DOCUMENT RESUME

ED 473 820 EC 309 433

AUTHOR	Miller, Phyllis, Ed.
TITLE	Does a High IQ Mean You're Smart?
INSTITUTION	American Mensa Education and Research Foundation, Arlington, TX.
PUB DATE	2002-00-00
NOTE	100p.; Published three times per year. Theme issue.
AVAILABLE FROM	Mensa Research Journal, 1229 Corporate Dr. West, Arlington, TX 76006-6103. Tel: 973-655-4225; Fax: 973-655-7382; Web site: http://wwww.us.mensa.org.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Mensa Research Journal; v33 n3 Fall 2002
EDRS PRICE	EDRS Price MF01/PC05 Plus Postage.
DESCRIPTORS	*Ability Identification; Cognitive Development; Elementary Secondary Education; *Gifted; *Intelligence; Intelligence Quotient; *Intelligence Tests; Metacognition; *Multiple Intelligences; Postsecondary Education

ABSTRACT

This journal issue presents articles that attempt to expand common ideas of intelligence and giftedness. Following a brief article offering "Notes, Quotes, and Anecdotes" by Francis Carter, the first major article is "Democratizing Our Concept of Human Intelligence" (by Wendy M. Williams). In it, the limitations of IQ tests are explored as are the different types of practical intelligence. The second article, by Wendy M. Williams and others, is "Practical Intelligence for School: Developing Metacognitive Sources of Achievement in Adolescence." It reports on an effort to develop practical intelligence with middle school students using a five-part program that stresses five sources of metacognitive awareness. The program successfully enhanced both practical and academic skills in each of the target skill areas. The third article is "Emotional Intelligence Meets Traditional Standards for an Intelligence" by John D. Mayer and others. It reports on two studies, one with adults and one with adolescents, using a new test of emotional intelligence, the Multifactor Emotional Intelligence Scale (MEIS). These studies showed that intelligence, as measured by the MEIS, meets three classical criteria of a standard intelligence. (Individual articles contain references.) (DB)

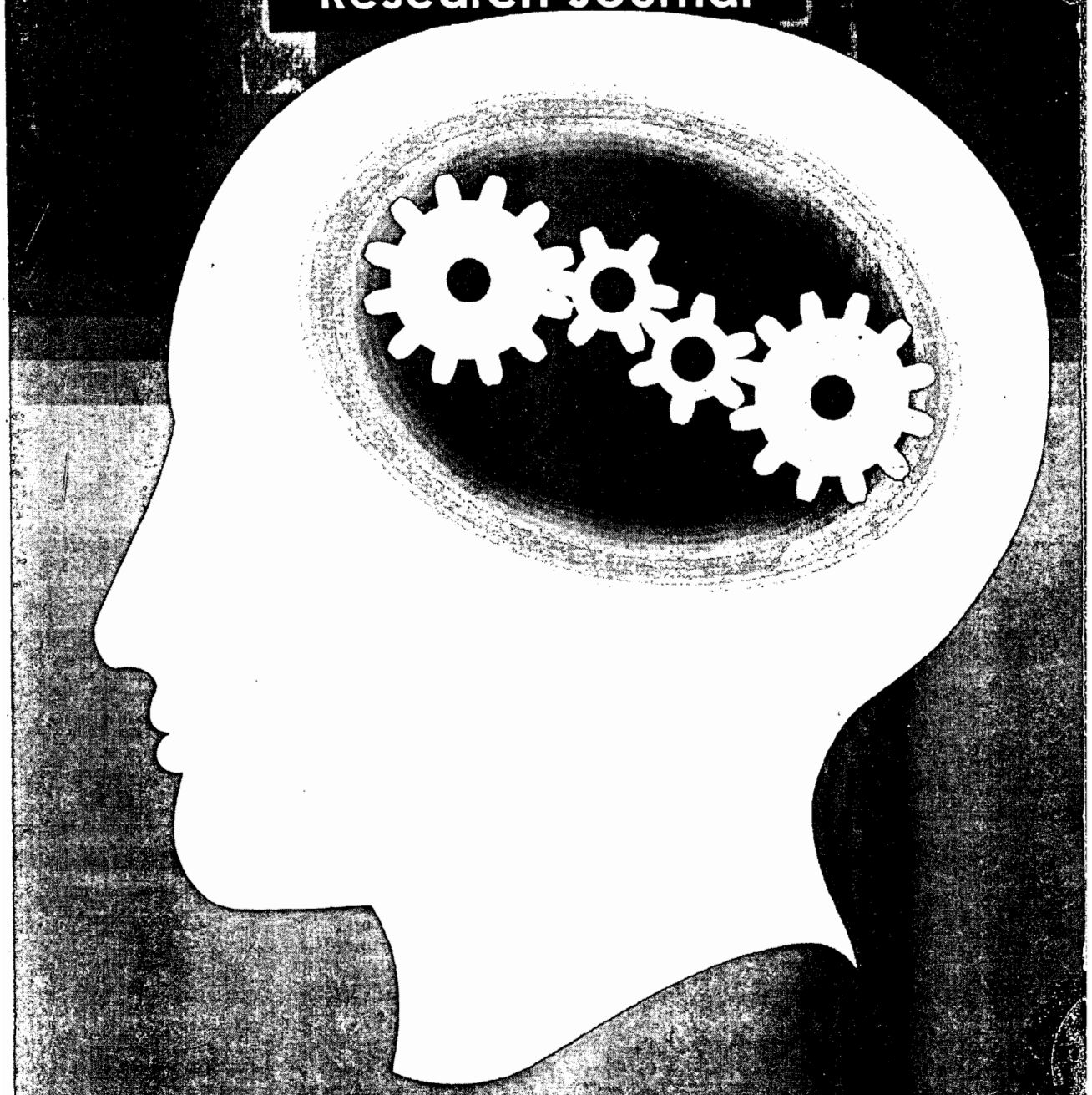
MENSA: Research Journal
Does a High IQ Mean You're Smart?

Edited By
Phyllis Miller

Vol.33, No.3/2002

Published by Mensa Education & Research Foundation and Mensa International, Ltd.™

MENSA Research Journal



Does a High IQ Mean You're Smart?

Vol. 33, No. 3

MENSA
Research Journal

Vol. 33, No. 3 Fall 2002

Table of Contents

Cover courtesy of Photosho

4

Mensa Research Journal is published three times a year, in conjunction with Mensa International, Ltd., by the Mensa Education & Research Foundation, Gregory Timmers, president, 2544 Chapel Hill Rd., Springfield, IL 62702.

Staff:

Editor • Phyllis Miller, 23 Lexington Road, Somerset, NJ 08873, MRJ@merf.us.mensa.org

Associate Editor • Francis Cartier, 1029 Forest Ave., Pacific Grove, CA 93950, fcar889755@cs.com

Circulation Manager • Jill Beckham, American Mensa, Ltd., 1229 Corporate Drive West, Arlington, TX 76006-6103, JillB@AmericanMensa.org.

Production Manager • Katie Hooten, American Mensa, Ltd., 1229 Corporate Drive West, Arlington, TX 76006-6103, KatieH@AmericanMensa.org.

Proofreader • Nancy Michell, 31 12th St., Somerset, NJ 08873, nancymic@microsoft.com.

The Journal is published three times a year. Subscribers will receive all issues for which they have paid, even if frequency of publication varies. Membership in Mensa is not required. To subscribe, order back issues, or report address changes, write to *Mensa Research Journal*, 1229 Corporate Drive West, Arlington, TX 76006-6103.

The Mensa Education & Research Foundation

Mensa, the high IQ society, provides a meeting of the minds for people who score in the top 2 percent on standardized IQ tests. As an international organization with thousands of members worldwide, Mensa seeks to identify and foster human intelligence; encourage research in the nature, characteristics and uses of intelligence; and provide a stimulating intellectual and social environment for its members.

The first two of these purposes are largely carried out by the Mensa Education & Research Foundation (MERF). MERF is a philanthropic, nonprofit, tax-exempt organization funded primarily by gifts from Mensa members and others. MERF awards scholarships and research prizes, publishes the *Mensa Research Journal,* and funds other projects consistent with its mission.

For more information about the Mensa Education & Research Foundation, write to MERF, 1229 Corporate Drive West, Arlington, TX 76006-6103. For information about joining American Mensa, call 800/66-MENSA or go to us.mensa.org. If you reside outside the U.S., see the Mensa International, Ltd., Web site at www.mensa.org.

To renew your subscription or subscribe to the *Mensa Research Journal*, just fill out the form on the next page (or a photocopy of it) and mail it with a check in the appropriate amount. Please refer to your mailing label now to determine how many issues of the *MRJ* you will receive before your subscription expires.

To: MERF, 1229 Corporate Drive West, Arlington, Texas 76006-6103

Please enter my subscription to the *Mensa Research Journal*.

❒ New ❒ 3 issues U.S. $21 (outside U.S. $25)
❒ Renewal ❒ 6 issues U.S. $42 (outside U.S. $50)
❒ Sample *Mensa Research Journal* U.S. $7 (outside U.S. $9)

Subscription amount enclosed $ _____
Please make payments in U.S. funds.

I would like to make the following tax-deductible contribution to MERF.

❒ Contributor ($25 to $99)	$ _____	Allocation (if desired)
❒ MERF Donor ($100-249)	$ _____	❒ General support
❒ Bronze Donor ($250-499)	$ _____	❒ Scholarships
❒ Silver Donor ($500-999)	$ _____	❒ Gifted Children Programs
❒ Gold Donor ($1,000-2,499)	$ _____	❒ Addition to endowed fund
❒ Platinum Donor ($2,500+)	$ _____	named _____

I would like my contribution to be a ❒ Memorial honoring _____
❒ Tribute honoring _____

Please send acknowledgement or notice to _____

Address _____

City/State/Zip _____

Name and address of contributor/subscriber:

Name _____ Member # _____

Address _____

City/State/Zip _____

For Office Use
Date received: _____ Total $ _____ Bank: _____ Date entered: _____
Sub code: _____ First issue number: _____ Date sent: _____

Editor's Preface

Everyone knows someone like Wendy Williams' grandfather. Never went to college, but he's the one you go to with a computer problem. Never finished high school, but writes the most beautiful letters, with a vocabulary far exceeding your own. Doesn't appear to be the sharpest crayon in the box, but reads incessantly and when you hunger for someone to discuss the book you just read you go to him.

My mother didn't finish high school but she did arithmetic in her head. Never wrote it down, and did it faster than those who did write it. My grandmother had no schooling whatsoever. Completely illiterate, she kept up with current events by having my grandfather read the newspaper aloud to her. Yet she could speak three languages fluently. And my friend, who is an artist and dyslexic, is the one I go to when I want to have a conversation about the deeper meaning of some scientific discovery.

And you have to wonder why so many supremely intelligent (high IQ) children have so much difficulty in school. Why so many intelligent teachers have problems teaching gifted children. Why so many intelligent parents have problems with their own intelligent kids.

In this issue, Dr. Williams and other scholars present their views on intelligence, and the one-sided approach we seem to take toward it. Perhaps there are other ways of being smart, and other ways to measure how smart we are.

Today, when so much depends on the SAT and related exams, maybe it's time to take a fresh look at these concerns. It's not bad to do really well on the SAT (Whew! That's good to hear!), but maybe it's not so bad to do less well.

Special Pat On The Back Department: *The Mensa Research Journal* has won an Honorable Mention award in the Association Trends national publications competition. As a reader, you should feel honored!

Phyllis Miller
Editor

Notes, Quotes and Anecdotes

• Phillip Morrison was writing about physics when he wrote these words in *Nature*, October 4, 2001, but they seem apt here, too.

> Science owes the outside world a clear, brief statement of its views, but these accounts will never be clear until we go beyond mere buzz-words and start conveying real ideas .... Perhaps we can enlist verbs to illuminate the nouns. Poetic, even ironic, tags catch on well, but name tags are just not enough.

I couldn't agree more. I may have said this here before but it bears repeating. Our minds are much influenced by the "parts of speech" (nouns, verbs, adjectives, etc.) we habitually use in talking and thinking about talent, intelligence, creativity, etc., as nouns. We might be liberated from some of our preconceptions by eschewing those reifying (thing-ifying) nouns and using only verbs, adverbs, or adjectives for such concepts. That won't be easy, as language habits are hard to break.

> Actions which were first voluntary, soon became habitual, and at last hereditary, and then came to be performed even in opposition to the will.

Oh, yes, and *in*actions, too. You can find the above quotation in Charles Darwin, *The Expression of Emotions in Man and Animals*, 1872.

• Research into "artificial intelligence" virtually stopped in 1989 when the Defense Advanced Research Agency (DARPA) gave up on the idea of a "smart truck." It turned out that none of the AI developers were able to devise any kind of machine that could adapt to a multiplicity of environments or rapidly changing circumstances. Even the so-called "smart" bombs used in the Gulf War eventually had to be guided by a human mind flying overhead. Here was another case where the English language was misused. If there really were an "artificial intelligence," it seems to me it would have to be called "intelligence," which is characterized by, among other things, its extraordinary flexibility.

• Dr. Robert J. Sternberg, whose name is well known to *MRJ* readers, is seeking to enter an interesting, new concept into research on intelligence. He calls it *wisdom*. The entire issue of *Educational Psychologist*, Fall 2001, was devoted to it.

• Here's a website you mustn't miss! Go to www.sengifted.org. SENG stands for Supporting Emotional Needs of the Gifted. SENG is at P.O. Box 6550, Scottsdale, Arizona. I don't know why I've never run across this organization before because it has obviously been around for quite a while. The 18th

SENG conference was held in July 2001 in Irvine, Calif. "It is the mission of SENG to support bright, talented, gifted individuals and their families, and the professionals who work with them, to better understand the nature of giftedness, and to help these individuals reach their personal potential." SENG's website has a wealth of information as well as links to more than 20 related websites, several publishers, and other organizations.

• Those interested in neuropsychology of the brain will be well served by www.brainsource.com/. It is not only for experts in that field.

• I've been a member of the American Education Research Association for perhaps 40 years and now of its fairly new special interest group, Research on Giftedness and Talent. You don't have to be a member, though, to access www.aeragifted.org. You can read or download many of the research studies given at the last AERA convention. The site also provides links to five other sources of information about the gifted and talented.

