. Although a book on the psychology of personnel selection, one may argue, should not devote much time to discussing these issues, it is quite clear that people's conceptions – right or wrong – of GMA and IQ tests threaten the use of these tests in personnel selection. In fact, given the direct economic implications of selecting or not on the basis of GMA, we feel compelled to discuss not only the major controversies surrounding IQ but also some of the claims associated with the idea of using or banning such tests altogether.

As noted above, the first hard-to-digest issue is that GMA is very general and has equally general effects in the workplace, such that it predicts virtually every relevant occupational outcome, and in most cases better than specific abilities do. Thus GMA refers to very *general* differences in aptitudes that are reflected in multiple indicators of job performance in a generic fashion.

The second inconvenient issue is even harder to digest, namely, the well-established fact that there are strong genetic influences on GMA. Correlations between genetic similarity and IQ scores approach r = .90 in mid to late adulthood, leaving little room for environmental influences (Bouchard, 1998; Plomin, 2001). Furthermore, most of the environmental effects on GMA are *non-shared* – in broad terms, non-shared environment is what exposes children in the same family to different environmental influences (e.g., differential treatment from parents, school mates and teachers) – and probably 'respond' to genetic influences (Chamorro-Premuzic, 2007, pp. 90–101). For example, if one of two siblings is musically talented, his/her teachers may insist that he/she takes up extra piano lessons, which, in turn, will enhance the differences in musical competence with the other sibling. Although even the most radical geneticists would acknowledge, and often emphasise, that the genetic basis of GMA is far from behavioural determinism, it is hard (and not just for the layman) to see how this could be the case. Occupational success – or failure – is obviously not encoded in the genes, but the psychological traits that facilitate or hinder it most certainly are. It is estimated that about 20–25 per cent of the observed variability in occupational level and income is due to genetic differences in GMA (Gottfredson, 2004), and there are even fewer doubts about the stability of GMA across the lifespan. Studies have

shown that, even over a period of sixty-five years, there is little variability in people's IQ scores (Deary, Whalley, Lemmon, Crawford & Starr, 2000), unless crystallised abilities, which increase until the age of 60 (Ackerman, 2008), are considered.

Although the stability of GMA has clear benefits for personnel selection (it would indeed be quite problematic to select someone on the basis of their IQ scores if their GMA levels changed drastically over time), it poses an important political question as to the limited lifelong job prospects of lower-GMA individuals even at an early age, particularly in industrialised nations. If, as we have seen is the case, IQ scores at age 12 are such strong predictors of occupational level and income later (Judge *et al.*, 1999) (see also Section 6.7), and GMA varies little over the lifespan, are lower-GMA individuals already 'condemned' or handicapped by the time they reach adolescence?

As if these two issues were not enough, there is the issue of *group differences* in GMA, which, in the context of personnel selection, tends to fall under the broader discussion of *adverse impact*. Thus '[C]ognitive ability tests represent the best single predictor of job performance, but also represent the predictor most likely to have substantial adverse impact on employment opportunities for members of several racial and ethnic minority groups' (Murphy, 2002, p. 173). For example, several robust studies, particularly in the US, report that whites tend to score higher than Hispanics, who tend to score higher than blacks on IQ tests. Estimates of white–black differences in IQ tend to give whites an average advantage of .85 to 1.00 standard deviation (that is, almost 15 IQ points), which is certainly 'not trivial' (Hough & Oswald, 2000, p. 636). Although group differences in job performance are somewhat less pronounced (Hattrup, Rock & Scalia, 1997; Waldman & Avolio, 1991), the mainstream view in intelligence research is that these differences are not caused by any test biases (Gottfredson, 1986, 1988, 2005; Jensen, 1980, 1998).

When the implications of the three inconvenient *g-facts*, or 'triple *g*' facts (general, genetic and group differences), are combined, they produce a chain of events that has clear social implications. First, it is evident that in globalised times industrialised societies are characterised by increasingly complex and less stable jobs that require more and more GMA. Newer jobs are more '*g*-loaded' because they require rapid adaptation and highly effective performers (think, for instance, of the downsizing levels caused by mergers and acquisitions) (Sackett & Lievens, 2008). Second, these jobs also tend to pay more, and have higher occupational status attached, than low-complexity jobs. As a consequence of these two points, it is harder and harder for lower-IQ individuals to be in the higher-earning spectrum of society and to aspire to higher-status jobs. In fact, individuals with lower GMA are also at greater risk of being unemployed as technological advances make it possible to replace repetitive (and unintellectual) manual tasks with machine operations.

