The predictive value of IQ

Article *in* Merrill-Palmer Quarterly · February 2001

DOI: 10.1353/mpq.2001.0005

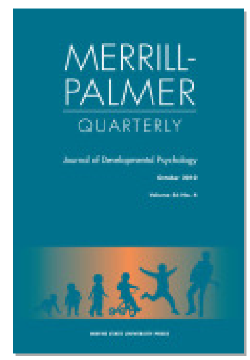
4 authors, including:

Robert Sternberg
Oklahoma State University - Stillwater

Elena Grigorenko
Yale University

The Predictive Value of IQ

Sternberg, Robert J.
Grigorenko, Elena.
Bundy, Donald A.



Merrill-Palmer Quarterly, Volume 47, Number 1, January 2001,
pp. 1-41 (Article)

Published by Wayne State University Press
DOI: 10.1353/mpq.2001.0005

Merrill-Palmer Quarterly, Vol. 47, No. 1

The Predictive Value of IQ

Robert J. Sternberg
Yale University

Elena L. Grigorenko
Yale University and Moscow State University

Donald A. Bundy
Oxford University and World Bank

This article reviews findings on the predictive validity of psychometric tests of intelligence. The article is divided into five major parts. In the first part, the issues with which the article deals are introduced. In the second part, we discuss what psychologists can learn about the predictive validity of intelligence tests from results obtained in the established market economies. Intelligence quotient (IQ) is considered in relation to educational achievement, employment prospects and wealth generation, career outcomes, and well-being. In the third part, the intelligence tests (primarily for infants and children) that yield the IQ scores are discussed. In the fourth part, constraints are presented on the interpretations of findings, including cross-cultural issues. We conclude that conventional tests of intelligence can be useful but only if they are interpreted very carefully, taking into account the factors that can affect them, and in conjunction with other measures.

In Kenya, those schoolchildren whose traditional skills are most prized by the community tend to do least well in school tests (Sternberg & Grigorenko, 1997; Sternberg, Nokes, et al., in press). In Brazil, street children who run a successful street business typically fail mathematics in the school setting (Ceci & Roazzi, 1994). In the West, school-based tests

Robert J. Sternberg, Department of Psychology, Yale University; Elena L. Grigorenko, Department of Psychology, Yale University and Moscow State University; Donald A. Bundy, Oxford University, Oxford, UK and World Bank.

Preparation of this article was supported by a grant from the World Bank. Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520–8205.

Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520–8205. Electronic mail may be sent to robert.sternberg@yale.edu.

Merrill-Palmer Quarterly, January 2001, Vol. 47, No. 1, pp. 1–41.

show correlations with career success, but they are also major gatekeepers of academic and vocational routes to advancement (Sternberg, 1997).

In this review we examine conventional and current approaches to the measurement of intelligence, paying particular attention to the extent to which these approaches are predictive of the later success in life. One of the main tasks psychology as a science sets for itself is to find ways to predict the future. Many psychologists have believed the general index of cognitive ability (intelligence quotient, or IQ) to be the best single predictor of virtually all criteria considered necessary for success in life in the Western part of the developed world (e.g., Herrnstein & Murray, 1994; Hunter & Hunter, 1984; Jensen, 1998; Schmidt, Ones, & Hunter, 1992).

Yet almost by definition, IQ is a culturally, socially, and ideologically rooted concept. It could scarcely be otherwise, as this index is intended to predict success (i.e., to predict outcomes that are valued as success by most people) in a given society (i.e., in a large social group carrying its own set of values). IQ has been most studied where it was invented and where it is most appreciated, that is, in the established market economies and especially in the United States. Oddly enough, the country where its testing originated—France—largely ignores it.

The use of a general index of cognitive ability raises technical issues that have attracted the attention of developmental researchers for many years. These issues are (a) whether IQ is an important developmental construct that is predictive of significant life outcomes; (b) whether IQ is changeable and whether changes in IQ are meaningful; (c) whether these changes are due primarily to error or are systematic; (d) the degree, if any, to which there is continuity or discontinuity in IQ during different developmental stages; and (e) whether other individual-difference variables are predictive of those life-quality indicators that are traditionally linked to IQ.

In approaching this review we have had to recognize that almost all the data available to us originate with studies in the established market economies of North America, Europe, and Australia. The major analyses presented here, therefore, should be interpreted in that context. These analyses form the first part of this review; they examine what IQ tests predict, and how the outcome of tests is modifiable by external factors. To adumbrate points made in more detail later: Childhood IQ is generally a fairly good predictor of many criteria, at least in high-income, industrialized countries. Infant IQ is not such a good predictor.

To attempt to place these findings in a context that is relevant to the majority of the world, and to low-income countries in particular, we review the somewhat sporadic evidence elsewhere. These studies suggest some general conclusions. First, views on smartness vary in different cultures; the majority of these views do not match Western views (Berry &

Bennett, 1992; Greenfield, 1997; Okagaki & Sternberg, 1991; Serpell, 1993; Yang & Sternberg, 1997). Second, instruments developed to quantify smartness are culturally based and cannot simply be "transplanted" to a culture with different values (Greenfield, 1997). The situation of testing itself (e.g., communicating with strangers regarding things and issues that lack context and that might appear to be meaningless) often results in the collection of unreliable data (e.g., Glick, 1968). Finally, psychological research in settings where there is no established psychological tradition should be preceded by ethnographic investigation involving firsthand experience in the settings in which the human activity of research interest occurs (Colby, Jessor, & Schweder, 1996).

These findings do not imply that tests of general cognitive ability have no place in low-income economies. Indeed, they confirm that IQ tests are predictive of some important outcome criteria. But such tests are not sufficient in themselves and, if used injudiciously and out of context, may be dangerously misleading. Factors that may be an important part of these measurements—in addition to the fact of so-called general cognitive ability—include adaptation to the social environment, skills for coping with novelty, self-efficacy, and persistence in the face of frustration. Measures of these abilities need to be developed with due regard for the local context.

In sum, despite the magnitude of the predictive power of IQ apparent from the findings presented later, this index might extend itself meaningfully only throughout its own kingdom—that is, only through selected segments of the Western part of the industrialized world. Where it extends itself best is in the area of educational achievement in Western-style schooling.

The remainder of this article is divided into five major parts. In the next and second part of the article, we discuss what psychologists can learn about the predictive validity of intelligence tests from results obtained in the established market economies. We consider the relation of IQ to educational achievement, employment prospects and wealth generation, career outcomes, and well-being. In the third part, consideration is given to how these IQ scores are obtained. Here, we discuss the intelligence tests that actually yield the IQ scores and review tests for infants as well as tests for children in some detail. In the fourth part, the issue is addressed of the necessity of evaluating IQ in its contexts, and thus, certain constraints on the interpretation of past findings. In this part cross-cultural issues are considered as well as general issues of evaluation. Finally, we conclude that conventional tests of intelligence can be useful, but only if they are interpreted very carefully, taking into account the factors that can affect them, and in conjunction with other measures.

WHAT CAN PSYCHOLOGISTS LEARN FROM EXPERIENCES IN THE ESTABLISHED MARKET ECONOMIES?

IQ and Educational Achievement

Vygotsky (1978) was one of the first psychologists to study systematically links between cognitive development and education. As a result of his influential work and work by other scientists, many developmentalists have adopted a view of reciprocal causality between cognitive development and education: Higher cognitive indices are predictive of more educational achievements and more education is predictive of higher intellectual outcomes (e.g., Brody, 1997; Ceci & Williams, 1997).

Tests of intelligence were originally devised specifically to predict educational achievement. In fact, they do a good job of prediction: The correlations between IQ scores and both school grades and achievement test scores average about .40 to .50 but have somewhat different ranges for different samples, different tests, and different areas of achievement. The correlations tend to be higher for the diverse groups serving as test-standardization samples. For example, in the Woodcock-Johnson-Revised standardization sample, the correlations ranged from .53 to .93, with a median correlation of .76 (McGrew & Knopik, 1993). But the correlations may be lower for specific populations. In a sample of 127 students enrolled in a private day school located in a large metropolitan area, the correlations ranged from .11 to .22, with the median of .18 (Novak, Tsushima, & Tsushima, 1991).

Correlations between IQ and achievement tests tend to increase with age (McGrew & Knopik, 1993). This increase may be due to the greater overlap in content between the two kinds of tests at higher levels, which in turn may reflect the greater overlap in the skills measured by the two kinds of tests at higher age levels. In a comprehensive study of 26,300 boys and girls from eight different ethnic backgrounds who were referred to and evaluated for the gifted program in the San Diego City schools, correlations between the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1991) and the Standard Raven Progressive Matrices (Raven, 1960, 1965), on the one hand, and the California Achievement Test, on the other hand, ranged between .32 and −.05, with a median of .17 (Saccuzzo & Johnson, 1995). The data on 29 studies[1] surveyed for this article suggest that, on average, 10% to 22% of the variance in subject-specific achievement scores is overlapping with the variance in IQs. The

[1] A complete list of references to the studies cited in this article is available from the authors upon request.

proportion of the explained variance is slightly higher (36%) when a composite measure of school achievement is analyzed. Finally, the data from 11 studies linking school grades and IQ(s) (see Sternberg, Grigorenko, & Bundy, 1997) suggest that a maximum of 18% of the variance in school grades can be attributed to the variance in IQs.

Note that, unlike with other cognitive skills (e.g., reading proficiency; Baldwin, Kirsch, Rock, & Yamamoto, 1995), not a single nationwide study has been conducted on a representative national sample to link academic performance to IQs. The proponents of the g-based theory of intelligence might point out that modern IQ tests have not been administered to a representative sample in studies independent of test-standardization efforts.[2] Therefore, an ideal data set is unavailable. Psychologists have used various corrections to allow for the fact that real data tend to produce lower correlations than would ideal data.

One response is to correct the correlations for attenuation (unreliability of measurements). Such corrections are useful so long as three things are kept in mind. First, the correlations are for an "ideal" measurement situation, not the actual measurement situation. In other words, the validity coefficient obtained after correction for attenuation describes a situation that might exist in theory but does not exist in practice. Second, the correction for attenuation makes psychometric assumptions that may not be fully met. Third, the greater the amount of correction applied (i.e., the lower the initial reliabilities of the measurements being considered), the less accurate the correction is likely to be. Thus, to take a somewhat extreme example, if two measures are correlated with each other, both of which have reliabilities in the .40s, the corrected correlation after assuming the reliabilities are 1.00 is likely to be at least somewhat suspect.