• One of those links led me to an organization that I can't recall hearing about before: The Davidson Institute for Talent Development. It was founded in 1999 "To recognize, nurture and support *profoundly gifted* young people and to provide opportunities for them to develop their talents in positive ways to create value for themselves and others." It provides publications, advisory services, speakers and much more for profoundly gifted young people, parents, teachers, counselors, etc. Go to www.davidson-institute.org. Even if you're merely curious, plan to spend up to half an hour exploring it.

• Take a look, too, at giftedpsychologypress.com.

• The American Psychological Association has long neglected the gifted but that changed last year. APA created the Center for Gifted Education Policy headed by Dr. Rena Subotnik, who says:

> Some disciplines [like music and sports] have highly developed structures for cultivating talent; but others take a more haphazard approach. I'd like to draw the attention of the latter group to how they can do it more consciously.

The center will publish information for parents and teachers, a book on "cutting-edge" research issues, etc. It also has a website with nearly 50 links to some you know of and some you might not. Try apa.org/ed/cgepwblinks.html.

• I just learned a new word: Hyperlexia. We all know about dyslexia and that it sometimes afflicts otherwise gifted children. Hyperlexia is a syndrome involving a very precocious ability to read words or an intense fascination with letters

and numbers. It is often coupled with difficulties in understanding spoken language (hyperlexics may even appear to be deaf) and in socializing. Go to hyperlexia.org.

♦ I sometimes wish that *MRJ* had the space to reprint everything in the *National Research Center on the Gifted and Talented Newsletter*. They permit reprinting free, asking only that you mention the source. Mensa local group editors take note! A subscription to the *Newsletter* is free for the asking from NRCGIT, 2131 Hillside Rd., Unit 3007, Storrs, CT 06269-3007.

♦ "The hallmark of a truly brilliant idea is as soon as someone expresses it, it seems obvious." David Funder, Ph.D.

♦ Want a TIGER in your think-tank? The Tennessee Initiative for Gifted Education (TIGER) publishes a free newsletter that can be obtained just by sending a blank email to TIGERadvocate-subscribe@yahoo.groups.com. TIGER is a "network of parents, educators, administrators, legislators, and the general public working to foster an understanding of gifted children and their exceptional needs and to advocate for the appropriate education of all students." As a vocabulary nitpicker, I'd rather they said "each student" instead of "all students," but I forgive them; their hearts are in the right place.

♦ Are you curious about curiosity? Todd Kashdan and Frank Fincham are. See their letter in *American Psychologist*, May 2002, page 373. "Specifically, we characterize curiosity as a self-regulatory mechanism that facilitates intrinsic goal effort, perseverance, personal growth, and, under the right conditions, creativity .... We are not suggesting that high curiosity leads directly to high creativity but that high curiosity is necessary, though not sufficient, for creativity." I say amen to that, and like them, I wish there were more research into curiosity, especially as it might apply to gifted children and adults. Is curiosity a uniquely powerful form of motivation? Where does it come from? What life factors seem to squelch a child's natural curiosity? Why do some people seem to have less curiosity than others? Is high curiosity one of the essential elements in the way a developing poet sees and hears and reacts? Think about curiosity a moment. I'll wait.

♦ *ADHD: The Great Misdiagnosis* by Julian Stuart Haber (Dallas, TX: Taylor Trade Publishing, 2000) is primarily for parents and teachers. Its major theme is the question: "Are too many children being unnecessarily diagnosed and treated for the disorder?" Haber disagrees with those who say there's an epidemic of attention deficit hyperactivity disorder. The symptoms of ADHD may actually indicate a variety of other disorders needing different treatments. Furthermore, "Many talented children with learning difficulties and ADHD can

advance in various areas of non-scholastic intelligence .... Finding the talent and gearing it to future skills is obviously very important." See Nikolay Zavadenko's review of the book in *Contemporary Psychology: APA Review of Books*, 2002, Vol. 47, No. 3, page 289.

• "In recent years, the science of intelligence has probably generated more 'bad noise' from what appears to be an overly excitable media than any other topic in psychology. Despite remonstrations that 'it doesn't matter,' or 'can't be measured,' intelligence does matter and it can be measured. In fact, excitement generated by intelligence research approaches that generated by issues of sex and money, because deep down people understand that intelligence is important, it always has been, and it always will be." David C. Geary, Deptartment of Psychological Sciences, University of Missouri.

• President George W. Bush has proclaimed March 10 to 16, 2003, Brain Awareness Week. I learned that from the Dana Alliance (dana.org), whose motto seems to be, "Just Use It." It's not clear to me how we are to celebrate (cerebrate?) that week, but mark it on your calendar anyway. Surely you'll think of something.

F. C.

Democratizing Our Concept of Human Intelligence

Wendy M. Williams, Cornell University

My grandfather was clever. With no formal schooling to speak of, he could build anything, from a dollhouse to a real house, from scratch, without plans. He also could fix anything — kitchen appliances, cars, children's toys, radios, televisions, you name it. He even published a book of his poems when he was in his 70s. He was not clever, however, at taking I.Q. tests, which he confronted in grade school, in the military, and when he looked for a job when he was in his early 20s. He hated taking the tests; he was made anxious by the clock ticking as he worked, and he found it confusing and unnatural to think in terms of abstractions, be they mathematical, pictorial, or verbal.

Because of his performance on tests, my grandfather did not consider himself very intelligent. Neither did the teachers, military recruiters, and job-placement personnel who used the test scores; They reduced my grandfather's intelligence to a simple, relatively low number on a page and labeled him "slow." The I.Q. tests that my grandfather took in the 1930s — versions of which are still in use today — were created to determine which children failing in school were doing so because of low intelligence, and which were failing for other reasons. Through questions about the meaning of words or paragraphs, mathematical problems, visual patterns, and so forth, these tests measured intelligence in terms of the number of problems a person could solve, compared with the average for other people of the same age.

Throughout our society, we still use I.Q. tests, and their close surrogates such as the SAT, in the belief that they provide a meaningful measure of a person's innate intelligence and capacity for success in intellectual tasks. We all know the considerable weight these tests are given throughout education, as well as in hiring and promotion decisions in the workplace. But scholars still have not explained how, if I.Q. tests tell us the most important things we need to know about a person's intelligence, we can account for my grandfather and the many others like him, who are competent and successful in so many domains in the real world.

This is the issue that my colleagues and I have studied in our attempt to democratize the concept of intelligence, by including in it more and different types of abilities and talents. While we have been conducting our research, other scholars working in the same area have demonstrated that I.Q. tests' reputation as an ultimate seal of approval was premature.

For example, consider the work of James Flynn, a political scientist at the University of Otago, in New Zealand. He proved that I.Q. scores have risen sharply over the past 60 or more years in all 20 nations for which data exist. In

Reprinted from *The Chronicle of Higher Education*, May 15, 1998

fact, a person born in 1877 whose score put him or her in what was then the 90th percentile on a widely used reasoning test would, with exactly the same number of correct answers, rank in only the 5th percentile of people born in 1967. (Flynn proved this by examining the raw numbers of correct answers on the same tests used over time. Most researchers rely on "normed scores," which are adjusted to keep the average score on a test constant from year to year, and which thus cannot accurately be compared over time.)

Are we really that much smarter than our grandparents? How could I.Q. scores change so much, so quickly?

We learned two things from Flynn's work: First, a high I.Q. score does not necessarily mean intelligence, nor does a low score mean stupidity. Second, whatever the test measures is highly mutable. Flynn is fond of saying that, if we take I.Q. scores seriously as meaningful predictors of intelligence, our grandparents would have been unable to understand the rules of baseball.

Given the rapidity of the changes Flynn reported, genetics could not be responsible, and so researchers have focused on aspects of culture, as well as on health and nutrition, in attempting to explain why people today are markedly outscoring their ancestors.

One possible cultural factor is that people are increasingly familiar with the material on certain types of I.Q. tests. My grandfather's generation rarely encountered anything in their everyday lives even remotely resembling the items on such tests. Today, however, mazes, puzzles, and other games that are thinly disguised versions of items from actual I.Q. tests appear on cereal boxes and on placemats at fast-food restaurants. People play with toys such as Rubik's Cube. Some computer screen-saver programs are strikingly similar to other kinds of intelligence tests: The complex patterns dancing around the screen closely resemble the Raven matrices, the most popular test of reasoning ability. Is it any wonder that today's kids outperform my grandfather's generation?

But the more important question is: Does this greater exposure to material similar to that on the tests make today's children and adults smarter in any meaningful way than earlier generations of test takers? I think not. The intellectual accomplishments of people in past eras are awe-inspiring, and the challenges and hardships that they had to overcome were extraordinary. Looking back on these accomplishments should make us cautious in interpreting the significance of I.Q. scores as predictors of likely success in the real world.

Perhaps the reason that so many individuals with low or moderate I.Q.'s, such as my grandfather, are so successful in their daily lives can be found in recent research that has broadened the concept of intelligence. Researchers today are demonstrating empirically the importance of many abilities that are not measured on I.Q. tests. Consider studies that my colleagues and I have conducted to assess practical and creative thinking in business, the military, and elementary and middle schools.

We wanted to know why some business managers with M.B.A.'s from pres-

tigious graduate schools alienate their subordinates virtually overnight, why some military leaders lose the respect of their soldiers and subordinate officers by adhering to formal doctrines even in situations where they are not adequate, and why some bright children hand in boring compositions after the deadline and then react with surprise when they receive low grades. We found that all of these people lack practical intelligence — an ability essential to success that differs from the more "academic" intelligence measured by I.Q. tests, and which is largely independent of it.

We learned that practical intelligence consists of three types of abilities — managing oneself, managing others, and managing the organization or environment in which one works, such as a school, corporation, or hospital. Each ability is important in a unique way, and each contributes to real-world success. People may be strong in one type of practical intelligence and weak in another, although, generally, being savvy about managing organizations builds on the abilities to manage oneself and others. Importantly, traditional measures of I.Q. tell us little about who has and does not have the three types of practical intelligence.

Where are scientists headed in our search to understand intelligence? Increasingly, we think in terms of types and facets of intelligence that lead to success in specific contexts: social intelligence, emotional intelligence, creative intelligence, and practical intelligence. We look at people's ability to manage their lives by motivating and organizing themselves to perform effectively. We consider people's ability to get along with their employees, peers, supervisors, and teachers. Often, it is those types of intelligence, as much as I.Q. scores, that determine success or failure in education and in the workplace, especially among people with a similar range of I.Q. scores.

Historically, a person's intelligence was reduced to a single number. Today, that number still holds sway in many admissions offices, but the realization is growing that we need to characterize and measure more of the abilities that are important to adult success. We owe the next generation a broader and more relevant battery of tests, designed to measure the many varied abilities that contribute to success in the real world. Better tests will lead to the admission of applicants with a wider variety of skills, thus diversifying further the pool of talent available to our society.

As we look ahead to the demographic changes under way and recognize the need to distribute educational and employment opportunities fairly and broadly, it becomes even more essential for us to assess people's capabilities accurately. We need a conception of intelligence that encompasses my grandfather's talents. The most successful leaders in business, the professions, and other enterprises know how to define workable goals and motivate themselves to accomplish them; they know how to "read" and motivate other people; and they know how to distinguish solutions that work in the real world from ones that work only in books — all abilities that current I.Q. tests do not measure.

This is not to say that success on an I.Q. test does not provide meaningful information; it is just that other types of success matter, too. It should not escape us that the technological developments on which our society depends may require types of intelligence — practical and creative, for example — that are different from those emphasized in our standardized tests. The science of understanding intelligence thus may progress farther and faster by recognizing the wisdom of our grandparents.

Practical Intelligence for School: Developing Metacognitive Sources of Achievement in Adolescence

Wendy M. Williams, Cornell University; Tina Blythe, Harvard University; Noel White, WestEd; Jin Li, Brown University; Howard Gardner, Harvard University; and Robert J. Sternberg, Yale University

We sought to boost school achievement by creating an intervention that would develop practical intelligence for school in middle-school students. Researchers worked with teachers in Connecticut and Massachusetts schools over a two-year period. Teachers were trained to deliver a five-part program developed to inculcate practical intelligence by emphasizing five sources of metacognitive awareness: knowing why, knowing self, knowing differences, knowing process, and revisiting. A broad range of assessments was administered in a pre-post design both to the children receiving the practical intelligence program and to matched control children. We found that the program successfully enhanced both practical and academic skills in each of the target skill areas (reading, writing, homework, and test-taking) in children from diverse socioeconomic backgrounds attending diverse types of schools. These results are discussed in terms of the acquisition of cognitive and metacognitive insights during adolescence and the promise such insights hold for enhancing adolescent achievement over and above traditional g-based approaches to learning. Finally, we discuss trade-offs between ecologically based and laboratory-based interventions.

A bright 9-year-old child shows one of his parents a paper he has written for his teacher. The parent comments that the paper seems to have some good ideas, but that it is messy and also full of spelling and punctuation errors. The child responds that the teacher does not care about those things, only about the ideas. The parent demurs, but does not argue. The child hands in the paper and is shocked to receive a low grade. It turns out, surprisingly to the child — but not to the teacher or parent — that the teacher *presumed* the students knew that neatness, spelling, and punctuation matter, and so the teacher focused instead on "coming up with good ideas" in the classroom discussion of the assignment. Somewhere along the way, however, the child was derailed.

This child is intelligent in the conventional sense of the term (see Neisser,

---

The Practical Intelligence for School Project was funded by a grant from The James S. McDonnell Foundation. We are grateful to Melanie Gordon Brigockas and Todd Lubart for their assistance with the Yale portion of the research. We also thank Jie Qi Chen, Mara Krechevsky, and Stephanie Weber for their work on the Harvard portion of the research.

Reprinted from *Developmental Review* 22, 2002.

1979; Neisser et al., 1996; Sternberg, Conway, Ketron, & Bernstein, 1981) and yet seems to lack some kind of intelligence relevant to the school environment. This kind of intelligence, termed *practical intelligence for school*, refers to the individual's understanding of, and ability to respond appropriately to, the demands of the school environment (Gardner, Krechevsky, Sternberg, & Okagaki, 1994; Sternberg, Okagaki, & Jackson, 1990; Williams et al., 1996). Included in this concept would be understanding and responding appropriately to things such as doing homework; taking tests; reading for understanding; and, as in the case of the child above, writing effectively.