Despite this divisive picture of intellectual potential and job opportunities, there is little evidence for the benefits of ignoring GMA when it comes to selecting employees. In fact, most studies report just the opposite, namely detrimental effects of banning IQ-based personnel selection (Hunter & Schmidt, 1996), and its macro-economical consequences have been discussed elsewhere (Hunter & Schmidt, 1996). Given that higher-complexity jobs require higher GMA, and that even in lower-complexity jobs brighter employees are more productive – meaning the higher the GMA of the workforce the smaller the workforce can be – it is easy to imagine what these effects may be. On the other hand, GMA-based selection is not necessarily a disadvantage for any group of society, as individuals would be rated on the basis of their own capability rather than their group membership. Furthermore, social policy can easily address the issue of group inequalities in GMA by providing additional support for individuals with lower GMA, and this can be done at the *individual* rather than group level. Thus, just as people are not selected for their membership of any age, gender or ethnic group, their learning would not be supported on the basis of group membership but individual competence. For example, people with an IQ < 80 (about 10 per cent of whites and 30 per cent of blacks in the US) are currently considered unsuitable for the US army by federal law and there are few civilian employers who would hire under this GMA threshold (Gottfredson, 2005). If these people were equipped with additional training that enabled them to develop extra skills and acquire knowledge in key areas, the performance or training potential gap that separates them from their higher-IQ counterparts (e.g., white *or* black, male *or* female, young *or* old) would probably decrease. Thus social policy can address these issues without causing IQ-based stagnation. It may also be possible to redesign specific jobs in a way that removes unnecessary complexity from them, such that lower-IQ individuals can attain the same level of performance (in training and at work) without scarifying productivity levels, though tailoring the instructions of ability tests to make them more user-friendly for disadvantaged groups does not seem to eliminate group differences (DeShon, Smith, Chan & Schmitt, 1998).

Recent reviews also suggest that GMA differences between blacks and whites may be smaller than previously thought (Sackett & Lievens, 2008), and certainly less marked for specific cognitive abilities. For instance, short-term memory ability (measured via digit span or symbol substitution) yields average differences of less than $\frac{1}{2}$ SD despite having high validities – correlating with job performance at r = .45 (Verive & McDaniel, 1996). In addition, it has been noted that selecting on the basis of specific cognitive abilities may also be beneficial for *higher* GMA candidates as specific abilities are more weakly intercorrelated at higher levels of GMA (Lubinski & Benbow, 2000).

More importantly, it is important to emphasise that the effects of GMA are especially consequential ceteris paribus (i.e., all other things being equal) (Ackerman, 2008; Murphy, 2002), but this is rarely the case. In fact, even if group differences in GMA are negligible, applicants' own assessments of what they can and indeed are likely to produce may push them toward, or away from, potential jobs (Wilk & Sackett, 1996), and it is questionable whether taking up jobs that are seen as too demanding has more beneficial than detrimental consequences. In that sense,

self-selection occurs regardless of personnel selection, and usually precedes it. This makes recent findings on individual differences in self-assessed intelligence very relevant in regards to career choices and personnel selection (Ackerman & Wolman, 2007; Chamorro-Premuzic, Harlaar & Plomin, 2008; Chamorro-Premuzic & Furnham, 2006b; Furnham & Chamorro-Premuzic, 2007).

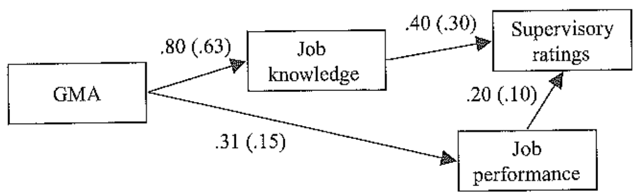
6.10 Searching for mediators: GMA predicts job performance and training, but why?