A second and related response to imperfect data is to correct for restriction of range. Range can be restricted by problems in distributions of grades and by narrow ranges in abilities of participants. Grade inflation has restricted the range of school grades, making higher correlations harder to obtain. One can correct for restriction of range in the grades, but such a correction assumes one knows what the true grade distribution should be. Specialized groups also result in restriction of range, especially when one is working with an ability-restricted group, such as the gifted, learning disabled, or mentally retarded. Corrections when working with such groups can be problematical because they raise the question of the population to which one wishes to generalize results. If the special group is the *population* of interest, correction for restriction of range is not

[2] Note that standardization samples might have problems of their own (e.g., Gudjonsson, 1995).

appropriate. If the special group is a *sample* of a population with normal mean and standard deviation, then obviously the sample is a highly non-representative one and the investigator must question the generalizability of the results to a normal population.

In specialized populations, correlations may be low not only because of restriction of range, but because other factors are more important than intelligence in predicting performance (e.g., Lyon, 1996). For example, motivation may be more important in certain groups than in others. Thus, blind application of corrections may overcorrect because these groups truly do not display predictor-criterion correlations typical of those that would be obtained in the general population. In sum, corrections for restriction of range, like those for attenuation, should not be applied blindly. If they are applied, then it needs to be made very clear what question is being answered (e.g., generalization to a hypothetical normal population with hypothetically ideal measurements).

Whereas correlations of ability tests with achievement tests tend to be lowered by attenuation and restriction of range, such correlations are raised by the fact that standardized achievement tests tend to overlap in content and format more than might be desired. Sometimes the two kinds of tests are even hard to distinguish. This difficulty is illustrated by the SAT, which measures a mixture of ability and achievement constructs. The uncertain nature of what the test measures is perhaps illustrated by the fact that its name has changed from Scholastic Aptitude Test to Scholastic Assessment Test to SAT (which now is the official term for the test—it no longer is an acronym). The degree of overlap between tests is consistent with Sternberg's (1998) notion of abilities as a form of developing expertise. According to this notion, both ability and achievement tests measure essentially the same mix of constructs, with the two kinds of tests differing only in their emphases. But if the correlations of ability with achievement test scores are interpreted as representing some kind of "pure" relation between hypothetical unadulterated and distinct ability and achievement constructs, the degree of overlap will appear unduly impressive.

In surveying correlations, one must be sensitive to the possible existence of interactions. A comprehensive survey of research addressing interactions between educational treatments and individual differences in abilities and aptitudes was conducted by Cronbach and Snow (1977). The conclusion from the evidence they surveyed was that general cognitive abilities, as captured by IQ, are strongly linked with achievement indices in the humanities, sciences, and social sciences. Since the late 1970s this finding has been replicated multiple times, both cross-sectionally and longitudinally. Wherever and whenever IQ has been studied, on average, children who score higher on intelligence tests administered after age $2\frac{1}{2}$; learn more and better in school than do those who score lower (for reviews, see Brody, 1997; Neisser et al., 1996).

The relationship between IQ and educational achievement is reciprocal. IQs respond to adequate intellectual challenges and grow as an outcome of successful educational experiences (e.g., Ceci & Williams, 1997). Each additional month in school may increase a student's IQ when compared with the IQ expected had the student dropped out of school (Ceci, 1991). It is also the case, however, that a student's IQ and pursuit of educational opportunities are determined, at least partially, by IQ. Moreover, correlations between IQ scores and total years of education have been found to be strong: .60 for white males (Herrnstein & Murray, 1994), .55 in the task force report by the American Psychological Association (Neisser et al., 1996), and ranging between .16 and .90 in recent studies (Ceci, 1991). Specifically, IQ was found to be the single best predictor of the decision to obtain postsecondary education (Rehberg & Rosenthal, 1978), and econometric analyses have shown that each additional IQ point may lead to a decision by a student to stay in school a little longer (Heckman, 1995).

Of course, it may not be IQ itself that is responsible for these effects, but rather the encouragement or opportunities given to individuals with high IQ. In other words, these phenomena may be created not so much by IQ per se, but by the implicit views of those individuals and institutions who value IQ and of those people who have more of it. For example, tall individuals may be encouraged to pursue basketball in a way that short individuals are not, providing the tall individuals with opportunities not afforded the short ones.

Still, all in all, IQ accounts for only about 25% of the variance in schooling outcomes. This percentage would be somewhat higher if one corrected for attenuation in the measures, but because the reliabilities of IQ tests and of school performance as measured by averaged grades are very high, the percentage would not change much. Successful learning also depends on many other factors, both individual and environmental (Sternberg, 1997).

One determiner of what happens to people is their set of beliefs in their capacity to exercise control over their level of functioning and over environmental demands. In other words, the belief that one can succeed is often a prerequisite for success. Such beliefs have been referred to as *self-efficacy beliefs* (e.g., Bandura, 1996). The various psychological processes through which self-efficacy beliefs exert their influence are intimately involved in cognitive development. Researchers have shown that children's beliefs in their ability to regulate their own learning activities and master difficult subject matter—in other words, their self-efficacy—affect their academic motivation, interest, and scholastic achievement (for a review, see Zimmerman, 1995). Moreover, even among high-ability students, those who were less certain of their abilities and were less internally motivated to succeed reported more school-related negative affect and withdrawal behaviors than did children who perceived themselves as having more ability, feeling in control, and being successful (Miserandino, 1996). These results must be interpreted with caution, however, as self-efficacy beliefs may themselves be affected by abilities, achievement, and motivation.

In addition, higher parental efficacy regarding the academic achievement of children has a significant positive effect on their children's educational outcomes. For example, when children are matched for level of ability, those children with parents who expect higher scholastic achievement and convey these expectations to the school system are placed in more challenging academic programs. These children also achieve higher scholastic success than do those children whose parents do not get involved in the educational process (Dornbusch, 1994). In low socioeconomic (SES) families, parents who have high academic aspirations for their children and who are more involved in school activities generally have children who are more academically successful (Kao & Trienda, 1995). When assessed together, children's education-related self-efficacy, parents' self-efficacy regarding parental influence on children's education, and a number of other efficacy-related variables (e.g., prosocial behavior and moral engagement/disengagement), accounted for 58% of the variance in academic achievement, in the absence of intelligence scores (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). These data suggest the possibility that part of the predictive success of IQ may be due to the feelings of self-efficacy that IQ creates in those who have more of it. Such feelings may also lead to enhanced employment prospects and potential for generating wealth.

EMPLOYMENT PROSPECTS AND WEALTH GENERATION

To what extent do IQ scores, obtained before individuals enter the labor force, predict such indicators of social status as employment and wealth? This question has no straightforward answer, because both the outcome measures and the predictor (IQ) are tightly linked to two other variables, namely, the SES of one's family and one's amount of schooling.

Jencks (1979) estimated that parental SES predicts about 30% of the variance in young adults' social status and about 20% of the variance in their income. However, in each case, approximately half of this predictive power is attributable to the link between parental SES and young adults' IQs, which, in turn, have their own predictive value for social outcomes. IQs themselves predict about 25% of the variance in SES and about 15% of the variance in income. Controlling for parental SES results in a reduction of about 25% in the predictive power of IQ. Jencks (1979) observed that if two brothers who grew up in the same family were compared on their SES as adults, the brother who had the higher IQ in adolescence would tend to have the higher adult social status and income. This path, however, is mediated by amount of education. The higher-IQ brother would be more likely to get more education and, correspondingly, to have a better chance of succeeding socioeconomically.

When IQ is statistically controlled, no more than 2% of the variance in obtained SES is accounted for by schooling (Schmidt, 1996). The reciprocity of the IQ-schooling relationship is complex, however. Higher IQ predisposes an individual to seek more schooling, and more schooling raises the individual's IQ. As Ceci and Williams (1997) and Sternberg (1997) have stated, schooling provides a person with a key to the gateway to certain high-paying jobs. However, once inside, success in these jobs may have little to do with either years of education or level of ability but may, rather, depend on different factors.

IQ seems to be predictive of the reaching of all steps of career life in a stable society, where Western schooling is valued and rewarded, income is scaled in rough correspondence to years of education, and highly skilled labor is needed. What happens, however, in societies passing through a stage of turmoil, where the social value of education is not high, income does not correlate well or at all with amount of schooling, and there is little demand for highly skilled employees? Very little research evidence is available that can help answer this question. Various writings on cultural psychology, discussed later, suggest that, in most developing societies, there is a conflict between the kinds of intelligence valued by schooling and the kinds of intelligence valued by local communities. Is there a conflict also with career outcomes?

IQ and Career Outcome

There are statistically significant correlations between "general intelligence" composites and performance within a job (for reviews, see Hunt, 1995; Wagner, 1997). Hunter and Hunter (1984) reviewed a large number of studies and found that, when corrected for sample biases, the correlations between intelligence and job performance ranged from .27 to .61 (see also Schmidt & Hunter, 1998). When supervisor ratings were used as a criterion, the mean population validity coefficient was .53; it was higher for more intellectually demanding jobs (e.g., the validity estimate for "manager" was .53) and lower for less intellectually demanding jobs (e.g., the estimate for "clerk" was .27). Hunter (1983) reported that differences in intelligence accounted for as much as 29% of the variance in job-performance ratings after the ratings were adjusted for error variance. The author found that IQ was more powerful as a predictor of success on the job than was any other variable considered.

Other authors have offered somewhat lower validity coefficients, suggesting that the average observed correlation between cognitive ability test scores and job performance is between .20 and .40 (Hartigan & Wigdor, 1989; Wigdor & Garner, 1982; Wigdor & Green, 1991). The difference between the estimates of Hunter and of others is due in large part to Hunter's generally correcting for attenuation (unreliability of measurement) and restriction of range, resulting in his use of elevated idealized correlations rather than actual correlations.

The concern has been raised (e.g., Wagner, 1997) that the link between cognitive ability and job performance may be artificially inflated due to the absence in the tested models of some variables that may moderate the link between intelligence and job performance. To resolve this issue, Wagner has suggested that causal modeling studies would be desirable. So far, only a handful of such studies have been done, however; moreover, those studies have included only a limited set of variables beyond cognitive ability and performance. Collectively, the results of these studies suggest that the magnitude of the direct impact of cognitive ability on job performance generally decreases when other variables are included in the model. For example, Ree, Carretta, and Teachout (1995) proposed a causal model of job performance for more than 3,000 U.S. Air Force officers in pilot training. The dependent variable was rated performance during checkout flights. Independent variables were cognitive ability, job knowledge, and prior checkout flight performance. All correlations were corrected to account for possible biases. No direct effect of cognitive ability on job performance was found. The best predictor of job performance was prior job performance. However, a significant indirect effect of .35 linked job performance and cognitive ability through job knowledge, again suggesting that IQ may exercise its effect indirectly rather than directly.

In sum, cognitive ability predicts anywhere from 4% to 30% of the variance in job performance. Even researchers who are strong believers in the utility of IQ agree that it is unlikely that any more improvement in conventional tests will result in substantially higher predictive validity to quality of job performance for the tests (Schmidt, 1994). The tests also show some predictive validity in predicting other life outcomes, such as overall well-being.