Practical intelligence for school is a specific aspect of practical intelligence, which has been studied by a variety of investigators in a variety of contexts (see Ceci, 1996; Sternberg, 1996, 1999; Sternberg, Forsythe, Hedlund, Horvath, Wagner, Williams, Snook, & Grigorenko, 2000; Sternberg & Wagner, 1986; Sternberg, Wagner, Williams, & Horvath, 1995, for reviews). The basic idea is that over the course of development from middle childhood to early adolescence, individuals acquire a set of procedural-knowledge skills (i.e., skills that tell people what to do and how and when to do it) that are relevant to their adaptation to real-world environments. This set of procedural-knowledge skills is not well or fully conceptualized by conventional notions of intelligence and is not well measured by conventional intelligence tests. Although not all investigators accept the existence of such a construct (e.g., Ree & Earles, 1993; Schmidt & Hunter, 1993), there is sufficient evidence that it is worth pursuing on an empirical basis. At minimum, it can be viewed as a part of general intelligence, broadly defined, whether or not it is viewed as wholly distinct from the academic aspects of intelligence. For example, Ceci (1996) speculated that the introspective awareness that undergirds practical intelligence (e.g., realizing that a noisy cafeteria is not as optimal a place to study as a quiet room) may be the basis of so-called general intelligence, or *g*, because such insights are important in most aspects of academic success. Thus, even if one does not accept the independence of practical intelligence from *g*, there may nevertheless be an important role for practical intelligence to play in a broader perspective on cognitive competence.

Can intelligence, in general, or practical intelligence, in particular, be taught? Few questions in the field of developmental psychology are more difficult to answer and more contentious in the answers they have generated. As explained below, the scientific evidence is mixed and subject to alternative interpretations. Part of the confusion has to do with the way practical intelligence is construed, with some viewing it as a largely independent type of cognition or thinking (e.g., Gardner, 1983, 1999) and others (e.g., Ceci, 1996) viewing it as the product of the application of other forms of cognition to practical problems. In the present study, we conceived of practical intelligence as a derivative of metacognitive awareness. This guided our development of thematic exercises that focused on various forms of introspective awareness.

Suggesting that intelligence can, in fact, be taught are a number of controlled studies, some of them yielding impressive gains. One of the most well known is the evaluation of the ODYSSEY project. This project was designed to raise the intellectual skills and school performance of Venezuelan school children of roughly middle-school age and was evaluated with highly favorable results by Herrnstein, Nickerson, DeSanchez, and Swets (1986). Ramey (1994; Ramey & Campbell, 1984, 1992; Ramey et al., 1992; Ramey & Ramey, 1998), studying younger children, also has accumulated substantial evidence that gains in intelligence and school performance are possible as a result of intensive interventions among high-risk preschoolers. Affirmative evidence also comes from the STAR project in Tennessee, in which reduction in class size in a large randomized design elevated kindergarten through third graders' reading scores from the 50th percentile to the 60th percentile (Finn & Achilles, 1990, 1999; Mosteller, 1995). Other programs have also shown at least limited, and sometimes quite impressive, success (Bereiter & Engelmann, 1966; Feuerstein, 1980; Garber, 1988; Nisbett, Fong, Lehman, & Cheng, 1987; Schweinhart, Barnes, & Weikart, 1993; see also Detterman & Sternberg, 1982; Honig, 1994; Nickerson, 1994; Nickerson, Perkins, & Smith, 1985).

The verdict is not unanimous, however. Suggesting a negative answer are Herrnstein and Murray (1994), whose review of the literature led them to conclude that little meaningful intellectual gain is possible in intervention studies short of early adoption. Others have come to the same conclusion (Jensen, 1969, 1989; McLaughlin, 1977; Spitz, 1986, 1992), suggesting, at the least, that the evidence can be interpreted in different ways.

Perhaps the most defensible interpretations of intellectual intervention studies are those that are mildly encouraging, but at least somewhat skeptical (Consortium for Longitudinal Studies, 1983; Lazar & Darlington, 1982; Snow & Yalow, 1982; Zigler & Berman, 1983). Indeed, any responsible study might be placed into this category because no serious psychologist has suggested that unlimited gains are possible. Rather, the best that can be said is that modest to moderate gains can be obtained in some cases under limited circumstances.[1] Typically, any initial gains fail to accumulate or "fade out" over time. For example, the STAR program's success was limited to the effect of a single year of reduced class size, with no systematic cumulative effects for subsequent years in smaller classes (Mosteller, 1995).

---

[1] Regardless of the specific conclusion, when evaluating research that attempts to increase intelligence, it is important to distinguish between increasing intelligence (broadly construed) and increasing IQ. Herrnstein and Murray (1994) reduce the first issue to the second; however, not all researchers would agree with this reasoning. Our focus in the present study was on enhancing those performances that reasonable individuals would consider to be intelligent for students in middle school (e.g., thinking critically, seeing oneself and one's intellectual products from other people's perspectives, managing one's intellectual performances, applying strategies appropriate to different intellectual tasks, and so on).

There could be various reasons why the evidence has been seen as mixed, some pertaining to the research itself and others pertaining to the ideological commitments of its interpreters. In the studies reported here, we tried to take into account two possible issues that might have engendered some of the mixed results of the past. First, our work was motivated by a clear guiding theoretical framework, a merger of Sternberg's (1985, 1996) triarchic theory of intelligence and Gardner's (1983, 1993) theory of multiple intelligences. Previous interventions have not typically been theory-driven (e.g., Lazar & Darlington, 1982; Zigler & Berman, 1983), instead relying on conventional activities and off-the-rack tests that are assumed to measure their impact. Second, we attempted to teach an extremely broad range of intellectual skills — those relevant to practical intelligence for school — rather than focus on skills targeted to more specific and, perhaps, clearly definable kinds of situations. By focusing on training a broad range of skills we hoped to improve students' performances on various types of tasks students complete every day. We also focused our evaluation and assessment of our program upon these diverse competencies. Because previous research on training more specific skills showed that some skills are more trainable than others, we hoped that an emphasis on a broad range of skills would be likely to demonstrate overall positive effects of training for various skills involved in general scholastic performance.

Theoretical Motivation

The theoretical motivation for our research is a combination of the theory of multiple intelligences proposed by Gardner (1983, 1993) and the triarchic theory of intelligence proposed by Sternberg (1985, 1996, 1999). The former theory, based on developmental, neurological, and various other forms of evidence,[2] proposes that intelligence can be understood in terms of seven distinct and relatively independent forms of intelligence: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal.[3] The latter theory proposes that intelligence has analytical, creative, and practical aspects. By viewing intelligence in each of the seven domains proposed by Gardner as having the three aspects proposed by Sternberg, we integrated these two theoretical frameworks.

For example, one could view linguistic intelligence, an important component of school success, as involving analytical aspects (e.g., of the author's intentions in writing a story or of whether a particular argument is persuasive), creative aspects (e.g., in writing a creative essay or a poem), and practical aspects (e.g.,

---

[2] These include evolutionary antecedents of a given intelligence; the existence of a symbol system (e.g., dance notation and maps); the existence of special populations (e.g., geniuses and savants); and, importantly, potential isolation by brain damage (e.g., language and spatial reasoning).

[3] Subsequently, Gardner (1999) expanded his theory of intelligence to encompass an eighth form of intelligence, termed naturalistic intelligence.

in writing or speaking persuasively so as to convince readers of the validity of one's arguments). One could view logical-mathematical intelligence as involving analytical aspects (e.g., in working accurately through the steps of a traditional mathematical proof), creative aspects (e.g., in discovering a novel approach for completing a mathematical proof), and practical aspects (e.g., in seeing a real-world analog of a mathematical equation and using this example to inform further thinking). The same merging applies to each of the other multiple intelligences.

Both theories have been applied in ways to boost both intelligence, in general, and school achievement, in particular (see, e.g., Gardner, 1993; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg & Spear-Swerling, 1996).[4] In this set of studies, we sought to boost school achievement by creating an intervention that would develop practical intelligence for school in middle-school students (Williams et al., 1996). This article reports on our evaluation of that intervention.

Method

The intervention program is a curriculum intervention designed to enhance the practical-thinking skills of early adolescents, especially those in the fifth and sixth grades. We focused on fifth- and sixth-grade students because we believed that the point at which a child leaves primary school and enters intermediate school (usually around sixth grade) is a time when the child is ripe for instruction in practical-thinking skills. The child at this juncture is old enough to assimilate and use the skills (many of which require a degree of metacognitive awareness not usually evident in younger children), but young enough to be open to learning them.[5] Around the time of entry into middle school, the child's practical thinking skills become more essential as he or she must change classes several times a day and deal with the demands of different teachers and, for the first time, a limited number of elective courses.

Materials

Themes of the practical intelligence for school (PIFS) curriculum. Based on extensive classroom observation, teacher and student interviews, and a review of the literature on metacognitive awareness (Metcalfe & Shimamura, 1996), our research team identified five major questions or themes that helped us understand the focus and mechanics of practical-thinking skills. These themes

---

[4] Although we distinguish between school-achievement-test performance and intelligencetest performance, this is not usually true among psychometric researchers (e.g., Scarr, 1997), who view *both* types of performance as *g*-loaded and hence correlated to the extent of their saturation on the first principal component or common factor.

[5] Our assumption is supported by Piaget's (1959) description of chilldren's entrance into formal operations in adolescence.

became the basis of our curricular intervention, which we called the Practical Intelligence for School (PIFS) program. These five themes were further refined through discussions with teachers during the implementation and during subsequent revisions of the curriculum. Each theme focuses on an aspect of metacognitive awareness that pertains to a particular academic domain.

The first theme is *Knowing Why*. We believed that students would benefit from understanding the purposes behind the tasks they are asked to do in school. (What is the point of homework? Why is it important to read?) They also need to know how their school work relates to their extracurricular lives. (How is the writing they do inside of school like the writing they do for other reasons? What kinds of tests do adults face in their work, and how do classroom and standardized school tests help adults to prepare for these tests in later life?) Students would be aided by an awareness of how learning is relevant to their lives now and how they can use it to improve their lives both now and later.

The second theme is *Knowing Self*. Assessing one's own work and one's personal strengths and weaknesses is an integral component of practical competence, both in school and in life. The ability to capitalize on strengths and compensate for weaknesses is a prime benefit of self-knowledge. Recognizing their own current homework or reading habits and identifying the strengths and weaknesses in those habits is the first step students must take toward improving their work.

The third theme is *Knowing Differences*. A working style that is very successful in handling math assignments may be ineffective in writing an essay. The differences among subject matters, and, therefore, the need to learn varied approaches to work, become more pronounced as students move through middle school and into high school. The PIFS lessons ask students to compare various kinds of assignments and subject matters to reveal their diverse demands. Once students can see differences, they can also begin to think critically and see connections between various aspects of school and also between school and life in general.

The fourth theme is *Knowing Process*. How should a student tackle an assignment or a problem? The classroom is a world of its own, and success in it requires an awareness of the academic process. The curriculum describes processes involved in completing work within four focal areas (discussed below) — reading, writing, homework, and testing. Students learn to recognize and define problems for themselves. Lessons help students plan their work, develop effective strategies for problem solving, and draw upon a wide variety of resources in overcoming difficulties.

The fifth and final theme is *Revisiting*. Each booklet in the PIFS curriculum focuses on a particular kind of reflection that is especially useful to its focal area: rereading texts, revising writing, reworking homework, and reviewing tests. Children learn that going the extra mile to review and revise is the hallmark of the successful student.[6] For example, metacognitive research by Beale

(1996) has demonstrated that students who lack the insight needed to revisit incomprehensible text material can be trained to recognize the value of "revision," and their scores benefit from this insight.

Contents of the practical intelligence for school (PIFS) curriculum. We focused the curriculum on four major areas in which students require practical intelligence to succeed in school. This is not to say that these are the only salient areas or that our curriculum addresses *all* the practical thinking skills children need. Rather, we focused on these four areas because our observations and interviews suggested that practical intelligence in each of these areas is important to school success. The four PIFS focal areas are reading, writing, homework, and testing. The PIFS curriculum contains five sets of lessons, organized into booklets: an introductory booklet, a reading booklet, a writing booklet, a homework booklet, and a testing booklet (examples of PIFS lessons from each of the five booklets are presented in Figs. 1-5; the reader should review these materials to gain an appreciation of the content of PIFS lessons before proceeding).

The *Introduction* to the PIFS curriculum (see Fig. 1) contains a series of lessons that asks students to think about the nature of school and why they attend it. Students also reflect on their own personal strengths and interests, developing a profile of themselves that they can use to help them draw their interests into future school assignments. The *Thoughtful Reading* booklet (see Fig. 2) invites students to consider the role that reading plays in adult life. They develop their own reading profiles and identify their best strategies for overcoming difficulties in reading. They also compare and contrast different kinds of reading. *Expository Writing* lessons (see Fig. 3) cover the practical side of planning, writing, and revising clear, coherent pieces that convey students' ideas, incorporate their particular interests, and capitalize on their strengths.

The *Homework Planning and Execution* booklet (see Fig. 4) covers issues such as understanding the purposes of homework, identifying personal strengths and weaknesses in homework habits, developing organizational skills, and making use of a variety of resources to overcome homework problems. Finally, *Preparing for and Using Feedback from Tests* (see Fig. 5) includes lessons in which students discuss the importance of and uses for tests; develop important classroom and study skills like listening carefully, taking notes, and memorizing; practice various study strategies in preparing for tests; anticipate and plan for problems they might encounter while testing; and, finally, consider how best to make use of feedback from tests.

---

[6] Readers familiar with other frameworks will recognize that our five-part approach covers similar conceptual ground. For example, Brown's (1975) "Knowing, Knowing about Knowing, and Knowing How to Know" approach is represented in our first four themes, as are metacognitive strategy-training approaches, such as those studied by Pressley and others (e.g., O'Sullivan & Pressley, 1984; Pressley & Ghatala, 1990; Pressley, Levin, & Ghatala, 1984).