> For practical purposes of prediction in personnel selection, it does not matter why GMA predicts job performance. However, scientific understanding requires theoretical explanation. Theoretical explanation is also required to gain acceptance of findings from those who question the plausibility of a central role for GMA in the determination of job performance. It is easier to accept an empirical finding when there is a theoretical explanation for that finding. (Schmidt & Hunter, 2004, p. 170)

The main reason why GMA predicts job performance, and related outcomes, is that it causes faster, better, more effective and enduring knowledge acquisition. This is evident in the strong associations between GMA and training success reviewed in Sections 6.4, 6.5 and 6.6. In simple terms, having a higher IQ means being able to learn faster. One of the first psychological figures to formalise this was Edward Lee Thorndike, whose *classic theory* of job performance posited that individual differences in learning are the main driving force of individual differences in job performance (Brolyer, Thorndike & Woodyard, 1927). In occupational settings, formal learning – through training and instruction – as well as learning on the job, is influenced by GMA. As noted, 'because the rate and amount of learning is determined by cognitive ability, the classic theory predicts a high correlation between cognitive ability and learning' (Hunter & Schmidt, 1996, p. 461).

Hunter (1983a) set out to empirically validate this hypothesis by examining fourteen studies that included measures of GMA, job knowledge (a proxy of learning) and job performance as assessed both objectively and subjectively through supervisory ratings. Figure 6.9, adapted from Hunter and Schmidt (1996), summarises the path models for both military and civil samples. As shown, most of the effects of GMA on job performance are explained by job knowledge. Thus, GMA was positively correlated with supervisory ratings because of its positive effects on job knowledge, which means that, once individual differences in job knowledge are taken into account, differences in GMA are not associated with differences in supervisory ratings, though they still translate into higher job performance. It is also clear that job knowledge has a greater impact on supervisory ratings of performance than on actual job performance (Ree, Carretta & Teachout, 1995).

Although the job knowledge mediation suggests that the effects of GMA on job performance, at least as rated by supervisors, could be attenuated by offering more



Note: Numbers in brackets show coefficients for military jobs; numbers outside brackets show coefficients for civilian jobs.

Figure 6.9 *Job knowledge mediates the effects of GMA on job performance and ratings (adapted from Hunter & Schmidt, 1996, based on fourteen studies analysed by Hunter, 1983a)*

training to lower-IQ individuals (a clear example of how the above-discussed psychological data could be used to *minimise* rather than maximise the effects of GMA on performance), this is probably not in the current agenda of many organisations, and not just in the western, industrialised world. Clearly, it is cheaper to hire fast learners or people who know their stuff already than to train slow learners or people who have yet to learn their stuff. Nonetheless, if the organisation's priority is to boost group parity, equity, non-discrimination and social diversity, a trade-off between validity and integration can surely be made. As recently noted:

> Reliance on cognitive ability measures in selection is likely to lead to more efficiency (i.e., higher average performance) and less equity (e.g., disparities in selection rates across racial and ethnic groups), whereas avoiding the use of cognitive ability measures is likely to lead to less efficiency and more equity. There is no scientific principle that can tell us which criteria *should* receive more evidence, but scientific research can clearly inform us of the consequences of emphasising one criterion or the other. (Murphy, 2002, p. 183, italics in the original)

Clearly, criteria are influenced by market demands and in the past decades technological advances in the workplace have certainly increased the importance of individuals' learning potential in comparison to the knowledge and experience they may have already gained.

It appears that even hiring on the basis of previous experience does not diminish the impact of GMA on subsequent training and job performance. In Hunter's (1980) seminal meta-analysis, mean job experience was over five years and GMA still correlated with job performance at r = .51 (see also Hunter, 1986). Indeed, Schmidt and colleagues (Schmidt, Hunter, Outerbridge & Goff, 1988) found that ability differences in job knowledge where as marked after five years as after a single year. On the other hand, research evidence shows that the experience–performance correlation decreases with time in the job (from r = .49 during the first three years to merely r = .15 after twelve years in the job) (McDaniel, Schmidt & Hunter, 1988), whereas the opposite occurs with the

GMA-performance correlation (from r = .35 during the first three years to .r = .59 after 12 years of age) (McDaniel, 1985). That said, knowledge acquisition is not only determined by GMA.

