

More than just test scores

Henry M. Levin

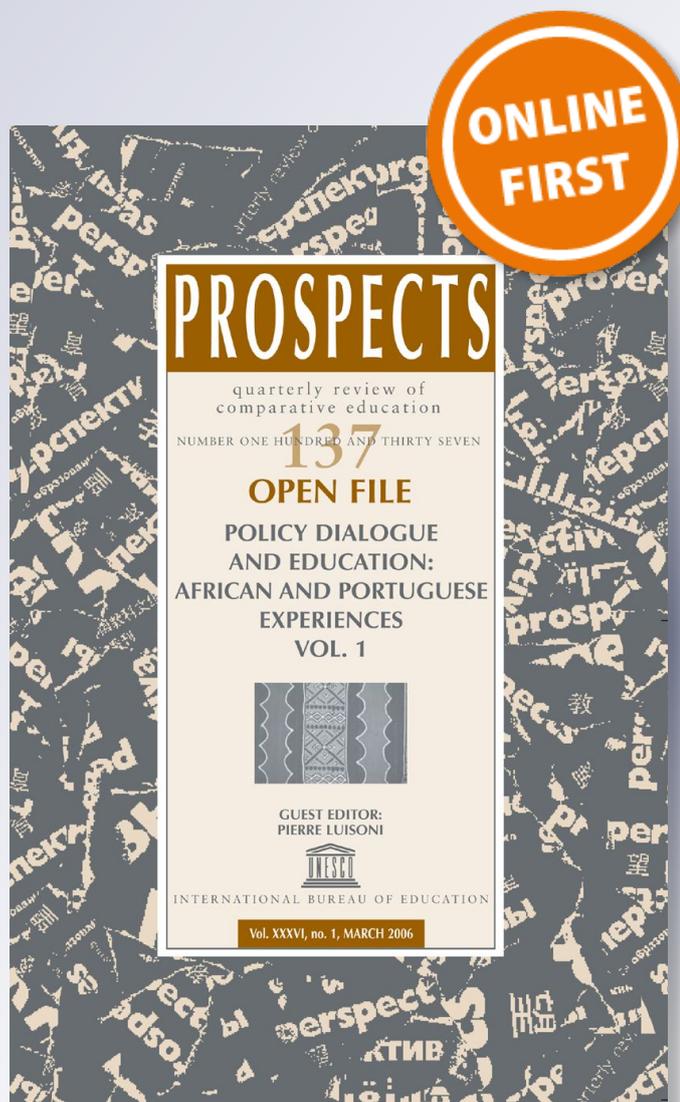
PROSPECTS

Quarterly Review of Comparative
Education

ISSN 0033-1538

Prospects

DOI 10.1007/s11125-012-9240-z



Your article is protected by copyright and all rights are held exclusively by UNESCO IBE. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

More than just test scores

Henry M. Levin

© UNESCO IBE 2012

Abstract Around the world we hear considerable talk about creating world-class schools. Usually the term refers to schools whose students get very high scores on the international comparisons of student achievement such as PISA or TIMSS. The practice of restricting the meaning of exemplary schools to the narrow criterion of achievement scores is usually premised on the view that test scores are closely linked to the provision of a capable labour force and competitive economy. In fact, the measured relationships between test scores and earnings or productivity are modest and explain a relatively small share of the larger link between educational attainment and economic outcomes. What has been omitted from such narrow assessments are the effects that education has on the development of interpersonal and intrapersonal skills and capabilities that affect the quality and productivity of the labour force. This article provides evidence on some of these relationships, on the degree to which the non-cognitive effects of schooling contribute to adult performance, and on the evidence that deliberate school interventions can influence non-cognitive outcomes. It concludes with the view that the quest for world-class schools must encompass a range of human development characteristics that extend considerably beyond test scores.

Keywords World-class education · Non-cognitive · Cognitive · Test scores · Interpersonal and intrapersonal characteristics · Socio-emotional learning · Economic productivity

Among countries around the globe, the term world-class education has become a common way to express educational aspirations. Precisely what “world-class” means has not been fully explored, but the slogan is widely used as a descriptor of what is sought. Underlying this term are two tacit or explicit assumptions. The first is that a world-class educational system is a prerequisite for the economic vigour and competitiveness of a nation or region. The concern about world-class education tends to be focused less on its importance for civic behaviour or cultural leadership than on its impact on future economic viability. The

H. M. Levin (✉)
Teachers College, Columbia University, 525 W. 120th Street, New York, NY 10027, USA
e-mail: HL361@columbia.edu

second assumption is that world-class education is largely measured by the high test scores of students in one country relative to those in other countries. That is, in evaluating world-class performance, those countries and schools whose students earn the highest scores on common achievement tests set the benchmarks for other countries. In this respect, a world-class educational system is judged strictly by measures of cognitive achievement, rather than on any of the other types of human development that schools produce.

As evidence, one needs only to consider the discussions that arise as reports and country rankings are published from the periodic surveys of the Program of International Student Achievement (PISA), the International Assessment of Literacy Survey (IALS), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). Countries take their rankings very seriously, and the media either praise their country's performance or decry it, calling for major educational reforms. At the same time, national and regional assessments compare different regions and educational entities on the quality of their educational systems, primarily using the metrics of student achievement as their guides. And these reviews of student achievement are used to draw conclusions about the quality of educational systems and the implications for the future workforce, considering little or no additional data on other attributes of school or student performances beyond the narrow range of test results.