FIG. 1. Excerpts from the Introduction of the *Practical Intelligence for School* (PIFS) program.

LESSON 1.1 Why School?

Synopsis

A brief discussion to raise some of the kinds of questions that will be an important part of the Practical Intelligence for School program throughout the year.

What To Do

1. Give students a chance to think individually about why they go to school. You might use one or two of the questions listed for discussion below. You might have students write in journals as homework or for five minutes before starting discussion. Or you might have them simply take a minute to think quietly before starting the discussion.

2. Lead a group discussion based on broad questions that students can answer from their own experience, such as:

- What is school?
- What would you be doing if you weren't attending school?
- Why do you think schools were started in the first place?
- Most children used to start work on farms or in factories at a very young age. What do you think those children missed by not going to school?
- In what ways might it be better not to have school?
- What kinds of things do you learn in school?
- What makes an excellent school?
- What do you need to know to do well in school?
- How does school prepare you for later life?
- What do you need to know to help you outside of school and later in life?

There is no need to cover all questions.

3. Tell students that these are the kinds of questions that the class will be considering through the year. You might want to use the metaphor of "looking behind the scenes." Through PIES, the class will learn about what is behind school success and will develop skills for making school a better experience. These skills are also essential outside of school, and in later life as well.

Connections

Homework Lesson 4.1: Purposes for Homework

Testing Lesson 5.1: Why Tests?

Reading Lesson 2.1: Why Read?

Writing Lesson 3.1: Writing In School and Out

Follow-Up Activity

Whenever students start a new activity or project, have them think about why they are being asked to do it. How does it fit in with other things they do for school? What might they learn? How is it like what people do outside of school?

FIG. 2 Excerpts from the "Thoughtful Reading" chapter of the Practical Intelligence for School (PIFS) program.

LESSON 2.1 Why Read?

Central Themes

Knowing Why
Knowing Self

Synopsis

A "reading inventory" and a discussion to help students think about the role that reading plays in their everyday lives and in the lives of adults.

What To Do

1. Ask students how much time they spend reading each day. Probably they'll tell you "Not very much." But don't take their word for it. Ask questions about their television watching: Do they read the credits at the beginning and end of shows (what Bart Simpson writes on the chalkboard for example)? Do they read the slogans that flash on the screen during commercials ("Just do it" at the end of the Nike commercials)? How about the comics; or the headlines in the newspaper that someone left on the bus seat next to them, or the sign on the front of the bus telling its destination? Continue to brainstorm with students for a few minutes until they seem to understand that everything counts — the backs of cereal boxes, the jacket of a new tape or CD, the billboards on the street, *TV Guide* — everything. (This list and other lists brainstormed in the lesson will be used in later lessons, so they should be recorded in a way that can easily be reproduced for the class. A flip pad of large sheets that can be posted on the walls of the classroom might be helpful.)

2. Now challenge them to keep track for one day (from one class period to the same class period the next day) of all the times they read anything. (You might set up a little competition by inviting other classes to carry out the same activity and then comparing whole-class totals the next day.) Students can keep a running tally on small pieces of paper or index cards.

3. Begin the next class by asking all students for their totals (this can be done anonymously if you prefer). Add them all up and write the grand total on the board. Then stand back and admire it with everyone. Discuss with students the following:

- Did they find themselves reading things that they hadn't listed in yesterday's brainstorm?
- Why do they think there are so many things that need to be read?
- What are the advantages and disadvantages of getting information through reading (as opposed to television or radio or other nonprint sources)?
- What would their lives be like if they couldn't read?
- What would the world be like if there were no such thing as reading or writing?

4. Help students to think about the role that reading plays in adult life by doing one or more of the following activities:

- If possible, invite community members to your class to talk about when reading is useful to them in their jobs (a volunteer from a literacy program is a possibility).
- Alternatively, the students could write letters to celebrities as well as local people to ask about their reading habits. (A similar activity is proposed in the writing section, so students might write a single letter asking both questions.)

FIG.2-*Continued*

- Or, students might simply suggest particular jobs and list the sorts of reading those jobs might entail.

5. Next, help students think about how the reading they do in school compares to the kinds of reading they (and adults) do outside of school. Have students list the different kinds of reading they do in school (encourage them to think about all subjects, including math) and how this reading differs from the reading they do out of school. Probe their responses — even the simple and superficial ones. If, for example, students claim that the reading they do in school is boring and pointless, ask them to make some guesses about why it is assigned. Ask them to compare their school assignments with some of the reading examples that they tallied, as well as with the types of reading adults say they do for their jobs. Which kinds of reading are most similar to their school reading? You might also ask them what in general would make school reading more tolerable for them, as well as what they themselves might do to make it better. (Lesson 3 takes up this point more fully.)

Connections

Introduction Lesson 1.1: Why School?

Writing Lesson 3.1: Writing In School and Out

Follow-up Activities

1. Ask students to survey people in their community (storekeepers, police officers, teachers, parents, siblings) about what they read and why. Each week, ask two or three students to interview one person each. At the end of the week, they can report their findings, which can be recorded in a public place in the classroom.

2. If you teach in a self-contained classroom, take a few minutes each week to discuss reading in one subject other than language arts. Invite students to make connections and find differences in the purposes for reading in their various subjects.

FIG. 3 Excerpts from the "Expository Writing" chapter of the *Practical Intelligence for School* (PIFS) program.

LESSON 3.2 Using Past Experiences in Writing

Central Themes

Knowing Self
Knowing Process

Synopsis

Journal writing, discussion, and brainstorms to help students reflect both on their past experiences with writing and on the extracurricular interests and strengths that might help them in their writing.

What To Do

1. Provide students with a copy of the list of different types of writing that they brain-stormed on the first day (Writing, Lesson 1). Have them refer to it when answering the following questions (in their journals) about their past experiences with writing:

- What kind of writing do you like doing the most?
- What is the easiest or most enjoyable thing about doing that kind of writing?
- What kind of writing do you like the least?
- What is the hardest or least enjoyable thing about doing that kind of writing?
- When and where do you do your best writing?
- What circumstances make it hard for you to write?
- What do you like to write about?
- Do you prefer the writing assignments of one subject to those of other subjects?
- Is there a specific kind of assignment that you enjoy doing no matter what class it's for?

2. Spend a few minutes talking with students about the things they find easiest and hardest about writing. Are there trends across the class? Students will probably find that the writing they do outside of school is almost always easier and more enjoyable than the writing they do in school. This is understandable, given that the writing they do outside of class is usually voluntary and done to accomplish purposes that they see as relevant to their lives. Not having chosen their school assignments, and perhaps not understanding their importance in either the long or short run, students find them more work and much less fun.

3. Ask students to look through the first pages of their self-profile journals, where they listed all their strengths and interests when they were doing activities from the Introduction lessons. Explain that they can use these interests to help them become more interested in, and so improve at, writing assignments. They can incorporate their interests, experiences, and strengths into their assignments.

4. Give the class a report topic that comes from the normal curriculum for this week (something that you would have assigned anyway). The best topics for this exercise are fairly focused, but not so much as to prevent the students from shaping the topics to their thoughts. Write the report topic on the board. For the purposes of illustration, let's say that your class is currently studying colonial America.

FIG.3-*Continued*

5. Have each student think of at least three ideas for a writing assignment (e.g., a writ-ing project or report) about this topic that could incorporate a personal interest or area of expertise. (You might want to do a few out loud to help students get the hang of it.) Lists might include ideas like these:

- *Drawing.* Focus on a topic that lends itself to illustrations (types of dress or housing).
- *Computer/video games.* Design and write a manual for a role-playing game about life in colonial times; players might win or lose points depending on how well they survive the boat trip, make friends with natives, build weather-resistant homes, and so on.
- *Acting.* Write a play about some aspect of life in colonial America.
- *Skateboarding.* Report on transportation in colonial times (how people got around, how much people traveled, who traveled).
- *Making up stories.* Write some diary entries of a colonial teenager.
- *Medicine.* Write about illnesses and cures common in colonial America.
- *Food.* Write a few entries for a colonial cookbook; recipes could include ingredients available to colonial Americans, tips for how to store food, cooking devices.
- *Getting into trouble with the law.* Describe what people in colonial America could be punished for, as well as how they were punished.
- *Entertainment.* Write about what young people did for fun in colonial America.
- *Music.* Describe the music people liked then, and how people listened to it. (If the writing requirements are loose, students might also write a rap or song about an event like the Boston Tea Party or the Salem witch trials.)
- *Animals.* What animals did people in colonial America use for pets? What animals did they hunt, or did they fear? How were animals a part of their daily lives?

6. Ask some of the students to share their ideas from this exercise. For each idea, have the class brainstorm a few possible sources of information (besides a textbook or en-cyclopedia) that might be helpful to consult.

7. Ask if anyone had an interest that he or she just couldn't relate to the topic in any way. If so, see if the class as a group has any ideas. Consider: Are there certain things

that simply won't relate to the given topic? Are there ways of looking at an interest in a broader light that might make it relate better to the topic (as in the earlier skateboarding example)?

8. Discuss the advantages and disadvantages of choosing these more unusual topics.

Conclude by stressing that every assignment has a certain amount of leeway in how it may be approached. It's up to the student to work a personal interest into the topic to enliven the assign-ment for both the writer and the reader.

Connections

Writing Lesson 3.7: Choosing a Format

Reading Lesson 2.3: Personal Reading Profiles

Homework Lesson 4.6: Making It More Interesting and Personal

Follow-Up Activity

Have the students log in their journals for each writing assignment how they have in-corporated their interests or how they have otherwise helped themselves relate to the topic.

Comment on Conducting Ecologically Valid Research in Public School Classrooms

Before we describe the details of the methods and procedures used to select sites, teachers, and students and to implement the PIFS program, we wish to consider the limitations and constraints inherent in conducting ecologically valid research in public school classrooms. Our goals included, first, implementing the PIFS program in a diverse group of schools (urban, suburban, and rural) containing students of diverse socioeconomic levels from various racial and ethnic backgrounds. Second, we sought to work with typical teachers using typical classroom curricula and facing typical constraints upon their teaching imposed by their schools and administrations, as well as by students' parents. Our third goal was to implement the program for an entire school year and not simply for a handful of tightly controlled lessons in a laboratory experiment setting. In general, we wanted the implementation of PIFS to take place in an ecologically valid and naturalistic setting.

Because of these goals, we did not attempt to force randomly selected teachers from randomly selected schools to participate in our program. First of all, it is difficult if not impossible — and possibly even unethical — to demand that a principal and his or her teachers participate in a year-long curricular intervention. Instead, we met with principals and asked them to review our materials and decide if they believed participation would be advantageous for their schools. Once a principal had agreed, we solicited the names of interested teachers and met with these individuals. From this group of interested teachers, the control-group and experimental-group teachers were assigned, at this point randomly, with control teachers being told they would receive all of the experimental materials after the study was completed.

We selected teachers whose students had been grouped into heterogeneous classes at the end of the previous school year. These classes were created by the principals and administrators to be equivalent with regard to gender and achievement level — thus, each experimental class and paired control class in our study contained equivalent numbers of gifted, average, and learning-disabled students. We recognize that our method did not constitute true random selection of schools, teachers, and students, but given the commitment we required from teachers, it would have been impossible to obtain a fair evaluation of the potential of our program by forcing teachers to comply — nor could we or even the administrators of the schools have successfully pressured experienced, tenured teachers in this way.

The students in our study were not told they were part of an experiment; their teachers simply taught for one school year from the perspective we advocated and using the materials we supplied. Thus, in the eyes of the students, this was business as usual. The students were also unaware that the assessments used by the researchers were being used outside of the classroom the students knew only that these assessments were being used by their teachers to measure

FIG. 4 Excerpts from the "Homework Planning and Execution" chapter of the *Practical Intelligence for School* (PIFS) program.

LESSON 4.1 Purposes for Homework

Central Themes

Knowing Why

Synopsis

A brief discussion on the purposes for homework.

What To Do

1. Conduct a class discussion about the purposes for homework. You might want to use some (or all) of these questions:

- Why do we do homework at all?
- How does homework change as you get older? Why?
- What would school be like without homework?
- What can you learn from doing homework?
- Can you think of anything that you do that is like homework?
- Can you think of anything that adults do that is like homework?

2. Write students' answers on the chalkboard. After enough discussion time to elicit a variety of responses, ask students to review some of the main themes of the discussion. (They might do this by categorizing the class's responses.)

3. Ask students to identify among their responses the ones that indicate how home-work can be important for more than just learning subject matter. Emphasize that doing homework builds responsibility and skills (similar to practicing for sports or rehearsing for a play) that will be helpful in adult life.

4. Optional: Allow five minutes for students to write in journals about their most im-portant reasons for doing homework.

Connections

Introduction Lesson 1.1: Why School?

Testing Lesson 5.1: Why Tests?

Follow-Up Activities

1. Each time students do a homework assignment, ask them to write the goal at the top of their paper. (Once they have finished the assignment, you might also ask them to evaluate briefly how well they've met that goal.)

2. Return to this discussion (and to what the students have written in their journals) after completing the lessons in the Homework chapter — or in a couple of months, whichever is later. Ask students to think about if and how their ideas have changed about the purposes for homework.

FIG. 5 Excerpts from the "Preparing for and Using Feedback from Tests" chapter of the *Practical Intelligence for School* (PIFS) program.

LESSON 5.1 Why Tests?

Central Themes

Knowing Why

Synopsis

Discussions designed to encourage students to develop a sound understanding of the role of tests both in and out of school.

What To Do

1. How are tests used outside of school? Ask students if they ever get tested outside of school. Although their initial response might be a resounding no, press them a bit.

- Do any of you play sports? How do you know when you're getting good? Could games be like tests?
- Have you ever given a musical performance for which you had to prepare?
- What about the first time your parents or older siblings gave you a certain privilege or special task (watching younger siblings, taking care of a pet, staying home alone)?
- Have you ever been given a new chore or a job working for a neighbor?

Might these instances be tests of students' sense of responsibility and trustworthiness? Encourage them to come up with more test-like situations from their extracurricular lives and list these on the board. Virtually any situation that involves performing or demonstrating something that has been practiced can count. As you compile the list, discuss the following questions:

- What purposes did these tests serve?
- How did you tell whether your practicing had been done correctly?
- How did you know whether you had done well?
- Did you learn anything about how to do it better next time?