Much of the theoretical and empirical work of Phillip Ackerman and Ruth Kanfer focuses on the 'non-ability' determinants of skills acquisition and adult knowledge. The underlying threat to this programme of research is the idea that GMA is mainly helpful at initial stages of skills acquisition and particularly in so-called 'inconsistent-tasks' (which require ongoing intellectual application). For 'consistent-tasks', however (e.g., typing, driving and answering phone-calls), *typical* levels of performance – what a person is motivated to do and normally does – are more important than *maximal* levels of performance – what a person *can* do (Ackerman, 1987; Cronbach, 1960; Sackett, Zedeck & Fogli, 1988). Performance on short tasks tends to be determined mostly by GMA, whereas performance on longer-term, everyday tasks is strongly influenced by an array of motivational and personality traits, which tend to fluctuate less in short-term performance assignments (Sackett *et al.*, 1988). As discussed in Chapter 7 (see also Hough, Oswald & Ployhart, 2001), these non-ability traits have been less frequent targets of political debate.

Thus, relying on non-ability traits and using multiple methods of assessment should reduce adverse impact without necessarily sacrificing predictive power, though the idea of not resigning predictive power if GMA measures are excluded is close to inconceivable (Gottfredson, 1996; Murphy, 2002).

Another reason why IQ tests predict job performance is that higher GMA is linked to higher *job role breadth*, enabling brighter employees to perform a wider range of tasks and, in turn, be rated more highly by their supervisors (Morgeson, Delaney-Klinger & Hemingway, 2005). These findings are consistent with the idea that higher GMA is associated with higher levels of intellectual curiosity, Openness to Experience and need for cognition (Chamorro-Premuzic & Furnham, 2004, 2005, 2006b), and the importance of problem-solving even beyond tested cognitive ability. For example, recent studies on Openness show that it is a critical determinant of cross-cultural training and coping with organisational change (George & Zhou, 2001), whereas studies on assessment centres show that problem-solving is the most important determinant of candidate's performance (Arthur, Day, McNelly & Edens, 2003).

6.11 GMA and counterproductive work behaviours

Although most applied psychologists employ cognitive ability tests to predict positive outcomes, such as learning, training and job performance (Ones, Viswesvaran & Dilchert, 2005), the importance of GMA at work is not limited to the wide range of desirable outcomes it predicts. Indeed, GMA has been negatively associated with workplace deviance, also referred to as *counterproductive work behaviours* (e.g., employee theft, high absenteeism levels and misuse of resources). These behaviours are commonly encompassed under the heading of workplace deviance, which has been defined in terms of 'voluntary behaviour that violates significant organisational norms and in so doing threatens the well-being of an organisation, its members, or both' (Robinson & Bennett, 1995, p. 556). It has been estimated that these behaviours could cost the US economy as much as $200 billion a year (Greenberg, 1997; Murphy, 1993; Vardi & Weitz, 2004; see also Bennett & Robinson, 2000).

After many decades of little research in this area, a recent US investigation of 1,799 police officers (of similar SES) (Dilchert, Ones, Davis & Rostow, 2007) reported substantial negative associations between IQ – measured during job applications – and several counterproductive behaviours at work – taken from the archive after several months of employment. Examples of counterproductive behaviours assessed in this study were racially offensive conduct, excessive use of force, misused official vehicles and destroying official property. In line with previous theoretical propositions (White, Moffitt & Silva, 1989) the authors interpreted the negative correlation between IQ and workplace deviance in terms of the protective or *inhibitory effects* of higher GMA, which enables brighter individuals to better predict or foresee the negative consequences of deviant behaviours, therefore inhibiting maladaptive behaviours. Particularly lower-GMA individuals would be characterised by 'a deficit in foreseeing temporally remote consequences of actions' (Lubinski, 2000, p. 430). As noted, however, this is a rather new area of research and replication of the negative link between GMA and deviant behaviours at work, particularly from longitudinal studies, is deemed necessary before implementing GMA measures solely for this purpose. Until then, it is more appropriate to associate counterproductive behaviours, and behavioural problems in general, with personality traits rather than cognitive ability, especially the personality trait of Conscientiousness and its primary facet of dependability (Dudley, Orvis, Lebiecki & Cortina, 2006) (see also Section 7.6). Special attention should be devoted to these non-ability traits as they are generally uncorrelated (Bobko *et al.*, 1999; Sackett & Lievens, 2008) – and often even negatively correlated (Ackerman & Heggestad, 1997; Chamorro-Premuzic & Furnham, 2006b) – with cognitive abilities.