In this article I argue that basing world-class standards on measures of student achievement is a very limited approach to evaluating an educational system. To meet the economic, political, social, and personal demand for competency, much more is required of students and adults than just cognitive proficiencies as measured by test scores. Individuals must develop interpersonal skills that enable them to relate to others in many different societal situations. They must also develop the intrapersonal skills that include good judgment and strategies for meeting their own needs in effective ways. I will suggest that a world-class educational system must consider these dimensions as well as cognitive ones, even though the latter have a much longer history of being identified with the measurement of proficiencies.

In what follows, I suggest some reasons why the focus has been almost exclusively devoted to the cognitive dimensions as measured by student testing rather than to what are broadly called the non-cognitive or socio-emotional outcomes of schools. I then provide a wider perspective and evidence that other human characteristics that are partially or primarily developed by schools carry considerable importance in guiding human and social performance and especially economic productivity. More detailed discussion of the evidence supporting these claims is found in Levin (2012). Finally, I suggest ways of incorporating these dimensions into international and inter-regional comparisons of educational systems.

Why have test scores been the primary focus?

When the field of human capital was developed in the middle of the twentieth century, it was predicated heavily on the close relationship between education and earnings. Presumably, persons with more education had more skills that were valued in labour markets, which contributed to their productivity and income. Human capital embodied in education was not limited just to the cognitive measures that were reflected in test scores. In fact, in his pioneering work on the subject, Gary Becker (1964, p. 20) interpreted the concept of human capital very broadly, to include knowledge, skills, values, and habits. Clearly, the skills that are valued in the workplace exceed those that are strictly cognitive and

measurable by standardized achievement tests. They include both attributes that are cognitive in nature, but not measured by most of the simplified testing formats, as well as so-called non-cognitive measures. For example, tests rarely assess the capacity to formulate and solve problems, or valuable interpersonal behaviours such as collaboration, listening skills, and the ability to communicate, and intrapersonal behaviours such as time management and impulse control.

This is not to dismiss measures of student achievement, but they should be viewed as one of many indicators of the potential productivity of a school system rather than the only important outcome. Many of the achievement domains that are tested have value in assessing students' present knowledge and predicting to some extent their further learning and future success in employment and civic roles, as well as their personal needs. But I will argue that conventional test scores have much less power than is commonly assumed to predict performance in these roles, and particularly to predict economic productivity. Perhaps the reason many economists narrowed the application of human capital to cognitive measures and test scores was the early finding that educational attainments (more years of schooling) produce higher income. Since a competitive economy pays workers according to their productivity, it was easy to assume that the higher test scores of those with more education were the key to productivity. The financial return to human capital and for society could be calculated as a return on investment; this was tacitly assumed to result from how much more a student learned in the cognitive domain, of which test scores were the measure of output.

Prior to the establishment of human capital theory, schools were commonly seen as having profound impacts on attitudes and behaviours as well as on those skills that tests could measure. For example, Alex Inkeles (1966) and other social psychologists emphasized the role of schools in creating competent adults; competence extended into a wide variety of personality characteristics that people needed to function successfully, both socially and individually. Few of these were measurable by standardized tests. Inkeles and Smith (1974) developed attitudinal measures that would predict successful performance in an industrial workplace and used these measures to assess "modernization" in six countries. They found that educational attainment was, by far, the most important predictor of the modern attitudes needed to function successfully in the workplace (Inkeles 1975). In his classic book, *What is Learned in School*, Dreeben (1968) focused on the specific characteristics of schools that differ from those of families and on why schools and families emphasize different values and interpersonal skills for creating competent adults.

But much of what schools actually do was ignored, as both national educational systems and international comparisons of educational systems focused almost exclusively on test results and omitted other important aspects that might impart value to educational development and outcomes. A good portion of what makes cognitive test scores an attractive way to assess schools is the field's relatively advanced development. A small sample of students' test performances can be obtained at low cost, and is believed to have predictive power for further education, occupational success, and earnings. These forms of testing have gone through more than a century of development and are highly sophisticated. In contrast, the specific non-cognitive or personality attributes required for successful adulthood are more diffuse and more contested and have not yielded to the straightforward measurement methods used for standardized tests. There is simply no global agreement on what is of consequence beyond student achievement and how it should be measured. For these reasons, and perhaps others, discussions of world-class education and educational systems have been limited to student achievement.

Evidence on the importance of cognitive and non-cognitive attributes

One way to examine whether achievement scores are given too much weight and whether the values, attitudes, and socio-emotional outcomes of schools are given too little is to review the importance of cognitive achievement in predicting important social outcomes. One of the main reasons for the obsession with test scores is the assumption that a nation's economic productivity and national economic success are largely determined by the test performances of its children. Certainly, there is some evidence that test scores are related to productivity and income, but it is important to ask how powerful that relationship is and whether other school outcomes also have important bearings on educational outcomes.

Do test scores explain earnings?

In a competitive economy earnings are based on worker productivity, so many studies have attempted to measure the determinants of earnings (Card 1999). In particular, the relationship between education and earnings has been measured statistically by relating educational attainments and work experience to earnings, using statistical controls for demographic and other pertinent variables. Almost invariably, a large expected increase in earnings is associated with additional years of education, about 10% for each year in industrialized countries. It has typically been assumed that this economic return is associated with the cognitive skills learned in school, rather than with other forms of learning, including non-cognitive ones.

Bowles and Gintis (1976) decided to add test scores to their statistical analysis to see how much they determined the returns to years of schooling. Presumably, test scores should have a closer relationship to skills and productivity than simply accounting for the time spent in an educational institution. However, a review of 25 studies in which researchers included test scores in their statistical analysis found that the impact of schooling on earnings retains more than 80% of its estimated effect on earnings, even after accounting for student achievement (Bowles, Gintis, and Osborne 2001). This finding suggests that test scores account for only a minority fraction of the impact of educational attainment on earnings, and that other important impacts of education must be accounted for.