2. How are tests used outside of school in adults' lives? Ask the students to list occupations and roles (airline pilots, police officers, teachers, doctors, truck drivers) for which testing is necessary. You might also have students break up into groups and see which group can come up with the most occupations. Pick out a few of the occupations they have suggested and talk about the following questions:

- Why do we test these people?
- What do people in these occupations need to be good at (what would they need to be tested for)?
- What sorts of tests might people in these occupations be given to see if they have the necessary skills?
- What could happen on the job if these people were not tested?

You might also discuss with them the less formal tests that adults encounter every day. A teacher teaching a class, for example, is undergoing a kind of test, as is the firefighter confronting a burning house. The process of applying for a job is also like taking a test. Share with the students some of the everyday tests that you go through, how you get feedback about your performance, and how that feedback helps you to make plans for future situations.

3. How are tests used in school? Ask students to list the sorts of tests they take in school. Discuss with them a series of questions like the ones listed in item 2:

FIG. 5-*Continued*

- Why are tests used in school?
- What do tests tell teachers? Parents? Students? How can they use this information?
- Do some subjects lend themselves to testing more readily than others (math versus English; science versus art)?
- What would school be like without tests? (Ask for both pros and cons.)
- Would you feel uncomfortable not knowing what your grades and academic standing are?
- What might happen in school if students weren't tested?

4. Finally, ask students to compare the list of school tests with the two lists of outside school tests. How are the lists different? How are they the same? The most obvious distinction is likely to be that most of the tests on the school list are written, and customarily receive a number or letter grade. On the other hand, many of the tests on the outside-school list will be "performance" assessments, in which the test-taker actually demonstrates a particular skill and is judged accordingly (perhaps pass-fail).

5. Ask students to think about testing situations that involve both a pencil-and-paper component and a performance component: driving tests, CPR certification tests, auto mechanic exams, surgeons' evaluations. What does each half of the test tell about the test-taker? If they were only given written tests (or only performance tests), what would be left out? Then ask students to think about school pencil-and-paper tests in this light: What can and can't they measure? Get concrete examples here. A spelling test, for instance, might demonstrate how well you can memorize, but won't show how well you can use a dictionary. A timed math test might show how quickly you can add and subtract, but won't tell how resourceful you can be in figuring out how to do a puzzling problem.

There are three important points to make from all of these discussions:

- *Tests aren't simply a school phenomenon.* They are an omnipresent part of life, so it makes sense to think about how to do them well and how to use them to advantage.
- *Tests are sources of information, not just stamps of success or failure.* Just as an altimeter reading lets a pilot know whether his plane is too high or too low, performances on tests can tell students what they know well, what they need to put more work into, and what kinds of tests and test questions they're most likely to have trouble with or do well on. (Forthcoming lessons will help students get better at "reading the altimeter.")
- *Tests can provide important information about students' progress in acquiring certain skills and knowledge, but often they measure only a part of that progress.* Different kinds of tests yield different kinds and amounts of information about achievement; students vary in how they best display their knowledge: Some do well on in-class tests, others on take-home essays. Some shine on short-answer quizzes but can't perform well on long multiple-choice tests. No student is perfect, but neither is any test.

Connections

Homework Lesson 4.1: Purposes for Homework

Follow-Up Activity

Ask students to think how homework is a form of testing. Have them list the ways homework can work like a test and how they can benefit from it. Have them also list ways homework cannot serve the same purpose as tests — why it is not always a good substitute. Ask students if they would like to have take-home tests occasionally in place of in-class tests.

their progress, as was the case with all classroom (nonstandardized) tests.

During the PIFS implementation, the researchers observed and met with teachers to ensure that PIFS lessons were being covered, but we made no attempt to force teachers to read from prepared scripts of curricula, word for word. To attempt to constrain teachers in this way might have created an artificial environment which the teachers would have resisted (appropriately so) and to which the students would have responded negatively. Thus, the classrooms in which we worked were not tightly controlled experimental settings; they were typical public school classrooms. Our belief was that differences in implementation across teachers — an inevitable result of teachers' differing styles, methods of expression, attitudes, and preferences, as well as of students' individual needs — would average out on balance across the experimental and control classrooms throughout the study. We recognize that some teachers emphasized slightly more heavily certain curricular themes that they felt more comfortable with; however, we believed that our naturalistic approach was not only more ethical than a tightly controlled experimental approach, but also more true to the eventual conditions under which PIFS would be used in the world beyond our study, in the event that the curriculum came into broad circulation. Our study was a first attempt at evaluating PIFS. Future research can benefit from the success of this early work and possibly explore individual differences among teachers in greater detail.

In sum: our goal was to conduct ecologically valid research in public school classrooms, not tightly controlled artificial research in a laboratory setting. Thus, our goal and the methods necessitated by this goal require us to openly acknowledge the experimental limitations of our approach: (a) self-selection of teachers who were then randomly divided into experimental and control groups; (b) lack of pure random assignment of students by researchers (rather, use of typical, intact classes comprised of heterogeneously grouped students); and (c) lack of word-for-word scripted reading of curricular materials, but, rather, individual tailoring of lessons to fit teachers' and students' needs, styles, and preferences.

Sites

The PIFS curriculum was implemented two separate times over two consecutive years. This implementation took place at sites in Connecticut and Massachusetts within reasonable driving distance of the Yale and Harvard research teams. The schools were chosen to span a range of environments and populations, with differing profiles of academic needs, so that we could evaluate the usefulness of this intervention in inculcating practical intelligence under varying circumstances. (Details of the site and teacher selection procedures are discussed below under "Procedure.")

Connecticut sites. For both Year 1 and Year 2 of the implementation, three

public schools in Connecticut contributed two sixth-grade classes and one fifth-grade class to the study. These schools were in rural and predominantly White communities, with a mix of socioeconomic levels. Three other classes within the same school district served as control groups not receiving the program; two of these classes were in two schools that also contributed experimental classrooms. All six classrooms were self-contained. The classes averaged 15 students each.

Massachusetts sites. For the first year of the implementation, the Massachusetts sites also contained three public schools, which contributed two combined fifth/sixth-grade classes and four sixth-grade classes. The combined fifth/sixth-grade classes were self-contained classes in elementary schools. The sixth-grade classes were taught by a team of four teachers (one language arts, one reading, one math, and one combined science-and-social studies teacher). All Massachusetts sites were urban. The students were White, African-American, Asian, and Hispanic and came from a broad mix of socioeconomic levels. Two comparable sixth-grade classrooms in a nearby urban, middle-, and lower-middle-class community served as control groups. The class sizes ranged from 15 to 25 students.

In Year 2, the Massachusetts sites included three public schools in the Boston area, serving lower- and middle-class communities with student populations that are African-American, Hispanic, and White. One teacher implemented the curriculum in her self-contained classroom that included both fifth- and sixth-graders. At another school, a team of three teachers implemented the program with their shared sixth-graders. Matched control students were selected from these same schools, but did not receive the PIFS curriculum. In addition, the Harvard research team worked with teachers who implemented PIFS for sixth- through eighth-graders in a suburban school 25 miles north of Boston.

Participants

Connecticut sites. In Year 1, 87 students from northwestern Connecticut served as participants (43 girls and 44 boys). Their ages ranged from 10 to 13 years with a mean of 11.33 (SD = .62). Participants were fifth- (n = 29) and sixth-grade (n = 58) students. Forty-four participants (in three classes) received the PIFS program and 43 participants (in three classes) served as controls. The PIFS and control groups were equivalent with respect to age, sex, and achievement level. This information is determined by the school principals and administrators at the end of each school year and is used to group students for the upcoming school year into heterogeneous classes (two classes per grade) with equivalent numbers of gifted, average, and learning disabled students.

In Year 2, 106 students from the same region participated (51 girls and 55 boys). Their ages ranged from 9 to 12 years with a mean of 10.68 (SD = .64). Participants were fifth- (n = 42) and sixth-grade (n = 64) students. Fifty-four participants (in three classes) received the PIFS program and 52 participants (in

three classes) served as controls. The PIFS and control groups were equivalent with respect to age, sex, and achievement level (as a result of the class-grouping procedure described above).

Massachusetts sites. In Year 1, 109 students from the Boston, Mass., area served as participants (59 girls and 50 boys). Their ages ranged from 9 to 13 years with a mean of 11.15 (SD = .70) years. Participants were fifth- (n = 17) and sixth-grade (n = 92) students. Sixty-one participants (in three classes) received the PIFS program and 48 participants (in two classes) served as controls. One of the PIFS classes was team taught. The PIFS and control groups were equivalent with respect to sex and achievement. The PIFS participants were younger (M = 11.02 years old, SD = .83) on average than the control participants (M = 11.31, SD = .47), unequal variance $t(98) = 2.35$, $p < .05$, although this difference represented less than 3.5 months.

In Year 2, 212 students from the same area participated (105 girls and 107 boys). Their ages ranged from 9 to 13 years with a mean of 11.33 (SD = .82) years. Participants were fifth- (n = 20), sixth- (n = 156), seventh- (n = 26), and eighth-grade (n = 10) students. There were 144 PIFS participants (in six classes) and 68 control participants (in two classes). One PIFS class and one control class were team taught. The composition of the PIFS and control groups was equivalent with respect to sex and achievement level (again, classes consisted of mixed-ability groups of students of different levels of ability and achievement). The PIFS participants were older (M = 11.41 years old, SD = .88) on average than the control participants (M = 11.19, SD = .67), unequal variance $t(167) = 1.97$, $p = .05$, although this difference represented less than 3 months.[7]

Procedure

Background and selection procedures. To select schools, the researchers first consulted demographic data to create a master list of acceptable sites. Next, the school principals were contacted by letter and asked for a meeting. At this meeting, the researchers described the program and asked if the principal would allow her or his school to participate. If so, the principal was asked to canvass the teachers to create a list of interested faculty. The researchers then met with interested teachers and explained what would be required to participate in the program. The most interested teachers offered their names for further consideration. These teachers were then randomly divided and assigned to either the experimental or control condition, with the control teachers being informed that they could have the program to teach after the study was concluded.

In both implementation years, experimental-group teachers approached the

---

[7] Due to limitations inherent in conducting research in inner-city, urban schools containing challenged populations, the Harvard research team was unable to achieve a perfectly matched experimental-control group design in which the numbers of students and classes were the same for both conditions and in which the classes were drawn from the same school districts.

teaching of PIFS lessons as though these lessons were a normal part of the curriculum. Students were not told the material was "extra work" or "enrichment;" they were simply taught from the PIFS perspective rather than from the more traditional purely academic perspective. The assessments used to evaluate PIFS, which were administered to all students (experimental and control), were also used by teachers to evaluate their students' proficiency in reading, writing, homework, and test taking. The use to which the researchers put these assessments was unknown to the students, who viewed the assessments as simply a part of normal classroom procedure.

Year 1. Throughout the first year-long implementation, the researchers worked closely with all teachers implementing PIFS. Teachers were introduced to the curriculum over the summer. The Connecticut teachers met with the Yale research team regularly over the summer to review the curriculum and refine the implementation plan. Similarly, the Massachusetts teachers met with the Harvard research team over the summer to review the curriculum and to make plans for the school year. During the school year, each teacher (or team of teachers, in the case of the Massachusetts middle-school site) met with a researcher approximately once a week to review the lessons taught; to consider the successes, problems, and questions that had arisen; and to plan how best to continue with the curriculum.

In Year 1, all teachers began the PIFS curriculum by working through the lessons in the introductory booklet. Once these introductions were completed, we encouraged teachers to use subsequent booklets (and lessons within the booklets) in the order that best met the needs of their students and their standard curriculum. For instance, one teacher felt that her students needed first to develop the organizational skills addressed in the homework booklet and so began by leading her class through all six lessons of that booklet. Another teacher wanted her students to practice note taking and reading skills as they carried out their fall social studies project. She began with a mix of lessons from the reading and testing booklets. Although such variability in the implementation of the program imposes inevitable constraints on generalizations from the present results, we felt this effect was more than offset by gains in ecological validity. Plus, in working with teachers over the course of a year-long program, it was necessary to be flexible in our demands, since each teacher had her or his own way of conducting classes and since teachers resist rigid outsiders who attempt to tell them exactly how to do their jobs. Through weekly meetings with and observations with teachers and classes, the researchers ensured that the PIFS curriculum was being taught appropriately by all teachers, despite the individual teachers' customization of the lessons to meet their own and their students' needs.

Teachers typically focused on one lesson per week. In the days following that lesson, teachers referred back to the particular skill or idea presented in that lesson as they worked with students on their standard curriculum. For example, on the day after presenting the lesson about recognizing differences in various

kinds of reading, one teacher asked her students to compare their science reading (a textbook passage) to the story they had read for homework: How were the two different? Did they read the textbook differently than the story? During a math lesson two days later, she asked the students to discuss why they did not read as much in math as they did in science or language arts. Such repetition in a variety of contexts fostered the students' deep and flexible grasp of PIFS ideas.

Year 2. In the second year of implementation, the researchers worked with a revised and streamlined version of the curriculum, which had benefitted from extensive comments and feedback generated by the teachers as they implemented the program in Year 1. By Year 2, the teachers were quite familiar with the program. Thus, throughout Year 2, the teachers worked more independently and attended fewer meetings with researchers.

In Year 2, the Massachusetts and Connecticut teachers followed somewhat different implementation procedures. The Massachusetts teachers implemented the PIFS program with an emphasis on customizing the PIFS lessons more than they had done in the previous year: They focused on the ideas or lessons that they believed to be most important for their students. Their instruction included infusing PIFS ideas into regular schoolwork as well as doing stand-alone lessons. The Connecticut teachers, on the other hand, emphasized broader coverage of the curriculum than they had been able to manage in the first year (an enrichment approach). Connecticut teachers typically covered one or two complete PIFS lessons per week, returning in the context of other subjects to the ideas introduced in the lessons. The Connecticut classes thus received about 90% of the lessons.