6.12 Retesting, practice and coaching effects

Despite their impressive validity as predictors of job outcomes, cognitive ability tests are not perfect measures of GMA because individuals' performance on ability tests is also affected by a number of non-ability factors (e.g., negatively by anxiety and worry, positively by motivation and confidence) (Chamorro-Premuzic & Furnham, 2004). In occupational settings, individuals may wish to retest in order to maximise their chances of employment or promotion, and indeed an estimated 25–50 per cent of people have been reported to retake ability tests in the US (Hausknecht, Halpert, Di Paolo & Di Paulo, 2007). The question therefore arises of whether a one-time measure of GMA can provide the most accurate

Table 6.7 *Explanations for practice effects (test score gains) on IQ*

Source	Explanation
Anastasi, 1981	Development of or improvement in actual abilities (require substantial intellectual investment from the individual)
Messick & Jungeblut, 1981	Decrease in test-anxiety due to previous test experience
Kulik, Kulik & Bangert, 1984	Memory of questions and answers in previous test
Sackett, Burris & Ryan, 1989	Improvement in test-taking strategies
Campbell & Kenny, 1999	Regression to the mean (i.e., the statistical sampling artefact that results from the fact that only people who scored relatively low in the initial test would be expected to request a second testing session, whilst those who did well in the first instance would not return for a second testing session)
Hausknecht *et al.*, 2007	Mere repetition (e.g., enhanced understanding of instructions or items without specific training); may occur even after completing other ability tests; another explanation is formal instruction (coaching) between sessions

indicator of an individual's potential to succeed in the workplace or elsewhere. This led researchers to examine the issue of retesting the same individuals on the same test or slightly modified versions of it, looking at practice effects (i.e., score gains from one test administration to the next) and coaching effects (i.e., assisted practice effects, that is, practice with formal instructions).

On the one hand, the reliability of IQ tests renders substantial intra-individual differences in IQ test performance unlikely, and, indeed, testing agents often invalidate second-attempted scores if they are significantly larger than the original score (Harcourt Assessment, 2005), whereas the US Air Force bans retesting in some of the components of its selection battery for pilots. Yet IQ and other ability tests measure maximal performance at a given time, and a single test administration may not capture the person's GMA accurately if underperformance occurred. Consequently (but in some conflict with the previous point), it has been recommended that 'employers should provide opportunities for reassessment and reconsidering candidates whenever technically and administratively feasible' (SIOP, 2003, p. 57) and that 'every test taker should have a fair chance to demonstrate his or her best performance on an assessment procedure', such that employers should 'consider retesting or using alternative assessment procedures before screening the individual' (US Department of Labor, 1999, Section 4, pp. 8–2).

Although several explanations for practice or unassisted training effects have been proposed (see Table 6.7), the fact is that practice *does* increase ability test performance, albeit marginally. Messick and Jungeblut's (1981) seminal study on the SAT concluded that score increases of approximately 3 per cent would require as much coaching or assisted practice time as full-time school education, and that the effects of coaching tend to decrease after initial coaching time. More recently, a meta-analysis of 107 samples (reported in fifty studies) revealed score gains of about 1/4 of a standard deviation between the first two test sessions and increases of over 1/2 of a standard deviation between the first and the third testing session (Hausknecht *et al.*, 2007). These results imply that candidates scoring in the 50th percentile when initially tested could be expected to score over the 70th percentile on their third attempt, which suggests larger practice effects than those reported by Messick and Jungeblut (1981), though the meta-analysis did confirmed that the effects of coaching diminish over time. In addition, the meta-analysis also highlighted the importance of having at least twelve months' interval between test administrations in order to reduce the risk of score gains due to memory of items.

6.13 Summary and conclusions

The present chapter has examined the validity of GMA – and IQ tests – in the prediction of major job outcomes, such as occupational level, job knowledge and job performance. Regardless of the criterion we choose, it is clear that GMA is a fundamental and very powerful predictor of positive job outcomes. Evidence also indicates that the predictive power of GMA is rivalled by no other psychological construct (see, for instance, Chapters 3 on biodata and 7 on personality to understand how much more valid cognitive ability tests are).