A second prevalent assumption that supports the intense focus on test scores as a criterion for world-class education is the view that in an information economy, test scores have an increasing influence on productivity and earnings. This view is supported by one influential study that compared the change in impact of mathematics scores on earnings between 1978 and 1986 and found an increasing impact (Murnane, Willett, and Levy 1995). But subsequent reviews of 24 similar studies over a much longer period found no rising effect of test scores and relatively small relationships between test scores and earnings overall (Bowles, Gintis, and Osborne 2001). A study in the UK based on changes in labour markets between 1995 and 2004 also found no evidence of increases in the importance of cognitive skills on earnings (Vignoles, De Coulon, and Marcenaro-Gutierrez 2011). Support for the rising impact of cognitive skills is absent, or mixed at best, and is beset with methodological issues (Cawley, Heckman, and Vytlačil 2001). Overall evidence of the weak relation between cognitive test scores and earnings is also evident in their low explanatory value in statistical studies (e.g., Murnane, Willett, Bratz, and Duhaldeborde 2001; Murnane, Willett, Duhaldeborde, and Tyler 2000).

A third assumption about the importance of tests is that they are useful for selecting employees; advocates claim that they have high predictive validity for worker productivity. In the late 1980 s, the states affiliated with the U.S. Employment Service were using the

General Aptitude Test Battery (GATB) to refer prospective applicants to jobs on the basis of test scores. Because of questions about this practice, and particularly the separation of results by race, a National Research Council (Hartigan and Wigdor 1989) panel of distinguished statisticians and other scholars was convened to assess the evidence on the relationship between this test and worker productivity. Not only did this study find that the link between the test scores and productivity (as judged by supervisors) was vastly overstated; the test was also found to explain only about 6% of the variance in productivity, leaving 94% to be explained by factors other than the test result. This finding was substantiated in a comprehensive survey of the general literature on employment testing (Sackett, Schmitt, Ellingson, and Kabin 2001).

Do non-cognitive measures explain productivity?

Thus far I have asserted that the impact that test scores have on economic outcomes tends to be viewed as almost the exclusive explanation for the impact that schools have on economic productivity and national competitiveness. Certainly, students performing below a reasonable test score threshold may find it difficult to gain and hold employment and benefit from training, and reducing these deficiencies should be an important focus in an educational system that aspires to a world-class rating. But the workplace evidence suggests that other aspects of school outcomes are also at play and need to be considered.

For example, many surveys of employers have tried to ascertain what is lacking in their workers. The usual expectation is that their cognitive preparation, such as reading and mathematics, is below the level they need to be productive workers, and that these dimensions are reflected on achievement tests. But, perhaps surprisingly, almost all employer surveys place such academic deficiencies well down their lists of concerns, which focus instead on appropriate work attitudes and behaviour such as self-discipline, punctuality, attendance, setting of goals, taking responsibility, and listening skills (National Research Council 1984; Zemsky and Iannozzi 1995). These skills are not measured by standardized tests. In a recent survey in the UK (Shury, Winterbotham, Davies, and Oldfield 2010), employers said they wanted their employees to have certain job-specific skills, which were based more on work experience than academic performance. They were also concerned with other attributes their employees lacked: motivation, customer-handling skills, problem-solving skills, and teamwork. Skills in numeracy and literacy ranked well below these areas in employer concerns.

A second example of the implicit evidence that something about schools beyond cognitive achievement and test scores is affecting economic outcomes is found in specific studies of school interventions that have modest effects on achievement scores, but substantial effects on educational and life outcomes. The Perry Preschool study was an experimental attempt to provide quality preschool experiences to an inner-city population; children were randomly assigned to either the Perry Preschool or a control group. Both groups were followed to the age of 40 to determine the apparent effects of participation in the preschool intervention (Belfield, Nores, Barnett, and Schweinhart 2006; Schweinhart 2010). Despite their relatively small advantages in achievement, the group that received the intervention completed more years of education, had less criminal involvement, needed less public assistance, and had higher levels of earnings and employment, generating a large social return (Heckman, Moon, Pinto, Savelyev, and Yavitz 2010). Their earnings were about one third higher, and their criminal convictions were at half the level of those in the control group. These differences were far greater than could be predicted by achievement differences.

The most important experimental study in education in the United States addressed the impact of reduction in class size on achievement (Finn and Achilles 1990; Mosteller 1995). This study showed similar long-range patterns on education that could not be explained by test scores alone. The Tennessee Class Size experiment randomly assigned students in grades kindergarten to three, across the state, into regular-size classes of 23 to 25 students or to small classes of 13 to 17 students. The study found achievement advantages in reading and mathematics for those in smaller classes, advantages that were sustained in later grades. But what was most surprising was the substantial difference in high school graduation rates almost a decade later. Compared to similar students who had attended regular classes, disadvantaged students who were in smaller classes for four years had graduation rates 18 percentage points higher: 88% compared to 70% (Finn, Gerber, and Boyd-Zaharias 2005). This was found to be well beyond the predictive effect of the achievement gains, suggesting that non-cognitive effects accounted for a large portion of the better graduation performance. A more recent study of reduced class sizes also found improved student learning behaviour as a consequence (Dee and West 2011).

Higher graduation rates can be linked directly to economic participation and performance. A recent study in the United States found that high school graduates earned a substantial amount more than similar high school dropouts, with differences by race and gender (Levin 2009). For example, both white and black males earned \$300,000 more over the course of their lifetimes if they graduated from high school rather than dropping out. (This figure is the present value at age 20 of the additional lifetime income). This additional lifetime income generated a present value of about \$140,000 in additional tax revenues. In addition, taxpayers saved about \$41,000 in public health costs, \$27,000 in reductions in the public costs of crime, and other savings in public assistance. Among five proven educational interventions for reducing dropouts, the benefits to the taxpayer alone averaged about \$209,000 in present value at age 20 for each additional graduate who would have otherwise dropped out. Thus, the benefits to the taxpayer alone averaged about \$209,000 in present value at age 20 for each additional graduate who would otherwise have dropped out.