IQ and Well-Being

Subjective well-being is traditionally defined as a predominance of positive thoughts, emotions, and attitudes about one's life (e.g., Myers & Diener, 1995). At the cognitive level, this concept refers to a global sense of satisfaction with various components of life (education, job, marriage, leisure, income, civic activities, etc.). At the affective level, higher levels of subjective well-being are characterized by primarily constructive and positive feelings, whereas lower levels of subjective well-being are related to feelings of depression, anxiety, and other forms of psychopathology.

Lower IQ has been suggested as perhaps the most significant factor associated with psychiatric disturbances in children (Howlin & Rutter, 1987). The link between low IQ and childhood psychopathology has been registered in epidemiological studies (Anderson, Williams, McGee, & Silva, 1989; Cook, Greenberg, & Kusche, 1994) as well as in studies of children evaluated because of their behavioral problems in schools (Carlson, Lahey, & Neeper, 1986). In addition, similar associations have been found for lower levels of intelligence and both internalizing (e.g., depressive and anxious symptomatology; Pianta & Castaldi, 1989) and externalizing (e.g., aggressive and delinquent behavior; Pianta & Caldwell, 1990) behaviors from kindergarten to first grade. Moreover, epidemiological studies have revealed links between lower intelligence and specific behavioral and emotional problems, such as hyperactivity (McGee, Partridge, Williams, & Silva, 1991) and delinquency (White, Moffitt, & Silva, 1989).

Early indicators of delinquency are crucial to understanding children's developmental trajectories and prospects. There have been numerous and consistent reports of a negative relation between IQ scores and delinquent behavior (see, for reviews, Hirschi & Hindelang, 1977; Wilson & Herrnstein, 1985). This finding is strong and robust and holds up when IQ is assessed over time (e.g., Moffitt, Gabrielli, & Mednick, 1981) and when delinquency is assessed through either official records of arrests and crimes or self-reports (e.g., Moffitt & Silva, 1988). This finding is also independent of social class (e.g., Wolfgang, Figlio, & Sellin, 1972), gender (Stattin & Magnusson, 1990), and race (e.g., Short & Strodtbeck, 1965). On average, people who demonstrate delinquent behaviors score

8 IQ points lower than do nondelinquents on standard intelligence tests (Lynam, Moffitt, & Stouthamer-Loeber, 1993). In addition, this relation is stronger for verbal tests than for performance tests (Prentice & Kelly, 1963; West & Farrington, 1973). There are at least three possible interpretations of this phenomenon: (a) the relationship between IQ and delinquency is spurious and caused by a third variable; (b) a delinquent lifestyle may result in lower IQ scores; and (c) low IQ scores may lead to delinquency. The last interpretation has received the strongest empirical support (Lynam, Moffitt, & Stouthamer-Loeber, 1993). The relationship also may be mediated by temperament.

A major assumption underlying the study of temperamental traits is that these early-emerging individual differences shape the course of personality development, its adaptive outcomes, and its problematic manifestations (Rutter, 1987). The pioneering work of Thomas and Chess (1977) introduced into developmental psychology the concept of *difficult temperament,* which refers to a specific temperamental profile, distinguishable in early childhood and predictive of academic difficulties and behavioral problems. Subsequent long-term longitudinal studies have revealed connections between early-childhood temperamental characteristics and adjustment problems in both later childhood and adolescence (e.g., Block, Block, & Keyes, 1988; Chess & Thomas, 1987). Such characteristics also are linked with lower school achievement (e.g., Caprara, Barbaranelli, Incatasciato, Pastorelli, & Rabasca, 1997). For example, a so-called temperamental factor of *lack of control* (i.e., emotional lability, restlessness, impulsiveness, and negativism) was initially identified at 3 years of age, and then reassessed at 5, 9, and 15 years of age (Caspi, Henry, McGee, Moffitt, & Silva, 1995). This factor not only was a strong single negative predictor of academic and behavioral adjustment, but it also reliably differentiated a group of violent offenders when reevaluation was done at the age of 18 (Henry, Caspi, Moffitt, & Silva, 1996). Several investigators reported that children with easy temperaments tend to have higher cognitive test scores (Belsky, 1980; Slomkowski, Nelson, Dunn, & Plomin, 1992). Although the causal nature of this link is still unclear, one possible explanation may lie in the moderating role of mastery motivation (for a review, see Shiner, 1998). For example, infants rated high on the mastery motivation questionnaire also were rated as more cooperative and less difficult on the temperament questionnaire, and they tended to be rated as more approachable and less irritable (Morrow & Camp, 1996). Therefore, it is possible that the link between easy temperament and higher cognitive test scores is attributable, among other factors, to children's persistence in acquiring a skill (due to their high mastery motivation) and to the higher degree of attention these children receive from adults who find them more approachable and less irritable.

DO COGNITIVE TEST SCORES REMAIN STABLE OVER TIME, OR DO THEY SHIFT AROUND?

Stability and Modifiability of IQ

In the previous discussion we showed that, in sections of the Western part of the developed world, IQ is predictive of global outcomes indicative of life success. Given that IQ is a consistent predictor of significant life outcomes, it is crucial that we examine the stability and modifiability of IQ. IQ does vary both longitudinally and as an outcome of controlled intervention. Two sources of evidence support this claim.

Evidence from studies of the natural course of development: Some get more intelligent, others get less intelligent. The Berkeley Guidance Study (Honzik, Macfarlane, & Allen, 1948) investigated the stability of IQ test performance over 12 years. The authors reported that nearly 60% of the sample changed by 15 IQ points or more from 6 to 18 years of age. A similar result was found in the Fels study (Sontag, Baker, & Nelson, 1958): Nearly two thirds of the children changed more than 15 IQ points from age 3 to age 10. Researchers also investigated the so-called intelligence lability score, which is a child's standard deviation from his or her own grand mean IQ. Bayley (1949), in the Berkeley Growth study, detected very large individual differences in lability across the span of 18 years. Rees and Palmer (1970) combined the data from five large-scale longitudinal studies, selecting those participants who had scores at both age 6 and age 12 or at both age 12 and age 17. They found that about 30% of the selected participants changed by 10 or more IQ points.

Sometimes, the effect of environment is dramatic. Two adoption studies were conducted in orphanages, one by Dennis (1973) in Iran and one by Rutter (1996) in Romania. Dennis found that children placed in Iranian orphanages had low IQs. Probably because they were reared in institutions of different quality, girls had a mean IQ of about 50 and boys of about 80. Children adopted out of an Iranian orphanage by the age of 2 had IQs that averaged 100 during later childhood; they were able to overcome the effects of early deprivation. Children adopted after the age of 2 showed normal intellectual development from that point but never overcame the effects of early deprivation; they remained mentally retarded. These results suggest that interventions to foster cognitive development need to start as early as possible.

Rutter's (1996) Romanian project showed increases in mean IQ from 60 to 109 for orphans who came to the United Kingdom before 6 months of age. These children showed complete recovery from early mental retardation. Those who came to the United Kingdom after 6 months of age showed, on average, continuing deficits.

Whereas the studies just summarized show what happens in the natural course of development, various intervention studies show what might happen under controlled conditions and targeted intervention.

Evidence from intervention studies: raising IQ by educational interventions. Stankov (1986) reported the results of a study in Yugoslavia conducted by Kvaschev, who exposed students in an experimental group to training in creative problem solving for 3 to 4 hours per week for 3 years. One year after the conclusion of the experiment, the students in both the experimental and control (untreated) groups were reevaluated. The students who participated in the training gained, on average, 8 IQ points more than did the students in the control group.

There also have been attempts to raise IQ scores through large-scale intervention programs. Such programs as Head Start and Head Jump are designed to enrich the school-related experiences of disadvantaged children. Usually the intervention program lasts for 1 or 2 years. The general trend observed in these programs is that test scores increase over the course of the program itself, but that, after the intervening forces are withdrawn, the gains fade with time. There are no statistically significant differences in the scores of experimental and control children by the end of elementary school. However, although there is no "direct" evidence of any impact of intervention on IQ gain, there is evidence of other kinds of gain: Compared with matched controls, the children from the intervention program are less likely to be retained, more likely to stay in mainstream classes and not end up in special educational settings, and more likely to obtain a high-school diploma (Consortium for Longitudinal Studies, 1983; Darlington, 1986).

It is possible that more extensive and individualized intervention might result in more pronounced and lasting effects. For example, the Carolina Abecederian Project for disadvantaged children—which started for a given child in the first year of the child's life—involved parents as well as children in at least one experimental group and continued throughout the preschool years. This program provided evidence for changes in IQ and academic achievement that are detectable up to age 15 (Campbell & Ramey, 1994, 1995). It is important to note, however, that even as it provided evidence for the modifiability of IQ and scholastic achievement, this project pointed to the limits of such interventions. These limits became apparent when the intervention children were compared with a community sample of children whose families had professional and academic backgrounds. A one-third standard-deviation increase in academic performance and intelligence associated with this preschool intervention eliminated only approximately 25% of the difference in performance between a selected group of disadvantaged children and the unselected community sample.

A number of targeted interventions also have been conducted in which, rather than engaging youngsters in positive educational and cognitive experiences, researchers attempted to teach school students specific cognitive and metacognitive skills (for reviews, see Perkins, 1995; Perkins & Grotzer, 1997). Although not all of these programs necessarily led to statistically noticeable IQ gains, they all improved children's school achievement and adjustment. In general, the evidence suggests that enhancements of intelligence in a broader sense can last for months and years, but there is no evidence that the obtained modifications are permanent without subsequent scaffolding or refresher intervention. Sternberg, Okagaki, and Jackson (1990) showed that a program to increase practical intelligence for school could have a significant positive impact on academic skills. In a related study, Sternberg, Torff, and Grigorenko (1998) showed that instruction emphasizing analytical, creative, and practical thinking and learning resulted in improved educational outcomes over memory- or critical-thinking based instruction. Children taught in a way that enables them to capitalize on their intellectual pattern of strength (analytic, creative, and/or practical) outperform students taught in a conventional way (Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999).

However, in spite of developmental and interventional fluctuations in IQ, it is well known that IQs are fairly stable during development. Multiple sources point to the relative stability of correlations between IQs registered at about 6+ years of age and subsequent indicators of intelligence registered later in life. For example, the classic study of Jones and Bayley (1941) presented correlations of IQ scores across successive years. Scores obtained at age 18 correlated ($r = .77$) with scores that had been obtained at age 6 and correlated ($r = .89$) with scores from age 12. To control for short-term fluctuations, scores were averaged across several successive years of testing. The mean for ages 17 and 18 was correlated ($r = .86$) with the mean for ages 5, 6, and 7. The correlation was even higher ($r = .96$) for averaged scores across ages 17 and 18 and across ages 11, 12, and 13. Table 1 (with data adopted from Sontag, Baker, & Nelson, 1958) shows stability of Stanford-Binet scale performance from 3 to 12 years of age. Two observations are noteworthy. First, the best predictor of IQ in a given year is the IQ from the previous year. Second, the predictive power

Table 1. Stability of Binet Scale Performance from 3 to 12 Years of Age

Age (in years)	*4*	*6*	*8*	*10*	*12*
3	0.83	0.73	0.60	0.54	0.46
5		0.87	0.79	0.70	0.62
7			0.91	0.82	0.73
9				0.90	0.81
11					0.90

of IQ in every subsequent year increases with the child's age. Similar data have been obtained for a variety of intelligence tests (e.g., Neyens & Aldenkamp, 1996; Schuerger & Witt, 1989).