In general, in Year 2, all of the teachers took control of the PIFS curriculum and became active leaders and adapters of the program, with the goal of meeting their students' needs. Although we acknowledge that this enhanced level of teacher control meant that we had less precise say in how PIFS was taught, we believed that the condition of greater teacher control more closely mimicked how our program would be used by teachers once our formal study and evaluations had ended. Thus, we believed that the data generated during Year 2 would be an ecologically valid test of the effects of the PIFS program under more typical conditions of use.

Assessments

Description: In Year 1, we developed pretests and posttests designed to assess the quality of students' practical knowledge in each of the four focal areas (reading, writing, homework, and testing). All of the tests were based on the kinds of tasks students are typically asked to do in school to make them fair to students not exposed to the curriculum.

We used two reading assessments, one based on a factual passage (e.g., "The Intelligence of Elephants") and the other on a passage of fiction (e.g., an allego-

ry about a talking elephant). Students read the passage and answered questions about their general understanding, their thinking processes while reading, the parts they found easy or hard to understand and why, how they would study for a test on the passage, the differences between the two passages, and so on. Most of the questions were open-ended. Each reading assessment took approximately 50 min.

The writing assessment involved two parts, each of which lasted 50 min. The first part of the pretest asked students to write a composition describing in detail a place they knew well. For the posttest, students described a person they knew well. Following the writing, students answered questions about their writing process — what was easy or hard, how they got their ideas and organized their presentation, and what their teacher's reaction might be (e.g., "What will your teacher like MOST about your composition? LEAST?"). For the second part of the writing assessment, students revised the composition they wrote on the first part. They then reflected on the revision process, indicating the parts they had added or deleted, explaining those changes, and predicting what the teacher might and might not like about the piece.

For the homework assessment, teachers began by assigning a typical homework task that was relevant to the standard curriculum. Students were not told that the assignment was anything out of the ordinary. The following day, students were given a sheet of questions asking them to reflect on the processes involved in doing the homework. The questions asked the students to write about why they thought the homework was assigned, how they gathered information to do the work, what was easy or hard (and why), how they dealt with the hard part, how they could make this homework more interesting to do, and how their teacher would assess the assignment.

The testing assessment was built into the other assessments in two ways: First, two of the three assessments were treated as tests when administered to the students; and second, some of the assessments included questions about how the student would take notes and study for a test on a given passage or topic (e.g., "Imagine that you will have a quiz on 'The Intelligence of Elephants' next week . . . Use the space below to make some notes on the article that will help you prepare for the quiz. Now describe what you would do to study for the quiz").

In Year 2, we developed refined versions of our Year 1 assessments that required fewer administration sessions (in response to the fact that the time requirement had presented a problem in Year 1). Year 2's assessments were completed in four days rather than five or six. To achieve this goal, we reduced the two-part assessments to single assessments that could be given in a single class period. This revision reduced the time burden on students and teachers associated with evaluating the curriculum and also reduced the likelihood that the researchers would be left with half-completed assessments due to absences (particularly common in the urban schools). Each individual assessment still

took 50 minutes.

As just noted, the reading assessment was reduced from two parts to one. Students read two brief passages, one fiction and one nonfiction on similar topics (e.g., the first battle of the American Revolution in 1775), and then answered questions that asked them to compare and contrast the two passages (e.g., "List at least three ways in which these two passages are the same. List at least three ways in which these two passages are different"). Students also described a personal experience that related to the topic, listed strategies for getting through a difficult reading assignment, and identified situations outside of school in which reading plays a crucial role.

The writing assessment also was reduced to a single task followed by questions. Students wrote a letter to their teacher convincing him or her not to give homework on weekends and answered the following questions about their work: What will their teacher like or not like about their letter? How do they overcome problems when they have trouble writing? How would they revise the letter if they had time? (e.g., "Name three or more things you could improve about your letter if you had time to revise it").

The testing assessment consisted of two 50-minute parts. The first part measured students' use of strategies on a standardized test: The assessment in-volved a lengthy set of multiple-choice questions of varying degrees of diffi-culty and worth varying point values. The most difficult questions were not always worth the most points (the point values were indicated next to each question). Students were told at the beginning that they would not have time to complete all questions, but that they should try to earn as many points as possible. (Due to time constraints, the first part of the testing assessment could be used only in the Connecticut schools). The second part of the testing assessment (given in all schools) asked students to describe their study strategies for a language arts test, to compare these strategies to ones they would use to study for a social studies test (e.g., "Name three strategies you should use to study for a social studies test that you should NOT use to study for an English test; Name three strategies you should use to study for BOTH tests"), and to contrast how they would study for a multiple-choice versus an essay test. Students also were asked how they would overcome difficulties if they got stuck on a test, and they were asked to identify situations in which adults have to take tests.

In the homework assessment, students considered the case of "Mike," a hypothetical student who has a variety of overnight assignments and small projects that need to be completed in a week (e.g., read an English story, define words in it, and write a paragraph about it; solve 10 math problems; read about a group of Native Americans in the text and then choose another group of Native Americans not covered in the text and write a report about it; and do an experiment at home consisting of placing cheese in a jar and observing it for four days and write a one-page report about it).

The students were asked to spell out what Mike should do each day to get all

his work done on time, given the other constraints on his time (e.g., "Starting after school at 3:30, Mike spends time with friends or on the phone; just before dinner, he watches his favorite TV show; after dinner until bed at 10, he usually does homework; on Monday he has to watch his baby brother from 4 to 9; on Thursday afternoon he has a basketball game"). Students also are told about how Mike works (e.g., "Math is his best and favorite subject; he does OK in science; English and social studies are the hardest subjects for him; if he likes the work he finishes with no problem; if he doesn't like it, he gets distracted easily and has trouble finishing"). Students then were asked to describe how they would approach the variety of homework assignments by filling in a planning grid for Monday through Friday stating "What Mike Should Do" and "Why" for each day. For the balance of the homework assessments, students described the various resources they would use for completing a report on Native Americans, and they described how they might incorporate their own interests within their report.

Administration. The pretests were given in October over the course of a week (and sometimes also on Monday of the subsequent week), approximately one assessment each day. The posttests were given in the same manner in June.

Scoring. Yale and Harvard researchers collaboratively developed a 5-point rating scheme for each test, in which a 1 was defined as "No Ability-very poor or essentially blank," a 2 as "Low Ability-poor but shows effort," a 3 as "Some Ability — generally good answer," a 4 as "Good Ability — very solid effort and performance," and a 5 as "High Ability — excellent in every respect." Each score was further defined with specific examples of student work that corresponded to that score. Separate scoring keys and sets of student work were developed for each academic and practical measure, reflecting the different emphases of academic and practical thinking skills. The same levels of relative and absolute performance were associated with the scores from 1 to 5 for both academic and practical measures, in both Years 1 and 2, at both the Connecticut and Massachusetts sites. Once the scorers had studied the anchors for the scale, they referred back during scoring to a key that contained the student examples corresponding to each score. Assessments were scored by three independent raters blind to student group membership (students' names, classes, teachers, and schools were masked and pretests and posttests were intermixed).

To clarify the details of the scoring procedure, consider the example of the writing assessment for Year 2. The task was to "write a letter to your teacher explaining why students should not have homework on weekends. Try to write a letter that will make your teacher agree with you." Following the composition, one of the questions (measuring practical intelligence) that the student was asked states, "Name three things your teacher will like about the way you wrote your letter." How was a practical-writing score arrived at for this particular question? Scorers looked for the degree to which the student had a sense of how to make an appropriate argument in writing and for how accurately the student

judged her writing. Scorers began by assigning a general score of high, middle, or low to each individual answer. For this question, a high score went to students who stated: "I showed the kid's point of view," "I explained my thoughts clearly," and "I went straight to the point." A middle score went to students who stated: "It was neat," "I wrote many reasons," "I put all my punctuation marks in," and "My vocabulary." A low score went to students who stated: "Do it right," "I don't know," and "Ask the teacher."

In general, for all of the practical questions on all of the assessments, "low" was assigned when students gave no response or an inappropriate or irrelevant response or wrote something such as "I don't know." "Middle" was assigned when a student's response consisted of a "verbal molecule," a simple statement that is appropriate, but obvious or stereotypical and indicates no deep thought or detailed explanation (such as "I would ask for help"). "High" was assigned when the response was original and personally appropriate, showing a knowledge of oneself, why the question was asked, how to think effectively about the question, seeing how one question differs from the others, and so on. After all practical-intelligence questions were scored in the same manner, the relative numbers of high, middle, and low scores were tallied across all writing questions. The tallying of all high, middle, and low scores was then used to compute the score on the 1-to-5 scale, which was done as follows: 1 meant that all responses were low, with the exception of at most one middle; 2 meant that roughly half of the responses were low and half middle; 3 meant that responses averaged to middle (allowing for roughly balanced low and high); 4 meant that roughly half the responses were middle and half high; and 5 meant that all responses were high, with at most one exception. Another way to think about this 1-to-5 scale is in terms of its being anchored from No Ability to High Ability (as described above).

Turning now to the scoring of the academic-intelligence portion of the writing assessment, the academic-intelligence score for writing was based on the actual letter the students composed, which was evaluated for format, language, and mechanics. Scorers looked for the degree to which students understood the purposes for giving and completing homework, wrote well-substantiated arguments with valid reasons, noted the benefits resulting from not giving homework and why, and wrote with clarity and good organization. Scores on the 1-to-5 scale were assigned, with a 1 being poor or unsatisfactory (the letter was poorly written with numerous grammatical errors and does not present a solid argument) and a 5 being excellent (the letter was organized and clearly written with persuasive and realistic arguments that explained points well). Intermediate scores corresponded to intermediate-quality answers. (As described above, elaborate answer keys with specific examples of student work corresponding to each score were used in all scoring.)

Consider one other example, that of the homework assessment for Year 2 (measuring practical intelligence). The task was to design a plan to accomplish a

week's worth of homework for a hypothetical student named Mike ("What Mike Should Do" and "Why" had to be filled in for each day). The assignments were clearly described: But how would Mike be able to accomplish them, given the other demands on his time (such as babysitting and playing basketball, also described)? The answers were scored by looking at the student's sense of the process of doing homework, as indicated by how well the plan takes into account all constraints upon Mike's time; whether the student's explanations of her or his choices were appropriate and detailed; and how well the different elements of the plan fit together.

For example, a high response included the following: "Monday: Set up science before babysitting — Have to start experiment. Do math while babysitting — Math can be done while babysitting cause it's easy. Read story from 9 to 10 — Gotta start English cause it's hard. Tuesday: Check cheese after dinner and take notes — Have to check science every day. Write English paragraph — English is due Wednesday. Read social studies and pick group — Social studies is big assignment so have to start early [and so on for Wednesday through Friday, completing all assignments on time.]" Middle and low scores were given for correspondingly weaker answers.

Before turning to the quantitative results, let us review all of the types of scores that were used in this research. In Years 1 and 2, for the reading and writing assessments, students received both a practical and an academic score. The practical score reflected performance on the practical-intelligence questions; the academic score reflected performance on the academic-intelligence questions.

In Years 1 and 2, for the homework assessment, students received only a practical score, because the nature of the questions (which involved reflecting about an actual homework assignment in Year 1 and how to plan to accomplish a week's worth of homework in Year 2) did not provide a basis for rating the academic quality of a sample of actual homework. (Academic scores for the actual teacher-generated homework assignments in Year 1, as graded by the classroom teachers, yielded incomparable data across classrooms and were not used. In other words, different teachers gave very different types of assignments that could not be comparably graded across classrooms: some classes did compositions, some did short answers to questions about readings, some did project work, etc. The actual homework assignments were thus used in this research solely as triggers of student reflection about their homework techniques.)

Regarding the academic- and practical-testing scores, no explicit academic-testing rating was given in Year 1 because the academic scores in reading and writing (i.e., the "academic summary" score) were equivalent to a general academic score for testing. (In other words, the reading and writing assessments were tests which yielded academic scores representing measures of test-taking ability). In Year 2, in the Connecticut schools, an explicit measure of academic-testing ability was administered, yielding an explicit academic-testing score. In Year 1, scores for practical-testing ability were based on answers to questions

regarding how a student would study for specific tests (these questions were included in the reading and writing assessments). In Year 2, scores for practical testing were based on students' performance on the testing assessment described above, in which students compared and contrasted their study strategies for various types of tests in various content areas.

Results and Discussion of Quantitative Data

Below we present results for two separate years of curriculum implementation and for two separate sets of schools, the Connecticut and the Massachusetts sites. First, we describe Years 1 and 2 results for the Connecticut sites ("Study 1"), and second, we describe Years I and 2 results for the Massachusetts sites ("Study 2"). The data were viewed as representing two separate studies because of the many differences between the Connecticut and Massachusetts schools (e.g., rural versus inner-city urban and White versus racially mixed) and because of the differing implementation procedures followed in the two states as a result of distinct academic needs and scheduling demands. Academic-intelligence measures (described above under "Assessments") consisted of assessments of reading, writing, and testing ability. We also computed an academic-intelligence summary measure representing the mean of reading and writing scores. Practical-intelligence measures (also described above) consisted of reading, writing, homework, and testing ability. We also computed two practical-intelligence summary measures — one representing the mean of reading and writing scores ("practical-2," comparable to the academic summary score) and the other representing the mean of reading, writing, homework, and testing scores ("practical-4," representing all practical-intelligence measures).

Before describing the quantitative results, let us reiterate the limitations of our approach: (a) self-selection of teachers, who were then randomly divided into experimental and control groups; (b) lack of pure random assignment of students by researchers (rather, use of typical, intact classes comprised of heterogeneously grouped students); and (c) lack of word-for-word scripted reading of curricular materials, but, rather, individual tailoring of lessons to fit teachers' and students' needs, styles, and preferences.

Study l: Connecticut Site Results

Rating reliability. Practical and academic performance ratings showed good interrater reliability for both Year 1 and Year 2. The median reliability for Year 1 was .85 (range = .66 to .91). The median reliability for Year 2 was .86 (range = .82 to .92).

Means and standard deviations. The mean practical and academic scores for Year 1 and Year 2 are presented in Table 1. The Year-1 means for individual variables at pretest and posttest ranged from 2.68 (*SD* = .47) to 3.73 (*SD* = .74) on the 5-point rating scale. For Year 2, the means of the same variables ranged

from 2.29 (*SD* = .70) to 4.09 (*SD* = .69). The Year-2 academic testing mean, based on a (very challenging) 100-point test, ranged from 35.06 (*SD* = 12:69) to 50.01 (*SD* = 13.37). The summary variables, also presented in Table 1, show a range of means and standard deviations that is similar to that of the individual variables.