There is also good evidence that high school completion improves non-cognitive outcomes for students in a way that improves their labour market prospects independently of their test performance. A study of students in the United States who dropped out of high school and took the GED exam to establish high school equivalency showed they had earnings patterns similar to those of dropouts rather than high school completers (Heckman and Rubinstein 2001). Although their test scores were about the same as those of high school graduates who did not pursue post-secondary education, their earning patterns were vastly inferior. The authors concluded that GED recipients have less of the non-cognitive skills desired by employers, such as persistence and commitment, attributes that are more likely to be found among high school graduates. As one indicator of the non-cognitive difference, the GED students were found to engage in higher rates of illicit activities than non-GED dropouts or high school graduates.

Can schools improve non-cognitive outcomes?

One advantage of the focus on test scores is that such tests are relatively low-cost and easy to implement, at least for those skills that the tests normally assess. And, presumably, schools can improve those scores with better teachers, curriculum, and instructional materials. In contrast, we lack a clear understanding of which interpersonal and intrapersonal attributes are key to productive human development, how to measure them

accurately and at reasonable cost, and how to help schools improve them. All of these limitations are good excuses not to include them in evaluating the quality of schools and school improvement.

However, recent evidence suggests that we can identify some of these promising non-cognitive characteristics and can improve them. One dimension that has been identified is that of self-regulation or executive function (EF), the ability to plan and monitor behaviour rather than to respond to impulse. Workplaces require the former and can be disrupted by the latter. To instill executive function in young children who might otherwise not gain this trait outside of the school, a curriculum called Tools of the Mind has been developed for preschool children. EF has been more strongly linked to school readiness than cognitive measures (Blair and Razza 2007). A recent experimental study confirms that in developing EF, the Tools of the Mind intervention improves children's social development, as well as their classroom experience (Barnett et al. 2008).

Overall summaries of the literature also confirm the importance of early childhood interventions on behavioural or socio-emotional change. Nores and Barnett (2010) reviewed a total of 38 studies evaluating 30 interventions in 23 countries that had utilized quasi-experimental or random assignment designs. They considered the type of intervention, sample size, study design and duration, country, target group, subpopulations, and dosage of interventions. They found both cognitive and behavioural benefits. Camilli, Vargas, Ryan, and Barnett (2010) undertook a meta-analysis of 123 comparative studies of early childhood interventions. All the programme evaluations that they reviewed had been designed using experimental principles. Although the largest effects were found for cognitive outcomes, preschool experience was also found to be associated with students' social skills and school progress. Important findings on these topics have also been found in the empirical studies of Duncan et al. (2007) and Duncan and Magnuson (2011).

The most extensive evaluation of teaching social and emotional skills and their impact is the meta-analysis (statistical summary) by Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011). This work is based upon a statistical survey of 213 school-based social and emotional learning (SEL) programmes from kindergarten through high school. Overall, the studies encompassed 270,000 children aged 5 to 18. The research team included only intervention studies that had control groups. They developed six categories of outcomes.

Social and emotional skills includes different types of cognitive, affective, and social skills related to such areas as identifying emotions from social cues, goal setting, perspective taking, interpersonal problem solving, conflict resolution, and decision making.

Attitudes toward self and others includes positive attitudes about the self, school, and social topics including self-perceptions (e.g., self-esteem, self-concept, and self-efficacy), school bonding (e.g., attitudes toward school and teachers), and conventional (e.g., pro-social) beliefs about violence, helping others, social justice, and drug use.

Positive social behaviour includes outcomes on getting along with others. These are based on data derived from the student, teacher, parent, or an independent observer, on the basis of daily behaviour, as opposed to hypothetical situations.

Conduct problems includes different types of behaviour problems, such as disruptive class behaviour, noncompliance, aggression, bullying, school suspensions, and delinquent acts.

Emotional distress includes internalized mental health issues, such as depression, anxiety, stress, or social withdrawal. Reports on these items were provided by students, teachers, or parents.

Academic performance includes standardized reading or math achievement test scores from such instruments as the Stanford Achievement Test and the Iowa Test of Basic Skills,

along with school grades in the form of students' overall grade point average (GPA) or their grades in specific subjects (usually reading or math). Only data drawn from school records were included.

The effects of the interventions to improve social and emotional skills were comparable to or exceeded the results found in the literature for improving student achievement. In terms of effect sizes (portions of a standard deviation), the results for the six outcomes were as follows: social and emotional skills, .57; attitudes, .23; positive social behaviour, .24; student conduct problems, .22; emotional distress, .24; and academic performance, .27. Of the academic performance studies, 33 had follow-up evaluations at least six months after the intervention ended, with a median follow-up time of about one calendar year. Effect sizes for the sub-group of longer-range studies remained at statistically significant levels with the effect size for academic performance at .32. This suggests that interventions to develop social and emotional skills have particular salience for improving student achievement.

A reasonable conclusion to be drawn from this literature is that non-cognitive skills can be taught through purposive interventions and that they can make a difference for many valuable social/behavioural outcomes and for student achievement. The latter is an important conclusion; not only are these outcomes important in themselves, but they also appear to have a positive impact on raising achievement. Durlak et al. (2011) found that the average effect size among the socio-emotional interventions on achievement is adequate to raise standardized student achievement scores by 11 percentile points. This is equivalent to about a 30-point increase in PISA scores: the difference between average scores for the United States and Canada. Or, it would raise the U.S. ranking from average among the industrialized countries to among the top performers. While this may not be a simple matter of policy, it does provide a framework for considering the potential of non-cognitive interventions as a strategy to raise cognitive results at the same time that they improve interpersonal and intrapersonal outcomes.