Between generations, IQ is highly modifiable. Environment has a powerful effect on levels of cognitive ability. Perhaps the simplest and most potential demonstration of this is called the "Flynn effect" (Flynn, 1984, 1994; see also Neisser, 1998). The basic phenomenon of the Flynn effect is an increase in IQ throughout successive generations around the world during the past 30 years.

The Flynn effect is powerful (Neisser, 1998), showing an increase in IQ of up to 18 points per generation for tests of fluid intelligence (Cattell 1971; Horn & Cattell 1966) such as the Raven Progress Matrices, which measure a person's ability to cope effectively with relatively new stimuli. The mean effect has been inexplicably greater for tests of fluid abilities than for tests of crystallized knowledge-based abilities.

This effect must be environmental because a successive stream of heritable changes could not have accumulated and exerted so much influence in such a short period of time. Many environmental factors (reviewed in Neisser, 1998) have been suggested as possible causes of the gain, such as better nutrition, increased and better schooling, and exposure to technology. Thus, scores on psychometric tests of intelligence indicate that environment must be exerting a powerful effect on intelligence; intelligence can be and is being modified.

In sum, IQ is changeable in the natural course of development and within the frame of targeted intervention, but it is usually fairly stable. This apparent contradiction makes sense when several additional issues are considered. First, correlations are usually obtained under a specific set of conditions, namely, ones that assume no specific intervention to change IQ. Interventions might lower the correlations. Second, there is no single definition of IQ. There is some consensus that IQ represents only one facet of intelligence and that this facet itself is a very complex reflector of multiple psychological functions (see Sternberg, 1982, 1994, 2000). Third, it has been suggested that IQ largely reflects a broad neuropsychological function, known as "executive function," that includes sustaining attention and concentration, reasoning abstractly, forming goals, anticipating and planning, initiating purposeful behavior, and self-monitoring (Lynam, Moffitt, & Southamer-Loeber, 1993). Executive-control deficits interfere with a person's ability to monitor and control his or her own behavior. They also influence, directly and indirectly, many aspects of the person's life. Finally, and most important, changes in absolute levels of a score are independent of the degree to which patterns of individual differences change. One could have large changes in levels of scores with anywhere from no change to major change in rank orders. The relative variability or stability of IQ also can depend on how it is tested.

INTELLIGENCE TESTING

We have shown that general indices of cognitive ability are predictive, to some degree, of broad life outcomes. These indices are relatively stable, even though they fluctuate in the normal course of development and are modifiable in controlled conditions. We now describe more fully the sources of these indices.

The most widely used source of information about both general and specific cognitive abilities is intelligence tests. The majority of intelligence tests have been developed within the psychometric paradigm, an approach based on the identification of abilities (verbal and spatial abilities, memory, reasoning, etc.) through the factor analysis of sets of diverse cognitive tasks. Most modern psychometric tests address both a general factor (the so-called *g*-factor, reflecting the positive manifold of correlations between various cognitive abilities) and distinct, though correlated, group factors. Whereas all of the psychometric tests have a full-scale or a composite index that, presumably, reflects the *g*-factor, no single test completely overlaps with any other test in terms of the precise set of cognitive abilities that is measured.

Despite the complex structure of the modern tests, at the applied, practical level, the *g*-factor remains the most-used attribute. Most of the studies, especially meta-analytic and longitudinal studies, have employed full-scale and composite scores, rather than group-factor or subtest scores. As a result, knowledge of the etiology, as well as of educational and vocational applications, of the broad abilities at the level below *g,* is remarkably limited. As Daniel (1997) has stated, at this point there are not yet enough raw data to evaluate the relevance of all the group (both broad and narrow) ability factors to schooling, career paths, and other areas of practical application. But how are any of these factors actually measured through tests?

This section provides a brief overview of the modern leading tests of intelligence and their predictive validity. To conduct the evaluation of the predictive validity of selected intelligence tests for children, we analyzed only those studies that met the following criteria: (a) the study was conducted with the goal of evaluating the construct validity of a given test; (b) the study was conducted no earlier than 1987 to avoid repetition with previous comprehensive reviews (e.g, Reynolds, 1987) and to cover the latest versions of most widely used tests (the latest version of the Stanford-Binet Intelligence Test appeared in 1986); and (c) the study was conducted on a normal (rather than clinical) population.

Indicators of Cognitive Abilities in Infancy

To evaluate whether developmental milestones are attained at an appropriate age, researchers have devised developmental schedules for infancy and early childhood. In essence, these schedules can be used to establish a series of evaluative normative assessments, according to which the adequacy of infants' sensory-motor and mental development can be evaluated. Altogether, roughly half a dozen major scales have been developed (e.g., Cattell, Gesell, and Merrill-Palmer tests; for a review, see Stott & Ball, 1965), but, in recent years, the Bayley Scales (Bayley, 1993; Nellis & Gridley, 1994) have been the most popular. The popularity of these scales is due primarily to their superior psychometric qualities and to their attention to questions of standardization and normalization (Columbo, 1993).

The *Bayley Scales of Infant Development-II* (BSID, Bayley, 1993) represent the first restandardization of the Bayley test in 25 years. The history of research with the BSID is replete with empirical demonstrations of both the usefulness and the futility of infant testing. On the one hand, the BSID has proven to be useful for the assessment of the current status of the infant (Lipsitt, 1992). On the other hand, the testing of children younger than 18 months of age with the BSID has yielded little predictive validity if one is interested in anticipating the later intellectual or cognitive development of a given child (Columbo, 1993). As a matter of fact, Dr. Bayley herself expressed reservations about the use of the test for predictive purposes, suggesting that researchers examine the mother rather than the child. Researchers have arrived at the conclusion that, for children younger than

18 months, the BSID does not yield consistent results. In addition to lacking predictive power in the domain of intelligence, the Bayley does not predict either child behavioral scores or psychiatric diagnoses (Dietz, Lavigne, Arend, & Rosenbaum, 1997). Burns, Burns, and Kabacoff (1992) made the case from their data that 3-month-olds are more like other 3-month-olds across a variety of tasks than they are like themselves over a long period of time. In other words, at 3 months of age, a normally developing and a mentally retarded child appear to be very similar in their capacities as assessed by the BSID, but when they reach the age of 3 years, they will be very different. To sum up, then, the predictive validity of the BSID scales is poor.

One of the most frequently cited summaries evaluating the predictive validity of infant tests is one by McCall (1979), in which the results of 20 studies conducted from 1933 to 1975 with normal infants were analyzed. McCall showed that predictive validity does not appear to vary as a function of the particular infant test and summarized the results across the tests. The most interesting observation derived from this analysis is that when a direct comparison is made of (a) correlations between various infant tests and childhood Binet performances and (b) correlations between childhood Binet performance at one point and childhood Binet performance at a later point, the mean correlational difference is 0.38 units. Similar results were obtained in a number of recent studies (for details, see Columbo, 1993), where the correlations were calculated between various age-appropriate cognitive scales in normal samples of children. The range of these correlations is over 0.50 units. The explained variability in cognitive performance of children older than 8 years of age proves to be the lowest (less than 1%) when the assessment was carried out at the ages of 1–6 months, and the highest (25%) when the assessment was conducted at the ages of 19–30 months. When combined, infant scales explained about 16% of the variability in intellectual performance of children younger than 4 years of age and about 8% thereafter (adapted from Columbo, 1993). Thus, these results point out that infancy-to-childhood correlations are considerably lower than are childhood-childhood correlations.

To summarize, for infants scoring in the normal range, the traditional infant tests do not predict childhood levels and certainly do not well predict adult levels of intelligence (for more detail, see the debate between McCall, 1972, and Wilson, 1972). Probably this lack of predictive validity is because such tests are based on the now largely discredited view that infant intelligence is largely sensorimotor in nature (Piaget, 1972). A different view of infant intelligence based on information-processing notions has had greater success.

Early Measures of Information Processing

Since the 1970s researchers have been emphasizing the relative lack of stability in the early conventional mental test performances of normal and at-risk infants (Kopp & McCall, 1982; McCall, 1979). At the same time, however, numerous studies have demonstrated that infant information processing (i.e., how infants store, retrieve, discriminate, and recognize information) predicts later intelligence somewhat better (Bornstein & Sigman, 1986; Fagan & Singer, 1983; McCall & Carriger, 1993). It is interesting to note that these measures do not correlate with contemporaneous measures of infant competence, represented by the Bayley Mental Development Index and the Psychomotor Developmental Index.

Two experimental paradigms are most frequently used to assess infants: the habituation and the paired-comparison paradigms (for a discussion of the stability and reliability of the paradigms, see McCall & Carriger, 1993; Rose & Feldman, 1987). Both paradigms have been used widely to study the development of early cognitive processes in infants (e.g., Bornstein, 1985; Rose, 1994).

Several recent studies have pointed to moderately high predictive associations with later cognitive measures using both the habituation paradigm (e.g., Caron, Caron, & Glass, 1983) and the paired-comparison paradigm (e.g., Fagan, 1984; see also review by Bornstein & Sigman, 1986). For example, research on infant attention differentiates "short-looking" babies (those who need only about 10 s of familiarization time to demonstrate preference for a novel stimulus) and "long-looking" babies (those who need about 40 s to process the same stimulus). In two follow-up assessments at ages 5 and 8, "long-lookers" scored lower on intelligence tests than did "short-lookers" (Sigman, Cohen, Beckwith, Asarnow, & Parmelee, 1991). The finding was replicated and extended at 12 and 18 years of age. The predictive patterns of infant fixation duration and inhibition were specific rather than general—the prediction held for tasks that were intellectually challenging, but it did not hold for the simple ability to inhibit responses to a previously correct stimulus and to shift to a different stimulus. In other words, these scores predicted intellectual performance on tasks that were intellectually nontrivial. Overall, a recent meta-analysis of 31 samples estimated the correlation between the infant information-processing measures and childhood IQ at .36 (McCall & Carriger, 1993). Although this correlation is modest in absolute value, its relative value is higher than that of an average infant-child correlation on tests of intellectual functioning. Moreover, the correlation demonstrates that other indicators of cognitive functioning in infancy are as informative about childhood functioning as are global indices such as the mental development index (MDI) or physical development index (PDI) of the Bayley.