Although GMA is a strong predictor of overall job performance (correlating at about $r = .50$, and thus explaining 25 per cent of the variance in job performance), it matters most in complex or intellectually demanding jobs (where it correlates at about $r = .80$ with job performance) and least in unintellectual or cognitively simple jobs (where it correlates with job performance at about $r = .20$). Objective measures of performance correlate more highly with GMA measures than subjective assessments of performance, such as supervisory ratings, do. This is because the latter – but not the former – are 'intoxicated' by a wide range of job-irrelevant variables, such as sociability, friendliness, looks and compatibility between the worker and the supervisor's personality traits. This is consistent with the finding that personality traits predict supervisory ratings better than objective performance tests (see Section 7.4).

Specific abilities, that is, variance in cognitive abilities unaccounted for by the general GMA factor, are insignificant predictors of job performance and related outcomes once GMA is taken into account. This is counterintuitive to most people because the layperson tends to overestimate the importance of situational and job-specific factors when interpreting the determinants of work performance. Likewise, most people tend to see their abilities as being heterogeneous, though in fact within-individual differences in abilities are marginal compared to interindividual differences.

Why does GMA predict job performance and related outcomes? Primarily because it accounts for individual differences in learning, specifically how quickly and in what depth people can learn a task, and acquire and retrieve information, both through formal training and informally through experience. Thus job knowledge has been found to partly mediate the effects of GMA on job performance, suggesting that the main – but not only – reason why higher-IQ individuals do better is because they are more knowledgeable in the job. The fact that not all the effects of GMA on job performance can be explained by GMA-related knowledge acquisition has been interpreted in terms of the beneficial effects of GMA on those aspects of jobs that are 'less routinised or less closely supervised; more fraught with change, ambiguity, unpredictability, and novelty (and hence are inherently less trainable); or otherwise require greater exercise of independent judgement and innovative adaptation' (Gottfredson, 2004, p. 175).

Despite the overwhelming body of evidence in support of the predictive validity of GMA in the workplace, the use of cognitive ability tests in personnel selection remains somewhat controversial because of the widely replicated *adverse impact* that GMA-based selection has on specific groups. Clearly, a key aim of civilised and industrialised economies is to provide work opportunities for its wide workforce but the fact that top jobs (which tend to be cognitively complex and require fast and efficient learning) seem out-of-reach for lower-GMA individuals, and that these people are found more frequently in some ethnic or gender groups than in others, creates a conflict between economic integration and productivity. That said, GMA is neither morally *wrong* or *right* but simply a useful – and accurate – indicator of people's ability to learn new things, solve complex problems and adapt to the environment. Furthermore, people's differences in GMA are mostly independent of any group differences as within each group individuals differ in cognitive abilities. More importantly, work performance is not only determined by people's cognitive abilities (even in top jobs), which makes it necessary to look at other constructs in other to explain, understand and ultimately predict occupational success.

Notes

1 For many years, however, there has been a tradition in occupational and industrial/organisational psychology to use 'intelligence' to refer to general mental ability after it reached a point of maturity (Hunter & Schmidt, 1996).
2 The fourth edition of the WAIS is in preparation.
3 It has been pointed out (Schmidt, Ones & Hunter, 1992) that some of the most important evidence for the validity of GMA with regard to job performance and similar outcomes is not available in the form of books and journal articles, but only reported in the form of technical reports or manuals, which are often more difficult to access. Anastasi presented a core summary of the main trends in GMA testing until the late 1980s (Anastasi, 1989), and a comprehensive review of the validity studies on GMA conducted on the Armed Services Vocational Aptitude Battery (ASVAB) has also been published (Welsh, Watson & Ree, 1990).
4 Its modern version is O'NET, a computerised delivery system that links jobs to specific attributes; see Hough & Oswald (2000).
5 Corrected only for unreliability in criterion.

7 Personality traits

7.1 Introduction

The previous chapter dealt with the validity of GMA (general mental ability) as a predictor of job and training performance, as well as other work-related outcomes. In the current chapter, we discuss the predictive power of personality traits.