Evidence of non-cognitive impacts on worker productivity

Given the broad preoccupation with establishing and maintaining world-class educational systems that can promote highly productive and competitive economies, it is useful to consider some of the emerging economic studies that link non-cognitive outcomes to productivity and earnings. This research is becoming more important, but it is still limited relative to what is needed.

Without question, the scholar who has done the most to develop an understanding of the role that non-cognitive skills play in educational and economic outcomes is James Heckman, along with his colleagues from the University of Chicago. Many of their most valuable studies are cited by Borghans, Duckworth, Heckman, and Weel (2008). Heckman's role is also central to the content of the Symposium on "The Noncognitive Determination of Labor Market and Behavioral Outcomes", a special issue of *The Journal of Human Resources* (Kniesner and ter Weel 2008); I cite many articles from that issue here. Heckman has not only called attention to the importance of non-cognitive skills, but has also worked with psychologists and neurologists to estimate which periods of childhood are optimal for investments in developing the different types of skills and their later impact on labour market returns (Knudsen, Heckman, Cameron, and Shonkoff 2006). His masterful study with Flavio Cunha (Cunha and Heckman 2008) is considered to be the

most ambitious and sophisticated attempt in two directions. First, they formulated a theory of optimal investment between cognitive and non-cognitive skills from birth to labour force entry; then, they applied that model to a specific longitudinal data set to measure the impact of cognitive and non-cognitive skill development on earnings. Drawing on their data set, they created a battery of non-cognitive scores, focusing particularly on an anti-social construct using student anxiety, headstrongness, hyperactivity, and peer conflict to go along with the cognitive test scores in this analysis. Based on the psychological, neurological, social, and other aspects of child development, they modeled the developmental path and estimated the impact that investments in developing both cognitive and non-cognitive skills would have on high school graduation and earnings (at age 23) at three different periods during the span from age 6 to 13. They found that, as children age, the impact of an investment return shifts markedly, from cognitive skills at the earlier ages (6 and 7 to 8 and 9) to non-cognitive skills during the later period.

Clearly, this analysis, if it stands up to replication, has profound implications for school policy and the construction of educational programmes. The work of Heckman and his students creates a benchmark for considering the optimal mix of interventions and policy implications that can enhance human development through a combination of appropriate strategies for developing both cognitive and non-cognitive skills. Moreover, many of the assumptions in this work seem to correspond with Sameroff's (2010) attempt to create a unified theory of child development; this suggests that the leading edge in all this research is moving in a similar direction. As with Heckman's programme, Sameroff has developed a conceptual approach that connects the individuals and their contexts in a dynamic way.

In an intriguing study in Sweden, Lindqvist and Vestman (2011) evaluated cognitive and non-cognitive dimensions of military enlistees; enlistment was mandatory for all Swedish males. All enlistees filled out an extensive questionnaire of 70 to 80 questions. A certified psychologist was given this information as well as measures of cognitive ability and other attributes. Following a specified set of procedures, the psychologist interviewed each enlistee and evaluated his perceived ability to cope with the psychological requirements of military service. Each enlistee was given a score according to the same distribution used for the cognitive ability score. Using a random sample of men born from 1965 to 1984, the authors evaluated the impact of cognitive and non-cognitive skills on wages, unemployment, and annual earnings. They found that the men who do poorly in the labour market are those who lack non-cognitive abilities. In contrast, cognitive ability is a stronger predictor of wages and earnings for the workers whose earnings are above the median. In Germany, Heineck and Anger (2010) also found that a combination of cognitive and personality measures were related to earnings.

Perhaps the best single source on the role of non-cognitive skills and the economy is the 2008 Symposium mentioned earlier. The unusually focused volume that resulted from it contains an article by Borghans, ter Weel, and Weinberg (2008), which analyses the tradeoffs in caring and directness in jobs that have different interpersonal requirements. Caring requires cooperation, whereas directness requires clear communication. The authors found that the returns to these attributes depend upon relative supply and demand: returns to these roles, which different individuals hold in different combinations, match their assignment models. Other articles in the volume, by Fortin (2008), Krueger and Schkade (2008), Segal (2008), and Urzua (2008), address other labour market consequences related to non-cognitive skills and the roles of workers, as well as the impacts of students' non-cognitive skills. Another useful resource is the presentations at a recent IZA (2011) workshop.

Specific non-cognitive characteristics

So many concepts, constructs, and names have been developed for the personality and social and behavioural characteristics that are referred to as non-cognitive that I will not allocate much space to listing or categorizing them. A major challenge is identifying those that are most important for predicting adult competencies and performance and finding the ways that schools and the educational system contribute to producing them. The most comprehensive analysis of personality and its roles in labour markets, health, crime, and civic behaviour is that of Almlund, Duckworth, Heckman and Kautz (2011), who attempted, ambitiously, to map personality traits into economic modelling. It is important, however, to provide at least a glimpse of how those traits have been referred to and used in the psychological literature.

For at least the last two decades, the five-factor model of personality has been used to relate non-cognitive skills to academic achievement, educational attainment, and other outcomes. Over time, independent researchers accumulated different hypotheses and empirical studies and used them to create dimensions for statistical factor analysis (Digman 1990). They consolidated many different dimensions of personality into a five factor model to find a basic structure for what was a highly disorganized and idiosyncratic set of measures and constructs. Accordingly, the so-called Big Five factors have been considered to constitute the basic structure that underlies all personality traits and that integrates a variety of findings in personality psychology. The Big Five factors are:

- Openness—the person is inventive and curious as opposed to consistent and cautious.
- Conscientiousness—the person is efficient and organized as opposed to easy-going and careless.
- Extraversion—the person is outgoing and energetic as opposed to solitary and reserved.
- Agreeableness—the person is friendly and compassionate as opposed to cold and unkind.
- Neuroticism—the person is sensitive and nervous as opposed to secure and confident.