Yet another frequently used indicator of early cognitive ability that has proven to be a valid predictor of future intellectual development is language delay. There are many tests of receptive and expressive language in early childhood, including both specially designed instruments, for example, *Sequenced Inventory of Communication Development-Revised* (Hedrick, Prather, & Tobin, 1984), *Preschool Language Scale-3* (Zimmerman, Steiner, & Pond, 1992) and subtests of global scales, for example, receptive and expressive language subscales of the *Mullen Scales of Early Learning* (Mullen, 1995). Results from three longitudinal studies (Fundudis et al., 1979; Richman et al., 1983; Silva et al., 1983) investigating the predictive power of language delay on further cognitive development were strikingly similar. All studies found that, in comparison with controls, children with early language delay had significantly lower IQs, especially verbal IQs, and either a significantly lower mean reading score or a significantly higher likelihood of being poor readers. We consider next some of the tests used to measure IQs.

LEADING TESTS OF INTELLIGENCE (2+ YEARS OF AGE): BRIEF DESCRIPTION

In contrast to the lack of predictive power exhibited by standardized tests for infants, standardized tests of development for children in middle and later childhood do a much better job of predicting (a) subsequent IQ scores, (b) scholastic achievement, and (c) school grades. These tests are based on related although nonidentical hierarchical theories of intelligence. What are the leading tests of intelligence in childhood?

The *Wechsler Preschool and Primary Scale of Intelligence—Revised* (WPPSI-R, Wechsler, 1989) is the most recent version of a test that was initially developed in the late 1960s. The test is an individually administered clinical instrument for assessing the intelligence of children aged 3 years through 7 years, 3 months. It yields Verbal and Performance IQs as well as Full Scale IQ.

The *Wechsler Intelligence Scale for Children—Third Edition* (WISC-III, Wechsler, 1991) is the most current edition of the test, which was initially developed in the late 1940s. This test is an individually administered clinical instrument for assessing the intellectual abilities of children aged 6 years through 16 years, 11 months. It yields the same scores as the WPPSI-R; in addition, although based originally on a conception of intelligence that emphasized the pervasive nature of general intelligence, the current edition of the WISC offers scores for four factors (Verbal Comprehension, Perceptual Organization, Processing Speed, and Freedom from Distractibility).

The Stanford-Binet Intelligence Scale: Fourth Edition (SBIS, Thorndike, Hagen, & Sattler, 1986a, 1986b) is an individually administered intelligence test used to assess the cognitive abilities of individuals from age 2 to adult. The fourth edition is the latest version of the Stanford-Binet, which was originally published in 1916. The SBIS is based on a three-level hierarchical model consisting of *g* (a general ability factor), and three second-order factors (Crystallized Abilities such as Verbal Reasoning and Quantitative Reasoning, Abstract—Visual Reasoning, and Short-Term Memory).

The *Differential Ability Scales* (DAS, Elliott, 1990) form an individually administered battery of cognitive and achievement tests for children and adolescents from ages $2\frac{1}{2}$; years through 17 years. The Cognitive Battery is organized into a set of core subtests that yield the General Conceptual Ability (GCA) score and a set of diagnostic subtests that provide additional information on specific abilities. There is also an intermediate layer of so-called cluster scores, linking specific subtests to the GCA score. The structure of the test is flexible and age-dependent. Thus, for children aged 2 years, 6 months to 3 years, 5 months, there are no cluster scores. Their absence is because abilities are relatively undifferentiated at this young age.

The Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) measures both intelligence and achievement. It is designed to assess both normal and exceptional children of ages $2\frac{1}{2}$ through $12\frac{1}{2}$ years. Four global areas of functioning are assessed: Sequential Processing, Simultaneous Processing, Mental Processing Composite, and Achievement. There are a total of 16 subtests (3 sequential, 7 simultaneous, and 6 achievement), but not all subtests are administered at every age. Unlike the other tests, this test draws on Alexander Luria's (1980) theory of the functional systems of the brain.

The Standard Raven Progressive Matrices (SPM; Raven, 1960), drawing on Spearman's (1923, 1927) theory of general ability, consists of 60 nonverbal matrix problems, which are separated into five sets of 12 designs each. Within each set of 12, the problems become increasingly difficult. Each individual design has a missing piece. The participant's task is to select the correct piece to complete the design from among six to eight alternatives. Another test, referred to as Coloured Progressive Matrices (Raven, 1965), has been developed for children in the 5–11 age range and the elderly (65+ years of age). Similarly, persons suspected to be of high intellectual ability can be administered the Advanced Progressive matrices (Raven et al., 1992). The SPM is considered to be one of the most reliable instruments for measuring general intelligence, especially in its fluid aspects (Court, 1988; Raven, 1989). The latest edition of the tests was published in 1995.

The *Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Cognitive Ability* (WJ-R COG; Woodcock & Johnson, 1989) is designed for use with individuals aged 24 months through 95+ years. The theoretical framework of this test is the Horn-Cattell G_f-G_c Theory (Horn & Cattell, 1966; Horn & Noll, 1997). The test contains 21 tests of cognitive ability measuring Fluid Reasoning, Crystallized Intelligence, Visual Processing, Auditory Processing, Short-Term Memory, Long-Term Retrieval, and Quantitative Knowledge.

A relatively new test is the *Cognitive Assessment System* (CAS—Naglieri & Das, 1997), measuring Sequential Processing, Simultaneous Processing, Planning, and Attentional functions. The test is so new that extensive data have yet to be collected.

Table 2 presents the results of summary analyses of 68 different studies performed on various samples of participants addressing the criterion validity of the major childhood IQ tests. The results presented in Table 2 suggest that composite scores on various IQ tests are convergent (the correlations range between .50 and .88). In other words, 25% to 77% of the variance in IQs obtained by different tests of intelligence is probably attributable to common individual variation in measured intelligence rather than to test-specific variance.

Generally, there is a vast amount of variation in such correlations when criterion validity coefficients of various tests are examined for groups of exceptional individuals. These coefficients range from low to high: for example, for two groups of children with learning disabilities, the correlations between the SBIS and WISC-R ranged from 0.49 (Brown & Morgan, 1991) to .92 (Phelps, Bell, & Scott, 1988). Moreover, the validity coefficients resulting from the studies of gifted children are consistently lower, such as .21 (McCall et al., 1989), .39 (Phelps, 1989), .70 (Hayden et al., 1988), than they are for broader samples, almost certainly in part because of restriction of range.

Thus, in general, the highest validity coefficients are associated with studies whose samples demonstrate more variability (age, race, gender, and ability) as a group. On average, when the sample is even somewhat restricted, the criterion validity indicators of *all* intelligence tests tend to drop.

Up to now, we have considered the success of various intelligence tests when used as predictors of future success. But the overwhelming majority of studies have been done in the developed world, and it is

Table 2. Intertest Correlations

	WPPSI	*DAS*	*K-ABC*	*WISC*	*SB*	*Ravens*
WPPSI *(Full Scale)*		85	99	238	260	
DAS *(General Conceptual Ability)*	.86		50	153	142	
K-ABC *(Mental Processing Composite)*	.50	.69		897	957	
WISC *(Full Scale)*	.85	.88	.69		1572	837
Stanford-Binet *(Composite)*	.72	.82	.75	.72		Not reported
Raven's Progressive Matrices *(any form)*				.59	.68	

Note: Based on 68 studies. Sample sizes are shown above the diagonal; correlations are shown below diagonal.

All sample correlations were subjected to an r-to-z' transformation. The normalized (z') coefficients were then weighted by the size of the sample minus 3, added up, and this quantity was then divided by the sum of all sample sizes minus 3 for the set of samples included in the analysis. The resulting z'-scores were transformed back into r(s). These procedures had the effect of normalizing the sample correlations and rendering them as unbiased estimators of the population correlation.

hazardous to extend these results to the developing world. Indeed, there are even constraints on these studies as regards the developed world. We consider these constraints next.

CONSTRAINTS ON FINDINGS REGARDING IQ: WHAT ARE THE IMPLICATIONS FOR TESTING IN LOW-INCOME COUNTRIES?

Taken at face value, the story of conventional tests of intelligence seems to be one of modest to moderate but unequivocal success. But do the data tell exactly the story they appear to tell? We believe they do not, and that unless one digs deeper—for the story behind the story—one runs the risk of telling the wrong story about the right data. Consider, for example, the role of practical intelligence.

The Role of Practical Intelligence

Practical intelligence is the ability to adapt to, shape, and select real-world environments (Sternberg & Wagner, 1986). Research from diverse sources suggests that practical intelligence is factorially distinct from the kind of academic intelligence measured by conventional tests of intelligence, such as the Wechsler or Stanford-Binet series (Sternberg, Forsythe, et al., 2000). It also predicts a wide variety of criterion behaviors at levels comparable to that of IQ, suggesting that tests of practical intelligence might provide useful supplements to conventional tests of intelligence.

Nuñes (formerly Carraher) has done a series of studies over the years investigating the mathematical skills of Brazilian street children (Carraher, Carraher, & Schliemann, 1985; Carraher, Carraher, & Schliemann, 1987; Nuñes, 1994; Nuñes, Schliemann, & Carraher, 1993). The example of Brazilian street children is an apt one for the illustration of the construct of practical intelligence, because as Nuñes points out, the survival of these children is threatened on a daily basis. If the children are unable successfully to run a street business, and lapse into crime, the chances of their being murdered are quite high. Nuñes has found that the same children who can do the mathematics to run a successful street business are often failing math in school or otherwise show only minimal competence in math in academic settings. Similar results have been obtained by Ceci and Roazzi (1994), suggesting the findings are generalizable across investigators. As pointed out by Anderson, Reder, and Simon (1996), one needs to be careful about the exact conclusions one draws from studies such as these. For example, the exact computations required in one situation may not be the same as the computations required in another. But in terms of *adaptive functioning,* the point is that the people who are best able to adapt in one circumstance often are not those best able to adapt in another.

In a related study conducted near Kisumu, Kenya, we found that children's knowledge of the use of natural herbal medicines to combat illness is significantly *negatively* correlated with scores on tests of crystallized (Mill Hill Vocabulary in English and a comparable test in Dholuo, the home language) abilities (Sternberg, Nokes, et al., in press). In other words, practical intellectual skills were actually inversely associated with academic intellectual skills.

Lave (1988) also did related studies among Berkeley, California, housewives. She found that the same housewives who had no trouble doing comparative price calculations in the supermarket (before the introduction of unit pricing) were unable to complete most of the problems on a standard paper-and-pencil test of mathematical knowledge given in a classroom.

Investigating a different population, Ceci and Liker (1986) found that men's handicapping abilities for predicting outcomes of horse races were unrelated to their IQs. Moreover, successful handicappers had an average IQ of only about 100, despite the complexity of the handicapping task.

Ceci and Bronfenbrenner (1985) looked at quite a different task. They gave children a time-estimation task either in a classroom or at home. Strategies and quality of performance were very different in the two settings, suggesting that the context in which the judgments were made had a major impact both on how they were made and how well they were made.