Gender differences. Although we had no gender-related hypotheses, we examined the data for gender differences because these differences, where they exist, are of interest to some readers. For Year 1, girls tended to score higher than boys at pretest and posttest. Several of these gender differences were significant. For example, on the homework-practical scores the means at pretest were 3.20 (*SD* = .60) for girls and 2.77 (*SD* = .75) for boys, $F(1, 77) = 11.07$, $p < .01$. The posttest means were 3.72 (*SD* = .78) for girls and 3.17 (*SD* = .72) for boys, $F(1, 77) = 10.69$, $p < .01$. However, there were no significant differences between boys and girls on pretest-posttest improvement and no significant interactions between gender and the PIFS treatment.

For Year 2, similar to the Year-1 results, girls tended to score higher than boys at pretest and posttest. Several of these gender differences were significant (academic writing and all practical variables). There were no significant gender main effects on pretest-posttest improvement and no interactions between gender and the PIFS treatment.[8]

Academic-intelligence variables. Analyses of covariance were conducted on all academic- and practical-intelligence variables using the pretest score for each measure as the covariate and comparing the fall-to-spring score changes for the PIFS and control group students. The resulting ANCOVA *F* values appear in Table 1 in the rows titled "Comparison of Differences." Data are reported separately for Years 1 and 2 of the study. The individual *t* tests conducted on the fall-to-spring changes *within* the PIFS or control conditions are also reported in Table 1 (for both the PIFS and control conditions) in the rows labeled "Difference *t*."

In general, for Year 1, there were pretest-to-posttest increases for both the PIFS and control groups. However, the PIFS group was not significantly different from the control group at posttest on the academic-intelligence variables. For Year 2, the PIFS and control groups were approximately equal at pretest. Both the PIFS and control groups increased at posttest. The PIFS group gains were significantly greater than the control group gains, showing beneficial effects of the PIFS curriculum on each type of academic work.

For the academic summary score (based on reading and writing ability), the covariate-adjusted posttest improvement was .47 points greater for the PIFS group compared to the control group, which is a 17% larger gain (above the sample's pretest mean) for the PIFS participants. For the academic testing score, the covariate-adjusted posttest improvement was 6.55 points greater for the

[8] The detailed data and analyses of gender differences are available from the first author.

TABLE 1
Study 1: Connecticut Site Means and Standard Deviations with Significance Tests (n = 193)

Group	Variable	Reading practical	Writing practical	Homework practical	Testing practical	Reading academic	Writing academic	Testing academic	Practical-2 summary (R, W)	Practical-4 summary (R, W, H, T)	Academic summary (R, W)
YEAR 1 (n = 87)											
PIFS (n = 44)	Pre	2.70 (.69)	2.76 (.56)	2.77 (.63)	2.89 (.54)	2.81 (.66)	3.08 (.81)	—	2.73 (.48)	2.78 (.49)	2.94 (.65)
	Post	3.68 (.69)	3.07 (.86)	3.53 (.86)	3.55 (.58)	3.33 (.73)	3.18 (.73)	—	3.38 (.53)	3.45 (.52)	3.26 (.56)
	Diff t	7.69***	4.06***	6.32***	9.20***	4.44***	<1.00 n.s.		9.10***	12.09***	3.72***
Control (n = 43)	Pre	3.22 (.65)	2.68 (.47)	3.20 (.73)	3.40 (.55)	3.22 (.86)	2.83 (.66)	—	2.95 (.45)	3.13 (.48)	3.02 (.64)
	Post	3.73 (.74)	3.02 (.70)	3.37 (.73)	3.51 (.80)	3.03 (.68)	3.03 (.68)	—	3.39 (.69)	3.41 (.63)	3.21 (.62)
	Diff t	5.66***	3.33**	1.63 n.s.	1.46 n.s.	1.25 n.s.	1.93 n.s.		5.44***	4.57***	2.21*
Comp. of diffs.	F	6.64*	<1.00 n.s.	8.41**	12.65***	<1.00 n.s.	<1.00 n.s.		2.21 n.s.	15.16***	<1.00 n.s.
YEAR 2 (n = 106)											
PIFS (n = 54)	Pre	2.75 (.91)	2.49 (.75)	2.53 (.88)	2.39 (.77)	3.17 (.98)	2.48 (.86)	35.06 (12.69)	2.61 (.74)	2.50 (.72)	2.81 (.82)
	Post	3.54 (.61)	3.31 (.75)	3.35 (.91)	3.08 (.77)	4.09 (.67)	3.16 (.84)	50.01 (13.37)	3.40 (.64)	3.28 (.64)	3.58 (.70)
	Diff t	7.52***	9.89***	6.99***	6.23***	6.38***	6.67***	8.42***	11.99***	14.32***	8.14***
Control (n = 52)	Pre	2.61 (.78)	2.60 (.66)	2.93 (.86)	2.29 (.70)	3.09 (.94)	2.58 (.58)	36.78 (11.25)	2.62 (.67)	2.67 (.65)	2.88 (.69)
	Post	2.95 (.79)	2.82 (.84)	2.86 (.76)	2.61 (.61)	3.58 (.72)	2.68 (.87)	44.46 (15.70)	2.89 (.75)	2.85 (.64)	3.15 (.70)
	Diff t	3.07**	2.61*	1.00 n.s.	3.39**	3.70***	1.08 n.s.	3.16**	3.66***	3.24**	2.99**
Comp. of diffs.	F	19.37***	25.33***	27.89***	10.36**	13.63***	16.49***	5.71*	29.98***	60.89***	17.44***

*p < .05; **p < .01; ***p < .001.

PIFS group compared to the control group, which is an 18% larger gain for the PIFS participants.

One note concerning summary scores: For summary scores (academic, practical-2, and practical-4), it should be noted that if a participant was missing one or more scores (e.g., homework-practical), then the summary score was the mean of the remaining scores. The summary score, as it now stands, can be viewed as the best available measure of the academic or practical effect of PIFS (vs no PIFS) because these scores use all of the available data for each participant. (However, a subnote is that for academic Year 2, testing-academic is not part of the academic summary score because it would make this score conceptually different from Year 1 scores and Massachusetts' academic scores for which no academic testing measure was collected.)

Practical-intelligence measures. For Year 1, there were pretest-to-posttest increases for PIFS and control groups. The PIFS group gains were significantly greater than the control group gains on reading, homework, testing, and the practical-4 summary scores (representing the mean of reading, writing, homework, and testing). On these variables, the PIFS participants began lower than the controls at pretest but PIFS participants overcame this initial difference and matched the posttest gains by controls. For the practical-4 summary score, the covariate-adjusted posttest improvement was .34 points greater for the PIFS group compared to the control group, which is an 11% larger gain over the sample's pretest mean for the PIFS participants.

For Year 2, the PIFS and control participants started equal at pretest and showed pretest-to-posttest increases. The PIFS group gains were significantly greater than the control group gains for all parts of the curriculum. For the practical-4 summary score, the covariate-adjusted posttest improvement was .56 points greater for the PIFS group compared to the control group, which is a 21% larger gain above the sample's pretest mean for the PIFS participants.

Teacher (or class) differences. As expected, there was some variability between teachers (or individual classes), presumably due to different instructional emphases and patterns of learning in different classes. In Year 1, for example, one PIFS teacher's class showed relatively large pretest-posttest gains on the homework-practical score (an increase of 1.60) compared to the other teachers' classes (which had an average increase of .37). However, this same teacher's class showed relatively small reading-practical gains (.55 compared to the other teachers' classes which averaged 1.17). For the PIFS teachers (classes), this variability was significant (at the $p = .05$ level) for reading-academic, writing-academic, reading-practical, and homework-practical scores. Control teachers also showed significant variability on reading-academic, academic-summary, reading-practical, homework-practical, testing-practical, and both

---

[9] Due to the space required to present these data in detail, and due to the fact that we do not interpret the teacher (class) differences beyond simply stating that some significant differences were observed, we do not present the pretest and posttest means, standard deviations, and comparative analyses for every teacher here. These data are available from the first author.

practical-summary scores (practical-2 and practical-4).[9] Perhaps the most important finding with regard to individual teachers (classes) was that PIFS teachers (classes) showed significantly greater gains from pretest to posttest, compared to control teachers, for all practical measures except writing.

For Year 2, similarly to Year 1, there was some variability between teachers (or classes). For the PIFS teachers (classes), this variability was significant only for homework-practical and testing-practical scores. For control teachers, this variability was significant for reading-academic and academic-summary scores. All PIFS teachers (classes) showed pretest to posttest gains for all parts of the curriculum.

It should be noted that we cannot distinguish teacher differences from class differences; each teacher has a certain selection of students (a class). The "teacher" differences may simply represent differences due to the samples of students that form each class. For example, one class of students may have been strong on homework regardless of who taught the class.

Student evaluation form. The student evaluation form was used only at the Connecticut sites. It was designed to be completed by teachers, and it described how well each child performed both generally and specifically with regard to PIFS skills. The form asked the teachers to rate each student on a 1-to-9 scale in terms of her or his grades, general behavior, and attendance. It then asked teachers to evaluate the extent to which each student recognized and incorporated personal strengths into schoolwork, compensated for weaknesses, questioned the purpose of school, made connections between subjects and ideas, used different approaches and a variety of resources, persevered with work, organized time, and sought feedback.

The student evaluation form results showed that PIFS teachers observed significantly more improvement in their students overall than did control teachers from October to June. The PIFS students' mean rating moved from 6.24 (Fall) to 7.16 (Spring), an increase of .92 on a 1-to-9 scale. The control students' mean rating moved from 5.55 (Fall) to 5.82 (Spring), an increase of .27 units. The relatively higher PIFS gain was significant, $t(68) = 3.50$, $p < .001$.

PIFS teachers observed significantly greater improvements than control teachers in the following areas: recognizing and making use of personal strengths and interests, noticing differences across subject matters, using different approaches in different subjects, using a variety of resources, persevering in the face of difficulties, organizing time and materials, and seeking and using feedback. There was no significant change in grades or general behavior in class. PIFS students showed less improvement than control students in attendance and the tendency to speculate about the purposes of studying particular topics or doing particular tasks. In general, however, the trend was positive for

[10] We acknowledge the limitations inherent in obtaining and comparing ratings of students generated by teachers who are participating in a study versus those of teachers who are not participating in a study. Obviously, demand characteristics could have affected the teachers' ratings.

the PIFS students as compared to the control students.[10]

Study 2: Massachusetts-Site Results

Rating reliability. Practical and academic performance ratings showed adequate interrater reliability for Year 1. The median reliability was .68 (range = .50 to .85). For Year 2, before full-scale scoring was undertaken, 20% of the assessments were scored by a second judge to establish reliability. Once the raters agreed on 80% of their scores, independent scoring commenced.

Means and standard deviations. The mean practical and academic scores for Year 1 and Year 2 are presented in Table 2. The Year-1 means for individual variables at pretest and posttest ranged from 2.01 (*SD* = .57) to 3.22 (*SD* = .78) on the 5-point rating scale. For Year 2, the means of the same variables ranged from 1.37 (*SD* = .49) to 2.82 (*SD* = .59). The summary variables, also presented in Table 2, show a range of means and standard deviations that is similar to that of the individual variables.

Gender differences. In Year 1, girls tended to score higher than boys at pretest and posttest. The differences were significant for the writing-academic, academic-summary, reading-practical, homework-practical, testing-practical, and the practical-summary scores. There also were significant gender main effects for pretest-posttest improvement favoring girls over boys on writing-academic, academic-summary, testing-practical, and practical-summary variables. There were significant interactions between gender and PIFS treatment for reading-practical, writing-practical, and the practical-summary scores.

In Year 2, girls tended to score higher than boys at pretest and posttest. The differences were significant for the academic variables, writing-practical, testing-practical, and the practical-summary scores. There also were significant gender main effects for pretest-posttest improvement favoring girls over boys on reading-academic and homework-practical variables. There were no significant interactions between gender and PIFS treatment.[11]

Academic-intelligence measures. For Year 1, both PIFS and control students tended to show pretest-posttest increases. As compared to the control group, the PIFS group showed a significantly greater increase for writing; the covariate-adjusted posttest improvement was .44 points greater for the PIFS group compared to the control group, which is an 18% larger gain (over the sample's pretest mean) for the PIFS group. The control group began higher than the PIFS group (significantly for reading and for the summary score), and the posttest position of the groups depended on the curriculum component: For the writing-

---

However, the student evaluation forms comprised the only data in this study that were teacher-generated. The bulk of our data consisted of assessments that were scored by individuals who did not participate in this research; who were blind to the hypotheses; and who did not know the students' names, classes, schools, or teachers. Hence, even if we ignore the teacher ratings, our conclusions are unaffected.

[11] The detailed data and analyses of gender differences are available from the first author.