Many researchers have used these categories to predict behaviour; they are prominent in the massive review by Almlund et al. (2011). For example, looking at four samples of university students, Nofle and Robins (2007) found that Conscientiousness predicted grade-point average and that Openness predicted SAT verbal score. Using the Big Five to measure workplace productivity, Neuman and Wright (1999) studied the relationship between the personality characteristics of human resource representatives at local units of a large wholesale department store enterprise. They found that Agreeableness and Conscientiousness predicted the peer ratings of team member performance after controlling for job-specific skills and general cognitive ability.

A team at the Research Division of the Educational Testing Service (Kyllonen, Lipnevich, Burrus, and Roberts 2008) has been further developing non-cognitive constructs and measures. They focus on both personality characteristics and motivation, reviewing studies that link these to educational outcomes. They also consider various measurement approaches and document particular interventions in developing personality facets that lead to higher achievement and productivity.

Implications for world-class education assessment

Modern societies demand much of their members, and competence in meeting these demands must be a high social priority. Among all of the vehicles for socializing the

young, schools are a very powerful one, because students spend considerable time there and schools have specific functions in preparing young people for adulthood. Clearly, knowledge and cognitive functioning are an important goal of schools and provide crucial skills for creating productive workers and citizens. But non-cognitive or behavioural and social skills and attitudes are also crucial. Even students with the same level of cognitive achievement differ in their levels of effort, self-discipline, persistence, cooperation, self-presentation, tolerance, respect, and other non-cognitive dimensions. All these dimensions play a role in forming healthy character and contribute to productive relations in workplaces, communities, families, and politics.

The almost singular focus on test score performance in educational assessments at both domestic and international levels does rest on some foundation. The cognitive domains being tested are important determinants of both educational outcomes and life chances, and the measurement technologies are well developed. Moreover, the process for assessing cognitive skills is parsimonious in that a valid sample of cognitive knowledge and behaviour can be obtained and evaluated at low cost. But the evidence does not support the assumption that cognitive skills are all that counts and that they alone can produce healthy and productive adult personalities. Although these skills are important determinants of productivity and income at both individual and societal levels, the empirical studies I have cited here show that their measurable influence is far more modest than generally assumed. Moreover, their impact does not seem to be rising, despite the conventional wisdom. Employers who describe skill shortages in their employees emphasize getting workers with proper attitudes and social behaviours, at least as much as they emphasize cognitive competencies. The studies by Heckman and his colleagues show that the connections between non-cognitive skills and workplace productivity are of comparable importance overall and are even more important than cognitive skills in the productive development and influence on wages and graduation of older children.

Cunha and Heckman (2010, p. 401) conclude that the non-cognitive variables contribute to the impact that cognitive variables have on earnings, but they found only weak evidence of the reverse. In general, I would leave this as an open question. Some four decades ago I used the Coleman et al. (1966) data to estimate the determinants of multiple school outcomes in a model that allowed me to estimate simultaneous equations addressing both cognitive and non-cognitive school outcomes and their influences on each other (Levin 1970). The Coleman et al. (1966) report represented the largest social science survey in the United States (about 70,000 teachers and 700,000 students), seeking to ascertain the determinants of student achievement. By using both student attitudes and achievement in combination, I was able to estimate an early picture of the contribution of non-cognitive variables to achievement and achievement to non-cognitive development. The results of that model estimation suggested reciprocal relationships, in which motivation and sense of efficacy influence student achievement and are also influenced by student achievement. Only now are we seeing an extension of this early work.

Thus, there are at least three reasons why the singular use of academic achievement measures to predict economic productivity and growth results in overstated findings, when non-cognitive measures are omitted. First, academic achievement is correlated with non-cognitive attributes and serves as a surrogate for them in efforts to predict economic outcomes, overstating the purely cognitive effects when the non-cognitive variables are omitted. Second, non-cognitive attributes are not merely correlated with cognitive attributes, but contribute to cognitive outcomes. For example, Duckworth and Seligman (2005) found that self-discipline was far more important than IQ in predicting academic performance.

The third reason is that aggregated attempts to connect academic test scores with economic growth at the country level suffer the same kind of upward bias that Hanushek, Rivkin, and Taylor (1996) stress when criticizing the upward bias in aggregate estimates of educational production functions. On this basis, it is highly likely that the dramatic and highly publicized estimates by Hanushek and Woessman (2008), that international achievement results make immense contributions to economic growth among countries, overstate the relationship, possibly by a very large magnitude. Unfortunately, the promise of massive gains in economic output from gains in test scores has been disseminated widely, even though those administering policy are neither aware of nor informed about what is missing and the fact that they are likely to be vastly overstated.

Far from being harmless, the focus on test scores and the omission of the non-cognitive impact of schools can create far-reaching damage. In recent years, in the United States and other countries, attempts have been made to develop evidence-based policies for education. But the evidence that is presented is limited to test score comparisons, with the explicit or tacit assumption that test scores are the crucial determinant of labour force quality. This message places pressure on schools, by leading citizens and government to focus exclusively on raising test scores. In particular, accountability sanctions pressure schools to raise test scores in the limited domains and measures used in the national and international assessments, typically in reading, mathematics, or sciences. Schools are pressed to use their time and resources to improve scores on these subjects at the expense of other activities and subjects including non-cognitive goals. And the instructional strategies used to raise test results, such as test preparation, cramming, tutoring, and endless memorization, may have little effect on the broader cognitive and non-cognitive skills that people need if they are to perform as competent adults contributing to a dynamic economy. Other goals may be as, or more, important in the long run in terms of creating productive, equitable, and socially cohesive societies and economic growth (Gradstein and Justman 2002).