In a very different context, Fiedler and Link (1994) reported that IQ positively predicted leadership performance under conditions of low stress but negatively predicted this same performance under conditions of high stress; in contrast, acquired knowledge of the kind that is essential for practical intelligence positively predicted leadership performance under conditions of high stress but negatively predicted under conditions of low stress.

Sternberg and his colleagues also examined practical intelligence in work settings. In a series of studies conducted over a period of about 15 years, Sternberg, Wagner, Williams, and Horvath (1995; see also Sternberg, 1997; Sternberg, Forsythe, et al., 2000) have reported that in tests of practical intelligence for U.S. managers, military leaders, salespeople, teachers, and children in school, measures of practical intelligence (a) do not correlate with IQ-based measures, (b) predict success in school or on the job as well as better than do IQ-based tests, and (c) can show changes as a function of learning from experience.

In conclusion, practical intellectual skills are, on average, relatively independent of academic intellectual skills and, in special circumstances, may even be inversely related to them. Yet these practical skills are essential for real-world adaptation, and in the long run may make more of a difference to everyday adaptation and economic productivity than do academic skills. Any one or several of these studies could be questioned on one or more grounds. For example, perhaps tests of practical intelligence are really tests of knowledge of some kind (Schmidt & Hunter, 1993). However, the plausibility of the Schmidt-Hunter account is undermined by the fact that tests of tacit knowledge tend not to correlate with IQ, whereas tests of job knowledge do so correlate. Moreover, *all* tests measure knowledge of some kind (Sternberg, 1998). We believe that, regardless of one's views of any single study, the studies taken together suggest that practical intelligence differs in major respects from academic intelligence.

The cross-cultural generalization of cognitive tests and scores on them is anything but straightforward. We may be eager to jump to conclusions on the basis of translated tests, only to find that such conclusions are false and without merit. But because such tests always yield scores—whether they are valid or not—we may be unaware of the falsity of the conclusions that can be drawn. Intelligence almost certainly has common elements across cultures (e.g., the need to recognize, define, and solve problems), but the appropriate content that instantiates those elements sometimes may differ from one culture to another.

Berry (1984) and more recently Sternberg and Kaufman (1998) have reviewed the literature on cross-country and cross-cultural conceptions of intelligence. Different cultures have different, and sometimes radically different, conceptions of the concept of intelligence. Indeed, cultures may differ in terms of whether they even have a word that provides a reasonable approximation to the concept of intelligence (as expressed in English). For example, in one such study, Yang and Sternberg (1997) found that Taiwanese conceptions of intelligence (for which there is no precise translation) included a general academic factor, but also included factors of interpersonal intelligence, intrapersonal intelligence, intellectual self-assertion, and intellectual self-effacement. Harkness, Super, and Keefer (1992) found that, in their studies in Kenya, parents defined intelligence among their children as the ability to do without being told what needed to be done around the homestead. Even more tellingly, given the American and Northern European emphasis on speed, certain Africans define intelligent people as slow in thought and action (Wober, 1974). Thus, a test that measures the Western conception of intelligence may be measuring something else, or, in a sense, nothing at all in a non-Western culture.

Even within a single country, different ethnic groups may have very different conceptions of what constitutes intelligence. For example, Okagaki and Sternberg (1993) studied parental conceptions of intelligence among different ethnic groups in San Jose, California, such as Latino, Asian, and Anglo parents. They found that the Latino parents emphasized the importance of social-competence skills in intelligence more than did the Anglo or Asian parents. But also the teachers' conception of intelligence was closer to the Asian and Anglo conception, with their emphasis on cognitive competence, than to the Latino conception. Moreover, the better the match of the parental conception of intelligence to that of the teacher, the better the children of the given ethnic group were doing in school. In other words, teachers have a set of values with regard to intelligent behavior but view it as "correct" rather than as culture-bound.

The same principle applies elsewhere in the United States. Heath (1983) studied conceptions of intelligence among African American and Anglo American groups of different socioeconomic classes in North Carolina and found again that the teachers' conceptions of intelligence were a much better match to the notions of the Anglo Americans than to the notions of the African Americans, possibly partly resulting in better achievement on the part of the Anglo Americans.

Sometimes members of other cultures interpret problems in ways that lead them to score poorly, even though their interpretations are valid within their own cultural context. For example, Luria (1980) found that central Asian peasants refused to accept syllogism problems as posed. When asked a question such as: "From Shakhimardan to Vuadil it is three hours on foot, while to Fergana it is six hours. How much time does it take to go on foot from Vuadil to Fergana?" the respondent might say, "No, it's six hours from Vuadil to Shakhimardan. You're wrong . . . it's far and you wouldn't get there in three hours" (p. 129). In a similar vein, researchers found that adult members of the Kpelle tribe, given a sorting task, tended to sort functionally, the way less intelligent adults in the United States or Europe would. The researchers were unable to get the Kpelle to sort taxonomically (i.e., supposedly more maturely), until they asked the Kpelle to sort in the way a foolish person would. The Kpelle then had no trouble sorting taxonomically (Cole, Gay, Glick, & Sharp, 1971; Glick, 1968). Similarly, Bruner, Olver, and Greenfield (1966) found that children of the Wolof tribe in rural Senegal preferred to sort by color rather than by taxonomic characteristics if they lacked Western schooling.

Greenfield (1997) has pointed out that, in collectivistic cultures, knowledge is not always viewed as residing in the individual. Rather, it may be viewed as residing in the collective. For example, among Zinacantecan Maya girls in Chiapas, Mexico, the notion that a girl would answer questions on her own from an independent perspective was largely incomprehensible. It was expected that when a question was asked, the mother, representing better the collective knowledge of the village, would be the one to answer. The idea that the child would take a test on individual knowledge or her own way of thinking seemed rather absurd.

In summary, scores from tests used in cultures or subcultures other than those for which the tests were specifically created are suspect, and probably of doubtful validity in many if not most cases.

CONCLUSION

What is to be concluded from all this? We believe that seven conclusions follow from the available data. First, relative levels of IQ show moderate to high consistency across individuals throughout childhood and early adulthood, although interventions may reduce the level of consistency. Second, infant IQ as traditionally measured is a relatively poor predictor of later performance, although newer information-processing measures show promise for improving prediction. Third, IQ is a relatively good predictor of many kinds of childhood and adult outcomes, although many other factors contribute to these outcomes as well. Broader tests of intelligence such as those being proposed and explored (e.g., Gardner, 1983; Sternberg, 1997) offer possibilities for increasing levels of prediction. Fourth, IQ is a better predictor of more academic kinds of performances than of less academic kinds of performance but shows some value in prediction to even nonacademic kinds of performances. Fifth, there is some degree of what has come to be called the "indifference of the indicator," in that a variety of different tests of IQ yield essentially similar results. Sixth, the quality of prediction to success in the developing world remains, for the most part, to be shown, because almost all the validity studies available have been conducted in Western settings. Finally, there is no reason for complacency. Levels of prediction have remained relatively stable over time, suggesting the need for broader kinds of measurements of all the varieties of skills that contribute to success. Thus, those who want to eliminate intelligence testing altogether bear the burden of proof to show that they have better or even equal measures. But those who want to improve intelligence testing have their work cut out for them, and everyone should wish them success, as science depends on building upon, rather than being fixed in, the past or even the present.

REFERENCES

ANDERSON, J. R., REDER, L. M., & SIMON, H. A., (1996). Situated learning and education. *Educational Researcher, 25,* 5–11.

ANDERSON, J., WILLIAMS, S., McGEE, R., & SILVA, P. (1989). Cognitive and social correlates of *DSM-III* disorders in preadolescent children. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 842–846.

BALDWIN, J., KIRSCH, I. S., ROCK, D., & YAMAMOTO, K. (1995). The literacy proficiencies of GED examinees. Princeton, NJ: Educational Testing Service.

BANDURA, A. (1996). *Self-efficacy: The exercise of control.* New York: Freeman.

BANDURA, A., BARBARANELLI, C., CAPRARA, G. V., & PASTORELLI, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development, 67,* 1206–1222.

BAYLEY, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *Journal of Genetic Psychology, 75,* 165–196.

BAYLEY, N. (1993). *Bayley scales of infant development* (2nd ed.). San Antonio, TX: The Psychological Corporation.

BELSKY, J. (1980). Child maltreatment: An ecological integration. *American Psychologist, 35,* 320–335.

BERRY, J. W. (1984). Towards a universal psychology of cognitive competence. In P. S. Fry (Ed.), *Changing conceptions of intelligence and intellectual functioning.* Amsterdam: North-Holland.

BERRY, J. W., & BENNETT, J. A. (1992). Cree conceptions of cognitive competence. *International Journal of Psychology, 27,* 73–88.

BLOCK, J., BLOCK, J. H., & KEYES, S. (1988). Longitudinally foretelling drug usage in adolescence: Early childhood personality and environmental precursors. *Child Development, 59,* 336–355.

BORNSTEIN, M. H. (1985). Human infant color vision and color perception. *Infant Behavior & Development, 8,* 109–113.

BORNSTEIN, M. H., & SIGMAN, M. D. (1986). Continuity in mental development from infancy. *Child Development, 57,* 251–274.

BRODY, N. (1997). Intelligence, schooling, and society. *American Psychologist, 52,* 1046–1050.

BROWN, T. L., & MORGAN, S. B. (1991). Concurrent validity of the Stanford-Binet, 4th Edition: Agreement with the WISC-R in classifying learning disabled children. *Psychological Assessment. A Journal of Consulting and Clinical Psychology, 3,* 247–253.

BRUNER, J. S., OLVER, R. R., & GREENFIELD, P. M. (1966). *Studies in cognitive growth.* New York: Wiley.

BURNS, W. J., BURNS, K. A., & KABACOFF, R. I. (1992). Item and factor analyses of the Bayley Scales of Infant Development. *Advances in Infancy Research, 7,* 199–214.

CAMPBELL, F. A., & RAMEY, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low income families. *Child Development, 65,* 684–698.

CAMPBELL, F. A., & RAMEY, C. T. (1995). Cognitive and school outcomes for high risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32,* 743–772.

CAPRARA, G. V., BARBARANELLI, C., INCATASCIATO, M., PASTORELLI, & RABASCA, A. (1997). Emotional instability, physical and verbal aggression, and prosocial behavior as precursors of scholastic achievement and social adjustment. In S. Feshbach & J. Zagrodzka (Eds.), *Aggression: Biological, developmental, and social perspectives.* New York: Plenum Press.

CARLSON, C. L., LAHEY, B. B., & NEEPER, R. (1986). Direct assessment of the cognitive correlates of attention deficit disorders with and without hyperactivity. *Journal of Psychopathology and Behavioral Assessment, 8,* 69–85.