The question then emerges as to what is the difference between GMA and personality traits, and this is a question for which only one simple answer exists: traditionally (in personnel selection as well as in the wider context of psychological assessment), GMA is measured or *tested* via objective performance tests (such as those discussed in Section 6.6 and 6.7), whereas personality traits are *assessed* via subjective inventories, notably self- or other-reports (but especially self-reports). In that sense, one can distinguish between cognitive abilities and personality traits on the basis of assessment methods, whereby the former reflect individual differences in the capacity to identify correct responses to a standardised test (verbal or non-verbal), whereas the latter reflect individual differences in general behavioural tendencies, assessed only subjectively, that is, through people's accounts (one's own or others'). This led to a now well-established distinction in psychology to refer to cognitive abilities in terms of *maximal performance* and personality traits in terms of *typical performance* (Cronbach & Gleser, 1965), though in the case of personality traits 'behaviour' is a more accurate term than 'performance'.

As much as this distinction is straightforward, things get more complex when we try to assess whether measures of ability and inventories of personality may or not be tapping into the same underlying constructs. That is, is the distinction between personality and abilities only relevant at the level of measured constructs, and therefore a purely methodological issue, or are personality inventories and ability tests also assessing similar constructs? Let us chose a simple example to illustrate this question. Let us say that we want to measure individual differences in running, specifically how fast people can run. One option would be to test a number of subjects in a 100 metre race and time how fast they run (using a stopwatch). Another option would be to *ask* people how fast they can run, and we would be willing to bet that at least 90 per cent of our readers would find the first option better than the second. Why? There are two main reasons. First, people may be unaware of how fast they can run. Second, even if they were aware, they

may be unwilling to tell us (especially if we were giving out medals or cash to the fastest runners). Thus, they may chose to exaggerate how fast they can run and try to deliberately mislead us. Of course, one may ask other people (their friends or indeed their enemies) how fast our candidates can run, but the same problems apply, i.e., the friends may be as unwilling or unable to provide this information as the runners themselves, and enemies can hardly be expected to be more accurate.

Our readers may have guessed that these two problems also apply to personality inventories, as personality traits are only assessed subjectively. Indeed, this has been the most common objection to using personality scales in personnel selection and, accordingly, a large part of this chapter is devoted to this issue. But let us make clear at the outset that there are some advantages in using subjective[1] reports rather than objective performance tests. In fact, these advantages highlight some of the limitations of objective assessment methods (briefly discussed in Section 6.13) and explain the importance of using personality inventories in personnel selection. The first issue is that subjective reports can take into account aggregate data, that is, how people have been performing or how they perform most of the time (this has already been shown in Chapter 3 on biodata). Thus, in the context of our running example, asking the runners or people who are familiar with them how fast they can run may provide information on how fast these people *usually* run, as well as how fast they have managed to run in the past. If we were only interested in how fast people *can* run, this may or not be a good approach. However, it is clear that if we were interested in how fast they *tend to run*, then objective performance tests would be very poor indicators of this: even Olympic-medal winners don't run as fast as they can most of the time.

There is also a second issue, which is that objective performance tests (i.e., timing people once, or testing their ability once) may not be accurate, especially if factors other than running ability interfere. Examples of such factors can range from fatigue or test-anxiety (including fear of evaluation, which may explain why even professional athletes may record faster times in training than in the actual competition) to a heavy hangover. This phenomenon, simply referred to as *underperformance* (understandably, 'over-performance' is rarely an issue), threatens the validity of objective tests but does not harm subjective reports. In fact, there are two main reasons underlying the fact that psychometric tests do not perfectly predict any outcomes: the first is that they may be failing to measure the construct in a completely accurate manner; the second is that, even if they do, that construct may be only one of the determinants of the outcome we wish to predict. In a sense, these two reasons are the same and merely one: tests that capture not just running ability but also fatigue, test-anxiety and hangover (to stay with the above example) should therefore predict running performance better than tests that capture only one of the predictors. This logic has been applied to personality inventories and ability measures (Chamorro-Premuzic & Furnham, 2006a; Wechsler, 1939). As shown in Figure 7.1, measures of academic performance, such as general point average (GPA), can be used to validate both