In the United States, the singular focus on a cognitive achievement gap created by national policy and funding has led to schools narrowing their curriculum and focusing on test preparation as a major instructional strategy (Rothstein, Jacobsen, and Wilder 2008). It is difficult for an evidence-based policy to embrace non-cognitive measures when the assessment practices exclude them from national and international studies. The obsession with the gap in test scores among races obscures the non-cognitive gap, which may be even more serious and a higher priority to address to improve various outcomes. For example, Fortin (2008) found that the effect on labour market outcomes of non-cognitive ability was stronger for blacks than whites and was a particularly strong predictor of the black-white gap in the incarceration rates of males.

A singular focus on students' scores on cognitive tests can also introduce instructional policies that ignore the importance of non-cognitive skills and fail to value the roles that teachers and schools play in developing students' non-cognitive skills. For example, many states and local school districts in the United States have adopted a "value-added approach" as a basis for teacher policy: Student gains on test scores ("value-added") are associated with individual teachers and become the basis for decisions on hiring, retaining, and remunerating teachers. With the recent cuts in public funding, school districts are considering laying off teachers, based on the value-added metric. But, in addition to the serious methodological issues surrounding the calculation of the value added for each teacher (Corcoran 2010; Harris 2009), an even more fundamental question arises. Why has the purpose of schooling and teacher productivity been reduced to the gains on narrowly construed math and verbal tests, if we expect so many other results from schools, including

non-cognitive outcomes? Even if a tradeoff can be demonstrated between teachers' effectiveness in developing cognitive and non-cognitive skills, educational policy must take account of both. That is the case for incorporating non-cognitive skill measurement in both large- and small-scale assessments and in considerations of what constitutes world-class status. This case has been recognized increasingly on both sides of the Atlantic. For example, Brunello and Schlotter (2010) prepared a report on the topic for the European Commission.

Next steps

To incorporate non-cognitive skills into assessments is a major challenge. Over ten years ago, Heckman and Rubinstein (2001) concluded this in their study of the GED:

We have established the quantitative importance of non-cognitive skills without identifying any specific non-cognitive skill. Research in the field is in its infancy. Too little is understood about the formation of these skills or about the separate effects of all of these diverse traits currently subsumed under the rubric of non-cognitive skills. (p. 149)

Fortunately, the research has exploded on this topic. Just seven years after the publication of this bleak statement, Cunha and Heckman (2008) were able to identify and employ specific non-cognitive measures in existing data sets that could be used for analysis augmented by further developments. As mentioned above, Almlund et al. (2011), Borghans et al. (2008) and Kyllonen et al. (2008) have developed rich literature reviews of non-cognitive skills, and their measurement, and linked these to specific school interventions that might raise non-cognitive performance in key areas.

My recommendation is to build on these efforts by selecting a few non-cognitive skill areas and measures that can be incorporated into research on academic achievement, school graduation, post-secondary attainments, labour market outcomes, health status, and reduced involvement in the criminal justice system in conjunction with the standard academic performance measures. The Big Five are certainly leading candidates with guidelines already suggested in the review by Almlund et al. (2011). Structural models and quasi-experimental designs might be used to understand the interplay of cognitive and non-cognitive skills in explaining particular outcomes for specific demographic groups. At some point we should learn enough to incorporate specific non-cognitive measures into both small- and large-scale assessments that can lead to a deeper understanding of school effects and school policy and a more inclusive framework for ascertaining what is, in fact, world-class education.

References

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011, January). *Personality psychology and economics*. (Working paper 16822). Cambridge, MA: National Bureau of Economic Research. <http://www.nber.org/papers/w16822.pdf>.
- Barnett, W. S., et al. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23(3), 299–313.
- Becker, G. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

- Belfield, C., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High Scope/Perry Preschool Program: Cost-benefit analysis using data from the age—40 follow-up. *Journal of Human Resources*, 41(1), 162–190.
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function and false belief understanding to emerging math and literacy in kindergarten. *Child Development*, 78(2), 647–663.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008a). The economics and psychology of personality traits. *The Journal of Human Resources*, 43(4), 972–1059.
- Borghans, L., ter Weel, B., & Weinberg, B. A. (2008b). Interpersonal styles and labor market outcomes. *The Journal of Human Resources*, 43(4), 815–858.
- Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America*. New York: Basic Books.
- Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, 39(4), 137–1176.
- Brunello, G., & Schlotter, M. (2010). *The effect of non cognitive skills and personality traits on labour market outcomes*. Analytical Report for the European Commission prepared by the European Expert Network on Economics of Education. http://www.epis.pt/downloads/dest_15_10_2010.pdf.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579–620.
- Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1802–1863). Amsterdam: Elsevier.
- Cawley, J., Heckman, J., & Vytlačil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4), 419–442.
- Coleman, J. S., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Office of Education, Government Printing Office.
- Corcoran, S. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: The Annenberg Institute for School Reform, Brown University.
- Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources*, 42(4), 738–782.
- Cunha, F., & Heckman, J. J. (2010). Investing in our young people. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life* (pp. 381–414). New York: Cambridge University Press.
- Dee, T., & West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, 33(1), 23–46.
- Digman, J. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440.
- Dreeben, C. (1968). *On what is learned in school*. Reading, MA: Addison Wesley.
- Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939–944.
- Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention and behavior problems. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality and the uncertain life chance of low-income children* (pp. 47–70). New York: Russell Sage.
- Duncan, G. J., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432.
- Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557–577.
- Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement and graduating from high school. *Journal of Educational Psychology*, 97(2), 214–223.
- Fortin, N. M. (2008). The gender wage gap among young adults in the United States: The importance of money vs. people. *The Journal of Human Resources*, 43(4), 884–918.
- Gradstein, M., & Justman, M. (2002). Education, social cohesion, and economic growth. *American Economic Review*, 92(4), 1192–1204.
- Hanushek, E., Rivkin, S., & Taylor, L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics*, 78(4), 611–627.
- Hanushek, E., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668.
- Harris, D. (2009). Would accountability based on teacher value-added be smart policy? An examination of the statistical properties and policy alternatives. *Educational Finance and Policy*, 4(4), 319–350.