CARON, A. J., CARON, R. F., & GLASS, P. (1983). Responsiveness to relational information as a measure of cognitive functioning in nonsuspect infants. In T. Field & A. Sostek (Eds.), *Infants born at risk: Psychological, perceptual and cognitive processes.* New York: Grune & Stratton.

CARRAHER, T. N., CARRAHER, D. W., & SCHLIEMANN, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Psychology, 3,* 21–29.

CARRAHER, T. N., CARRAHER, D. W., & SCHLIEMANN, A. D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education, 16,* 37–44.

CASPI, A., HENRY, B., McGEE, R. O., MOFFITT, T. E, & SILVA, P. A. (1995). Temperamental origins of child and adolescent behavior problems: From age three to fifteen. *Child Development, 66,* 55–68.

CATTELL, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

CECI, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722.

CECI, S. J., & BRONFENBRENNER, U. (1985). Don't forget to take the cupcakes out of the oven: Strategic time-monitoring, prospective memory and context. *Child Development, 56,* 175–190.

CECI, S. J., & LIKER, J. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General, 115,* 255–266.

CECI, S. J., & ROAZZI, A. (1994). The effects of context on cognition: Postcards from Brazil. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context.* New York: Cambridge University Press.

CECI, S. J., & WILLIAMS, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52,* 1051–1058.

CHESS, S., & THOMAS, A. (1987). *Origins and evolution of behavior disorders.* Cambridge, MA: Harvard University Press.

COLBY, A., JESSOR, R., & SHWEDER, R. (Eds.). (1996). *Ethnography and human development.* Chicago: University of Chicago Press.

COLE, M., GAY, J., GLICK, J., & SHARP, D. W. (1971). *The cultural context of learning and thinking.* New York: Basic Books.

COLUMBO, J. (1993). *Infant cognition: Predicting later intellectual functioning.* Newbury Park, CA: Sage.

CONSORTIUM FOR LONGITUDINAL STUDIES. (1983). *As the twig is bent . . . lasting effects of preschool programs.* Hillsdale, NJ: Erlbaum.

COOK, E. T., GREENBERG, M. T., & KUSCHE, C. A. (1994). The relations between emotional understanding, intellectual functioning, and disruptive behavior problems in elementary-school-aged children. *Journal of Abnormal Psychology, 22,* 205–219.

COURT, J. H. (1988). *A researcher's bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary Scales* (7th ed.). Cumberland Park, South Australia: J. H. Court.

CRONBACH, L. J., & SNOW, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

DANIEL, M. (1997). Intelligence testing: Status and trends. *American Psychologist, 52,* 1038–1045.

DARLINGTON, R. B. (1986). Long-term effects of preschool programs. In U. Neisser (Ed.), *The school achievement of minority children.* Hillsdale, NJ: Erlbaum.

DENNIS, W. (1973). *Children of the creche.* New York: Appleton-Century-Crofts.

DIETZ, K. R., LAVIGNE, J. V., AREND, R., & ROSENBAUM, D. (1997). Relation between intelligence and psychopathology among preschoolers. *Journal of Clinical Child Psychology, 26,* 99–107.

DORNBUSCH, S. M. (1994, February). *Off the track.* Paper presented at the Society for Research on Adolescence, San Diego, CA.

ELLIOTT, C. D. (1990). *Differential Ability Scales: Administration and scoring manual.* San Antonio, TX: The Psychological Corporation.

FAGAN, J. F. (1984). The intelligent infant: Theoretical implications. *Intelligence, 8,* 1–9.

FAGAN, J. F., & SINGER, L. T. (1983). Infant recognition memory as a measure of intelligence. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 2). Norwood, NJ: Ablex.

FIEDLER, F. E., & LINK, T. G. (1994). Leader intelligence, interpersonal stress, and task performance. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context.* New York: Cambridge University Press.

FLYNN, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

FLYNN, J. R. (1994). IQ gains over time. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1). New York: Macmillan.

FUNDUDIS, T., KOLVIN, I., & GARSIDE, R. F. (Eds.). (1979). *Speech retarded and deaf children: Their psychological development.* London: Academic Press.

GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic.

GLICK, J. (1968, February). *Cognitive style among the Kpelle of Liberia.* Paper presented at the meeting on Cross-Cultural Cognitive Studies, American Educational Research Association, Chicago.

GREENFIELD, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52,* 1115–1124.

GUDJONSSON, G. H. (1995). The Standard Progressive Matrices: Methodological problems associated with the administration of the 1992 adult standardization sample. *Personality and Individual Differences. 18,* 441–442.

HARKNESS, S., SUPER, C., & KEEFER, C. (1992). Culture and ethnicity. In M. D. Levine, W. B. Carey, & A. C. Crocker (Eds.), *Developmental-behavioral pediatrics* (2^{nd} ed). Philadelphia: Saunders.

HARTIGAN, J. A., & WIGDOR, A. K. (1989). *Fairness in employment testing.* Washington, DC: National Academy Press.

HAYDEN, D. C., FURLONG, M. J., & LINNEMEYER, S. (1988). A comparison of the Kaufman Assessment Battery for Children and the Stanford-Binet IV for the assessment of gifted children. *Psychology in the Schools, 25,* 239–243.

HEATH, S. B. (1983). *Ways with words.* New York: Cambridge University Press.

HECKMAN, J. J. (1995). Lessons from *The Bell Curve. Journal of Political Economy, 103,* 1091–1120.

HEDRICK, D., PRATHER, E., & TOBIN, A. (1984). *Sequences inventory of communication development* (rev. ed.). Seattle: University of Washington Press.

HENRY, B., CASPI, A., MOFFITT, T. E., & SILVA, P. A. (1996). Temperamental and familial predictors of violent and nonviolent criminal convictions: Age 3 to age 18. *Developmental Psychology, 32,* 614–623.

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

HIRSCHI, T., & HINDELANG, M. J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review, 42,* 571–587.

HONZIK, M. P., MACFARLANE, J. W., & ALLEN, L. (1948). The stability of mental test performance between two and eighteen years. *Journal of Experimental Education, 17,* 309–329.

HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253–270.

HORN, J. L., & NOLL, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues.* New York: Guilford.

HOWLIN, P., & RUTTER, M. (1987). The consequence of language delay for other aspects of development. In W. Yule & M. Rutter (Eds.), *Language development and disorders.* London: MacKeith.

HUNT, E. (1995). *Will we be smart enough?* New York: Russell Sage Foundation.

HUNTER, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory.* Hillsdale, NJ: Erlbaum.

HUNTER, J. E., & HUNTER, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–78.

JENCKS, C. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books.

JENSEN, A. R. (1998). *The g factor.* Westport, CT: Praeger.

JONES, H. E., & BAYLEY, N. (1941). The Berkeley Growth Study. *Child Development, 12,* 167–173.

KAO, G., & TRIENDA, M. (1995). *Educational aspirations of minority youth.* Unpublished manuscript.

KAUFMAN, A. S., & KAUFMAN, N. L. (1983). *Kaufman assessment battery for children: Interpretive manual.* Circle Pines, MN: American Guidance Service.

KOPP, C. B., & McCALL, R. B. (1982). Predicting later mental performance for normal, at-risk, and handicapped infants. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol 4). New York: Academic Press.

LAVE, J. (1988). *Cognition in practice.* New York: Cambridge University Press.

LIPSITT, L. P. (1992). Discussion: The Bayley scales of infant development: Issues of prediction and outcome revisited. *Advances in Infancy Research, 7,* 239–245.

LURIA, A. R. (1980). *Higher cortical functions in man* (2nd ed., rev.). New York: Basic Books.

LYNAM, D., MOFFITT, T., & STOUTHAMER-LOEBER, M. (1993). Explaining the relation between IQ and delinquency: Class, race, test motivation, school failure, or self-control? *Journal of Abnormal Psychology, 102,* 187–196.

LYON, G. R. (1996). Learning disabilities. *Future of Children, 6,* 54–76.

McCALL, R. B. (1972). Similarity in developmental profile among related pairs of human infants. *Science, 178,* 1004–1005.

McCALL, R. B. (1979). The development of intellectual functioning in infancy and the prediction of later IQ. In J. D. Osofsky (Ed.), *Handbook of infant development.* New York: Wiley.

McCALL, R. B., & CARRIGER, M. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64,* 57–79.

McCALL, V. W., YATES, B., HENDRICKS, S., TURNER, K., & MCNABB, B. (1989). Comparison between the Stanford-Binet: L-M and the Stanford-Binet: Fourth Edition with a group of gifted children. *Contemporary Educational Psychology, 14,* 93–96.

McGEE, R., PARTRIDGE, F., WILLIAMS, S., & SILVA, P. A. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child & Adolescent Psychiatry, 30,* 224–232.

McGREW, K. S., & KNOPIK, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. *School Psychology Review, 22,* 687–695.

MISERANDINO, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational Psychology, 88,* 203–214.

MOFFITT, T. E., GABRIELLI, W. F., & MEDNICK, S. A. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90,* 152–156.

MOFFITT, T. E., & SILVA, P. A. (1988). IQ and delinquency: A direct test of the differential detection hypothesis. *Journal of Abnormal Psychology, 97,* 330–333.

MORROW, J. D., & CAMP, B. W. (1996). Mastery motivation and temperament of 7-month-old infants. *Pediatric Nursing, 22,* 211–217.

MULLEN, E. M. (1995). *Mullen Scales of Early Learning.* Circle Pines, MN: American Guidance Service.

MYERS, D. G., & DIENER, E. (1995). Who is happy? *Psychological Science, 6,* 10–19.

NAGLIERI, J. A., & DAS, J. P. (1997). *Das-Naglieri Cognitive Assessment System.* Itasca, IL: Riverside.

NEISSER, U. (Ed.). (1998). *The rising curve.* Washington, DC: American Psychological Association.

NEISSER, U., BOODOO, G., 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.

NELLIS, L., & GRIDLEY, B. E. (1994). Review of the Bayley Scales of Infant Development-Second Edition. *Journal of School Psychology, 32,* 201–209.

NEYENS, L. G. J., & ALDENKAMP, A. P. (1996). Stability of cognitive measures in children of average ability. *Child Neuropsychology, 2,* 161–170.

NOVAK, P. A., TSUSHIMA, W. T., & TSUSHIMA, M. W. (1991). Predictive validity of two short-forms of the WPPSI: A 3-year follow-up study. *Journal of Clinical Psychology, 47,* 698–702.

NUÑES, T. (1994). Street intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 2). New York: Macmillan.

NUÑES, T., SCHLIEMANN, A.D., & CARRAHER, D. W. (1993). *Street mathematics and school mathematics.* New York: Cambridge University Press.

OKAGAKI, L., & STERNBERG, R. J. (1991). Cultural and parental influences on cognitive development. In L. Okagaki & R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking.* Hillsdale, NJ: Erlbaum.

OKAGAKI, L., & STERNBERG, R. J. (1993). Parental beliefs and children's school performance. *Child Development, 64,* 36–56.