- Hartigan, J., & Wigdor, A. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). A new cost-benefit and rate of return analysis for the Perry Preschool Program: A summary. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life* (pp. 366–380). New York: Cambridge University Press.
- Heckman, J., & Rubenstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED Testing Program. *American Economic Review*, *91*(2), 145–149.
- Heineck, G., & Anger, S. (2010). The returns to cognitive ability and personality traits in Germany. *Labour Economics*, *17*(3), 535–546.
- Inkeles, A. (1966). The socialization of competence. *Harvard Educational Review*, *36*(3), 265–283.
- Inkeles, A. (1975). Becoming modern: Individual change in six developing countries. *Ethos*, *3*(2), 323–342.
- Inkeles, A., & Smith, D. (1974). *Becoming modern: Individual changes in six developing societies*. Cambridge, MA: Harvard University Press.
- IZA [Institute for the Study of Labor] (2011). Workshop: Cognitive and Non-Cognitive Skills, Bonn, 25–27 January. <http://www.iza.org/link/CoNoCoSk2011>.
- Kniesner, T. J., & ter Weel, B. (Eds.) (2008, Fall). Noncognitive Skills and Their Development (Special issue), *Journal of Human Resources*, *43*(4). <http://jhr.uwpress.org>.
- Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *PNAS [Proceedings of the National Academy of Sciences]*, *103*(27), 10155–10162.
- Krueger, A., & Schkade, D. (2008). Sorting in the labor market: Do gregarious workers flock to interactive jobs? *The Journal of Human Resources*, *43*(4), 859–883.
- Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2008). *Personality, motivation, and college readiness: A prospectus for assessment and development*. Princeton, NJ: Educational Testing Service.
- Levin, H. M. (1970). A new model of school effectiveness. In A. M. Mood (Ed.), *Do teachers make a difference?* (pp. 55–78). Washington, DC: Office of Education, U.S. Department of Health, Education and Welfare.
- Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher*, *38*(1), 5–20.
- Levin, H. M. (2012). The utility and need for incorporating noncognitive skills into large-scale educational assessments. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The Role of international large-scale assessments: Perspectives from technology, economy, & educational research* (pp. 67–86). New York: Springer.
- Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, *3*, 101–128.
- Mosteller, F. (1995). The Tennessee Study of Class Size in the Early School Grades. *The Future of Children*, *5*(2), 113–127.
- Murnane, R., Willett, J., Bratz, M., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later: Evidence from the NLSY. *Economics of Education Review*, *20*, 311–320.
- Murnane, R., Willett, J., Duhaldeborde, Y., & Tyler, J. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, *19*(4), 547–568.
- Murnane, R., Willett, J., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, *77*(2), 251–266.
- National Research Council (1984). *High schools and the changing workplace: The employers' view (Report of the Panel on Secondary School Education for the Changing Workplace)*. Washington, DC: National Academy Press.
- Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*, *84*(3), 376–389.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130.
- Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) investing in the very young. *Economics of Education Review*, *29*, 271–282.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York: Teachers College Press.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, *56*(4), 302–318.

- Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, 81(1), 6–22.
- Schweinhart, L. J. (2010). The challenge of the High Scope Perry Preschool Study. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life* (pp. 366–380). New York: Cambridge University Press.
- Segal, C. (2008). Classroom behavior. *The Journal of Human Resources*, 43(4), 783–814.
- Shury, J., Winterbotham, M., Davies, B., & Oldfield, K. (2010). *National employer skills survey for England 2009: Key findings report*. South Yorkshire: UK Commission for Employment and Skills.
- Urzua, S. (2008). Racial labor market gaps: The role of abilities and schooling choices. *The Journal of Human Resources*, 43(4), 919–971.
- Vignoles, A., De Coulon, A., & Marcenaro-Gutierrez, O. (2011). The value of basic skills in the British labour market. *Oxford Economic Papers*, 63(1), 27–48.
- Zemsky, R., & Iannozzi, M. (1995). *A reality check: First findings from the EQW National Employer Survey* (EQW Issue no. 10). Philadelphia: National Center on the Educational Quality of the Workforce, University of Pennsylvania. <http://www.eric.ed.gov/PDFS/ED398385.pdf>.

Author Biography

Henry M. Levin (United States) is the William Heard Kilpatrick Professor of Economics and Education at Teachers College, where he also directs the National Center for the Study of Privatization in Education (NCSPE) and co-directs the Center for Benefit-Cost Studies in Education. He is also the David Jacks Professor of Higher Education, Emeritus, at Stanford University, where he served on the faculty for 31 years, with a joint appointment in the School of Education and Department of Economics. He has been president of the Palo Alto, California, School Board; the Comparative and International Education Society (CIES); and the American Evaluation Association, and is a member of the Board of Trustees of the Educational Testing Service (ETS). He is an elected member of the National Academy of Education, the author of about 300 articles, and the author or editor of 20 books.