PERKINS, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence.* New York: The Free Press.

PERKINS, D. N., & GROTZER, T. (1997). Teaching intelligence. *American Psychologist, 52,* 1125–1133.

PHELPS, L. (1989). Comparison of scores for intellectually gifted students on the WISC-R and the fourth edition of the Stanford-Binet. *Psychology in the Schools, 26,* 125–129.

PHELPS, L., BELL, C. M., & SCOTT, M. J. (1988). Correlations between the Stanford-Binet: Fourth Edition and the WISC-R with a learning disabled population. *Psychology in the Schools, 25,* 380–382.

PIAGET, J. (1972). *The psychology of intelligence.* Totowa, NJ: Littlefield Adams.

PIANTA, R. C., & CALDWELL, C. B. (1990). Stability of externalizing symptoms from kindergarten to first grade and factors related to instability. *Development and Psychopathology, 2,* 247–258.

PIANTA, R. C., & CASTALDI, J. (1989). Stability of internalizing symptoms from kindergarten to first grade and factors related to instability. *Development and Psychopathology, 1,* 305–316.

PRENTICE, N. M., & KELLY, F. J. (1963). Intelligence and delinquency: A reconsideration. *Journal of Social Psychology, 60,* 327–337.

RAVEN, J. C. (1960). *Guide to the Standard Progressive Matrices.* London: H. K. Lewis.

RAVEN, J. C. (1965). *Guide to using the Coloured Progressive Matrices.* London: H. K. Lewis.

RAVEN, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement, 26,* 1–16.

RAVEN, J. C., COURT, J. H., & RAVEN, J. (1992). *Standard Progressive Matrices; 1992 Edition.* Oxford: Oxford Psychologists Press.

REE, M. J., CARRETTA, T. R., & TEACHOUT, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80,* 721–730.

REES, A. H., & PALMER, F. H. (1970). Factors related to change in mental test performance. *Developmental Psychology, 3*(2, Pt. 2), 1–57.

REHBERG, R. A., & ROSENTHAL, E. R. (1978). *Class and merit in the American high school: An assessment of the revisionist and meritocratic arguments.* New York: Longman.

REYNOLDS, C. R. (1987). Playing IQ roulette with the Stanford-Binet, 4th Edition. *Measurement and Evaluation in Counseling and Development, 21,* 52–63.

RICHMAN, N., STEVENSON, J., & GRAHAM, P. (1983). The relationship between language, development, and behavior. In M. H. Schmidt & H. Remschmidt (Eds.), *Epidemiological approaches in child psychiatry II.* New York: Thieme-Stratton.

ROSE, S. A. (1994). Relation between physical growth and information processing in infants born in India. *Child Development, 65,* 889–902.

ROSE, S. A., & FELDMAN, J. F. (1987). Infant visual attention: Stability of individual differences from 6 to 8 months. *Developmental Psychology, 23,* 490–498.

RUTTER, M. (1987). Temperament, personality, and personality disorder. *British Journal of Psychiatry, 150,* 443–458.

RUTTER, M. (1996, April). *Profound early deprivation and later social relationship in early adoptees from Romanian orphanage followed at age 4.* Paper presented at 10th Biennial International Conference on Infant Studies, Providence, RI.

SACCUZZO, D. P., & JOHNSON, N. F. (1995). Traditional psychometric tests and proportionate representation: An intervention and program evaluation study. *Psychological Assessment, 7,* 181–194.

SCHMIDT, F. L. (1994). The future of personnel selection in the U.S. Army. In M. Rumsey, C. Walker, & J. Harris (Eds.), *Personnel selection and classification.* Hillsdale, NJ: Erlbaum.

SCHMIDT, F. L., & HUNTER, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge, *Current Directions in Psychological Science, 1,* 8–9.

SCHMIDT, F. L, & HUNTER, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

SCHMIDT, F. L., ONES, D. S., & HUNTER, J. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

SCHMIDT, W. H. O. (1996, December 10). Letters to the editor: New police test will be a disaster. *The Wall Street Journal,* p. A23.

SCHUERGER, J. M., & WITT, A. C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology, 45,* 294–302.

SERPELL, R. (1993). *The significance of schooling: Life-journeys in an African society.* New York: Cambridge University Press.

SHINER, R. L. (1998). How shall we speak of children's personalities in middle childhood? A preliminary taxonomy. *Psychological Bulletin, 124,* 308–332.

SHORT, J. F., & STRODTBECK, F. L. (1965). *Groups process and gang delinquency.* Chicago: Chicago University Press.

SIGMAN, M., COHEN, S. E., BECKWITH, L., ASARNOW, R., & PARMELEE, A. H. (1991). Continuity in cognitive abilities from infancy to 12 years of age. *Cognitive Development, 6,* 47–57.

SILVA, P. A., MCGEE, R. O., & WILLIAMS, S. M. (1983). Developmental language delay from three to seven years and its significance for low intelligence and reading difficulties at age seven. *Developmental Medicine and Child Neurology, 25,* 783–793.

SLOMKOWSKI, C. L., NELSON, K., DUNN, J., & PLOMIN, R. (1992). Temperament and language: Relations from toddlerhood to middle childhood. *Developmental Psychology, 28,* 1090–1095.

SONTAG, L. W., BAKER, C. T., & NELSON, V. L. (1958). Mental growth and personality development: A longitudinal study. *Monographs of the Society for Research in Child Development, 23,* no. 2 (Whole No. 68).

SPEARMAN, C. (1923). *The nature of intelligence and the principles of cognition.* New York: Macmillan.

SPEARMAN, C. (1927). *The abilities of man.* New York: Macmillan.

STANKOV, L. (1986). Kvaschev's experiment: Can we boost intelligence? *Intelligence, 10,* 209–230.

STATTIN, H., & MAGNUSSON, D. (1990). *Pubertal maturation in female development.* Hillsdale, NJ: Erlbaum.

STERNBERG, R. J. (Ed.). (1982). *Handbook of human intelligence.* New York: Cambridge Univerisity Press.

STERNBERG, R. J. (Ed.). (1994). *Encyclopedia of human intelligence.* New York: Macmillan.

STERNBERG, R. J. (1997). *Successful intelligence.* New York: Plume.

STERNBERG, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 11–20.

STERNBERG, R. J. (Ed.). (2000). *Handbook of intelligence.* New York: Cambridge University Press.

STERNBERG, R. J., FERRARI, M., CLINKENBEARD, P. R., & GRIGORENKO, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly, 40,* 129–137.

STERNBERG, R. J., FORSYTHE, G. B., HEDLUND, J., HORVATH, J., SNOOK, S. WILLIAMS, W. M., WAGNER, R. K., GRIGORENKO, E. L. (2000). *Practical intelligence in everyday life.* New York: Cambridge University Press.

STERNBERG, R. J., & GRIGORENKO, E. L. (1997). The cognitive costs of physical and mental ill health: Applying the psychology of the developed world to the problems of the developing world. *Eye on Psi Chi, 2,* 20–27.

STERNBERG, R. J., GRIGORENKO, E. L., & BUNDY, D. A. (1997). *Measuring Human Development: The role of general indices of cognitive ability in assessing the impact of education and HNP interventions on child development.* Washington, DC: World Bank.

STERNBERG, R. J., GRIGORENKO, E. L., FERRARI, M., & CLINKENBEARD, P. C. (1999). A triarchic analysis of an aptitude interaction. *European Journal of Psychological Assessment, 15,* 3–13.

STERNBERG, R. J., & KAUFMAN, J. C. (1998). Human abilities. *Annual Reviews, 49,* 479–502.

STERNBERG, R. J., NOKES, K., GEISSLER, P. W., PRINCE, R., OKATCHA, F., BUNDY, D. A., & GRIGORENKO, E. L. (in press). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence.*

STERNBERG, R. J., OKAGAKI, L., & JACKSON, A. (1990). Practical intelligence for success in school. *Educational Leadership, 48,* 35–39.

STERNBERG, R. J., TORFF, B., & GRIGORENKO, E. L. (1998). Teaching for successful intelligence raises school achievement. *Phi Delta Kappan, 79,* 667–669.

STERNBERG, R. J., & WAGNER, R. K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world.* New York: Cambridge University Press.

STERNBERG, R. J., WAGNER, R. K., WILLIAMS, W. M., & HORVATH, J. (1995). Testing common sense. *American Psychologist, 50,* 912–927.

STOTT, L. H., & BALL, R. S. (1965). Infant and preschool mental tests: Review and evaluation. *Monographs of the Society for Research in Child Development, 30*(3, Serial No. 101).

THOMAS, A., & CHESS, S. (1977). *Temperament and development.* New York: Bruner.

THORNDIKE, R., HAGEN, E., & SATTLER, J. R. (1986a). *Guide for administering and scoring the Stanford-Binet Intelligence Scale* (4th ed., Pt. 2). Chicago: Riverside.

THORNDIKE, R., HAGEN, E., & SATTLER, J. R. (1986b). *Technical manual: Stanford-Binet Intelligence Scale* (4th ed). Chicago: Riverside.

VYGOTSKY, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

WAGNER, R. K. (1997). Intelligence, training, and employment. *American Psychologist, 52,* 1059–1069.

WECHSLER, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence* (rev.). San Antonio, TX: The Psychological Corporation.

WECHSLER, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-III.* San Antonio: The Psychological Corporation.

WEST, D. J., & FARRINGTON, D. P. (1973). *Who becomes delinquent?* London: Heinemann.

WHITE, J. L., MOFFITT, T. E., & SILVA, P. A. (1989). A prospective replication of the protective effects of IQ in subjects at high risk for juvenile delinquency. *Journal of Consulting & Clinical Psychology, 57,* 719–724.

WIGDOR, A. K., & GARNER, W. P. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies* (Pt. 1). Washington, DC: National Academy Press.

WIGDOR, A. K., & GREEN, B. F., JR. (1991). *Performance assessment in the workplace* (Vol. 1). Washington, DC: National Academy Press.

WILSON, J. Q., & HERRNSTEIN, R. J. (1985). *Crime and human nature.* New York: Simon & Schuster.

WILSON, R. S. (1972). Twins: Early mental development. *Science, 175,* 914–917.

WOBER, M. (1974). Towards an understanding of the Kiganda concept of intelligence. In J. W. Berry & P. R. Dasen (Eds.), *Culture and cognition.* London: Methuen.

WOLFGANG, M. E., FIGLIO, R. M., & SELLIN, T. (1972). *Delinquency in a birth cohort.* Chicago: University of Chicago Press.

WOODCOCK, R. W., & JOHNSON, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

YANG, S., & STERNBERG, R. J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence, 25,* 21–36.

ZIMMERMAN, B. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies.* New York: Cambridge University Press.

ZIMMERMAN, I. E., STEINER, V. G., & POND, R. E. (1992). *Preschool Language Scale-3.* San Antonio, TX: Psychological Corporation.