TABLE 2
Study 2: Massachusetts Site Means and Standard Deviations with Significance Tests (n = 321)

Group	Variable	Reading practical	Writing practical	Homework practical	Testing practical	Reading academic	Writing academic	Testing academic	Practical-2 summary (R, W)	Practical-4 summary (R, W, H, T)	Academic summary (R, W)
YEAR 1 (n = 109)											
PIFS (n = 61)	Pre	2.20 (.64)	2.01 (.57)	2.38 (.46)	2.02 (.58)	2.53 (.82)	2.39 (.72)	—	2.11 (.57)	2.13 (.52)	2.45 (.69)
	Post	2.47 (.55)	2.50 (.64)	2.48 (.58)	2.38 (.59)	2.92 (.71)	2.92 (.62)	—	2.46 (.53)	2.44 (.48)	2.88 (.60)
	Diff t	3.56***	6.17***	1.22 n.s.	4.52***	3.27**	6.10***		5.82***	5.72***	5.41***
Control (n = 48)	Pre	2.64 (.69)	2.29 (.57)	2.50 (.63)	2.27 (.63)	3.02 (.73)	2.60 (.74)	—	2.46 (.54)	2.42 (.51)	2.81 (.65)
	Post	2.52 (.73)	2.24 (.74)	2.23 (.56)	2.08 (.69)	3.22 (.78)	2.60 (.74)	—	2.36 (.68)	2.26 (.60)	2.91 (.64)
	Diff t	1.38 n.s.	<1.00 n.s.	2.64*	2.28*	1.51 n.s.	<1.00 n.s.		1.29 n.s.	2.52*	1.10 n.s.
Comp. of diffs.	F	3.76 n.s.	12.26***	6.94**	17.21***	<1.00 n.s.	14.73***		12.66***	22.38***	2.03 n.s.
YEAR 2 (n = 212)											
PIFS (n = 144)	Pre	2.30 (.75)	2.12 (.78)	2.44 (.76)	2.16 (.82)	1.95 (.72)	1.78 (.60)	—	2.23 (.70)	2.26 (.65)	1.90 (.58)
	Post	2.82 (.59)	2.54 (.76)	2.73 (.71)	2.66 (.83)	2.20 (.53)	2.01 (.60)	—	2.71 (.61)	2.71 (.59)	2.14 (.50)
	Diff t	7.20***	6.16***	4.29***	5.80***	4.05***	4.44***		8.37***	9.47***	5.59***
Control (n = 68)	Pre	1.88 (.73)	1.74 (.67)	1.95 (.59)	1.63 (.66)	1.50 (.60)	1.37 (.49)	—	1.78 (.60)	1.80 (.50)	1.39 (.46)
	Post	2.34 (.71)	2.23 (.76)	2.17 (.81)	1.92 (.80)	1.67 (.54)	1.49 (.57)	—	2.26 (.69)	2.13 (.60)	1.58 (.49)
	Diff t	4.70***	4.70***	2.20*	2.69**	1.93 n.s.	1.55 n.s.		5.51***	5.04**	3.02**
Comp. of diffs.	F	10.13**	1.09 n.s.	8.11**	14.93***	25.54***	12.37***		8.44**	18.62***	22.50***

*p < .05; **p < .01; ***p < .001.

academic score, the PIFS group was higher at posttest, for reading-academic the control group remained higher, and for the academic-summary variable there was no significant difference between the PIFS and control groups at posttest.

For Year 2, the PIFS group showed significant pretest-posttest improvements. The control group did not change significantly, except for the academic-summary score. The PIFS group began significantly higher than the controls and ended significantly higher. Significantly greater improvement occurred for PIFS on reading, writing, and the academic-summary score. For the academic-summary score, the covariate-adjusted posttest improvement was .33 points greater for the PIFS group compared to the control group, which is a 19% larger gain (over the sample's pretest mean) for the PIFS group.

Practical-intelligence measures. In Year 1, the PIFS group showed pre-test-posttest gains. The control group declined from pretest to posttest, with significant declines on some variables. There was a significant PIFS advantage for all practical variables (marginally significant for reading). The control group began higher than the PIFS group (significantly for all variables except homework), and at posttest the PIFS group generally exceeded the control group (significantly for testing and homework). For the practical-4 summary score, the covariate-adjusted posttest improvement was .38 points greater for the PIFS group compared to the control group, which is a 17% larger gain (over the sample's pretest mean) for the PIFS group (note that the control group declined).

In Year 2, the PIFS and control groups showed significant pretest-posttest improvements. The PIFS group began significantly higher than the controls and ended significantly higher. Significantly greater improvement occurred for PIFS in reading, homework, testing, and the practical-summary scores. Essentially, as with academic scores, the gap between PIFS and control groups grew wider from pretest to posttest. For the practical-4 summary score, the covariate-adjusted posttest improvement was .33 points greater for the PIFS group compared to the control group, which is a 16% larger gain (over the sample's overall pretest mean) for the PIFS group.

Teacher (or class) differences. In Year 1, except for the reading-academic variable for PIFS teachers (classes), there were no significant differences between PIFS teachers (classes) on the pretest-posttest effects. All PIFS teachers (classes) showed positive pretest-posttest gains. There were no significant differences among control teachers. In Year 2, there was significant variability among PIFS teachers (classes) on the academic and practical variables, except for homework-practical scores. There was also significant variability among control teachers (classes) on the academic variables, reading-practical, and practical-2 summary scores.[12] However, all PIFS teachers showed positive pretest-posttest gains.

---

[12] As was the case with the Connecticut data, and for the same reasons, we do not present detailed analyses of teacher (class) differences here. These data are available from the first author.

Overall Summary of Studies 1 and 2

In general, the PIFS program successfully enhanced practical and academic skills. Positive results were observed in both years of the program in Studies 1 and 2 at the Connecticut and Massachusetts sites. The first-year results varied depending on the curriculum component, with PIFS advantages occurring for practical variables in particular. Consistently significant benefits of the PIFS curriculum were found for practical and academic variables in Year 2 at both the Connecticut and Massachusetts sites. The PIFS effect occurred across a variety of initial conditions in which the PIFS group at pretest scored lower than, equal to, or higher than the control group. In general, there was no interaction between PIFS treatment and gender (with a few exceptions for the Massachusetts site's Year-1 data), and there was consistency of the pretest-posttest gains across the PIFS teachers (classes).

Results and Discussion of Qualitative Data

The pretests and posttests provided a way to quantify the enhancement of students' practical intelligence over the course of the year. However, these numbers provided us with only part of the picture. We also wanted to track gradual changes in students' actual behavior, in their attitudes toward school and toward themselves, and in classroom dynamics. We wanted to gauge teachers' changing sense of the project and how they used it to enhance their normal curricula. In addition, we wanted to understand the conditions that best promote a PIFS implementation and, by extension, any attempt to train practical thinking skills.

To harvest this kind of information, we observed classrooms regularly, writing detailed ethnographic descriptions of what we saw. In informal conversations, we reflected frequently with teachers on the PIFS program and its effects on students. At the end of the school year, we also conducted formal interviews with all the teachers and with students selected at random from PIFS classrooms. We asked both teachers and students to reflect on the quality of their PIFS experiences. Students also were asked a series of general questions about PIFS skills and each of the focal areas. The student interviews each lasted about 30 minutes, while the teacher interviews were more free-ranging, lasting up to 2 hours. From all of these data, we gleaned insights about the value of PIFS and the conditions that foster the most effective implementations.

Effects of the Curriculum

1. *Acknowledging students' outside-of-school skills enhances student self-esteem and respect for peers.* Lessons in the introductory booklet ask students to reflect on things they do well outside of school and to demonstrate for the class a special talent or skill. In virtually every class, these lessons yielded an array of interesting and adept demonstrations from many students and many surprisingly energetic and talented performances from students who usually did not partici-

pate in class activities.

These lessons allowed teachers to see previously unacknowledged strengths in students and enabled students to build more distinctive and accurate pictures of themselves and of one another. The respect and esteem engendered by such demonstrations was evident both in the interviews and in classroom interactions. Teachers and students referred to these lessons as being the most memorable and among the most important of the curriculum. Several teachers commented with surprise on the skillfulness of their students' demonstrations, saying that the performances helped them to appreciate better the students who were less academically inclined. In their interviews at the end of the year, students recalled with great accuracy the demonstrations given by their classmates at the beginning of the year. Commented one poor academic performer: "It's wrong to think we're all smart in the same way. I can fix things and that makes me smart in my own way. And it's an important way." Another student remarked: "I always thought John was just kind of dumb. I didn't know he could put together tape recorders like that." One teacher pointed out, "Those lessons were crucial to the development of students' self-esteem. They legitimized what students *can* do well, rather than focusing on things they need to learn to do better."

A good example of the potential value of acknowledging students' outside skills was provided by one boy, who entered class late in the term and who immediately showed evidence of physical and emotional neglect. Academically, the boy was performing poorly, and he suffered from outbursts of temper when he became frustrated. During a PIFS lesson, the boy impressed the teacher and the class by disassembling, fixing, and reassembling an electronic appliance, while explaining what he was doing (at the teacher's urging). This experience of success encouraged the boy to work harder in other subjects as well, and the teacher was gratified to be able to praise the child for his legitimate (although usually untapped) abilities. Another child, a sixth-grade girl, came out of her shell when she brought to class a miniature and elaborate dollhouse she had built. The recognition and encouragement she received for her first-rate performance in building the dollhouse encouraged her to invest more energy in other more academic curricular areas in the future.

2. *Harnessing extraschool strengths and interests in the service of school work can serve as a powerful motivator, particularly for students not inclined to academics*. While the simple acknowledgment of their strengths and talents did much to improve students' self-esteem, those strengths also served as a hook to engage students in school work. Several teachers allowed students more freedom to be guided by their own interests in carrying out school work: An avid fan of airplanes did a book report on the Wright brothers; an artistically inclined student used drawings to take notes, to organize her thoughts before writing, and to illustrate her essays. In such instances, students reported greater satisfaction in doing school work, and teachers usually were pleased with the results from typically lackluster performers. In situations in which students normally

would have "tuned out" and simply not bothered to do homework, creating an opportunity to write about what interested them was the hook that kept them involved.

One child with a lust for video games and computers was encouraged by the teacher to harness his fascination with computers in service of his schoolwork. The boy researched the background leading up to the computer revolution for history class, wrote a short story about a boy who became addicted to computer games for writing class, and designed his own mock computer game for science class. The value of the PIFS approach here was in helping the boy benefit from doing what he enjoyed by borrowing from his topic area of interest while completing other assignments in seemingly unrelated areas.

When it works, this activation potential is probably the most powerful aspect of the PIFS curriculum: its capacity to help otherwise-disengaged learners to involve themselves in the standard curriculum in a way that is meaningful to them. However, this is not an easy tactic to negotiate, either for students or teachers, and it happens less frequently than we would like. It requires the teacher to allow students a certain amount of freedom in carrying out their assignments. It requires that students have enough self-confidence and self-knowledge to use this freedom to advantage. While the results of this stretch from school to personal lives are impressive, getting such results with any regularity probably involves more than a year's worth of practice for teachers and students too used to fulfilling the letter, and no more than the letter, of the school law.

3. *Basing skill-building on students' own experiences and inclinations gives all students the same opportunity to participate in class and to develop those skills*. Consistent with the above-mentioned emphasis on helping students assume responsibility for developing their own best ways of doing things, many PIFS lessons ask students to talk about their particular strengths, problems, and past experiences in carrying out various kinds of assignments (e.g., studying for tests and reading difficult passages). Since all students — not just the strong academic performers — have such experiences to share, PIFS lessons often afforded low-achievers the chance to speak and be heard. Rather than the best academic students dominating class time, PIFS helped put other students on a more equal footing with them.

In every class, teachers and researchers observed usually passive students — students who never offered much in standard class discussions — taking active roles in PIFS discussions, sharing their experiences with a particular problem or solution, and suggesting strategies to other students. The effect of this change on classroom dynamics was tangible. PIFS seemed to perform a similar function in reducing gender differences in typically male-dominated classes such as mathematics. Despite the fact that girls are not usually as assertive as boys in math classes, girls at several of the sites showed an increase in their participation in math over the course of the year.

Compared to the abilities emphasized by traditional academic curricula, PIFS focuses on a more robust group of abilities as being relevant for school and life performance. Thus, with PIFS, students who are not usually as involved become more involved — they have something to contribute, rather than sitting on the sidelines and letting the "brains" do all the work. The "brains" also learn that others have worthy ideas. Some of these "brains" also realize that their ability in academics is not mirrored in other practical skills areas. This is a positive aspect of the curriculum for all students, since everyone must learn to respect others with varying talents and since both the academically and the practically competent children often have significant, although different, skills.

4. *PIFS skills become lasting and useful tools for learning only when students can see the tangible results of their efforts to employ them.* PIFS runs the risk of becoming like other aspects of the school curriculum that students memorize dutifully, regurgitate faithfully on a test, and promptly forget. PIFS skills become useful and lasting tools only for those students who see concrete benefits from using these skills. Students continue to draw on memory techniques when they see their scores on vocabulary tests improve. They continue to take notes long after the note-taking lessons are over when they learn that it makes report writing easier.

But such changes and improvements in students' work come only gradually over the course of the school year. This fact emphasizes the importance of the many opportunities PIFS gives students to reflect on their work and their progress. Through journal writing, reflection discussions, peer feedback, and the like, the PIFS curriculum offers students many ways to track their small improvements, allowing them to see progress before it yields the impressive differences that many students ultimately describe.

5. *PIFS skills seem to endure.* We have evidence suggesting that PIFS skills endure and that the curriculum creates lasting impressions on students. Teachers we interviewed whose students were taught PIFS during the previous year noted that these students remembered and applied aspects of their PIFS experiences to their new classroom setting. These students suggested research and project topics that called on outside interests. They exhibited well-developed and effective note-taking, studying, and test-preparation strategies. They saw the big picture when they wrote — understanding the needs to plan, create, and revise. They asked teachers to be specific about course and instructional objectives and requirements. In general, PIFS students a year later seemed to have been assisted in their transfer to middle school by their PIFS skills. (Of course, these are impressionistic data, and we lack a direct quantitative comparison of the tendency of PIFS graduates to display these behaviors as compared to control-group graduates.)

Lessons about Implementation

In the process of working with teachers, we learned about the circumstances

that fostered or hindered the implementation of the PIFS curriculum.

1. *Flexibility in implementing the curriculum is essential.* Whereas all teachers completed the introductory lessons first, they then followed the dictates of their own needs in deciding how and when to carry out lessons in the other focal areas. This flexibility in the implementation of the curriculum had several advantages: First, it invited teachers to take an active role in planning and carrying out the PIFS curriculum. Second, since the PIFS lessons were being used to meet real and pressing needs, such flexibility helped to secure sufficient time for PIFS in the always-busy school day. Finally, the fact that teachers could choose lessons to complement their standard curriculum ensured that students would use PIFS in the context of their standard school work — a necessary condition for the students to internalize the principles and skills advocated in the PIFS curriculum.

2. *PIFS is a tool for teacher refection and change.* The process of adapting PIFS to the standard curricula proved to be an important, difficult, and thought-provoking venture for teachers. Many commented that PIFS helped them systematically to address issues they always had believed were important but did not always have time to teach. Several found that having to be explicit about their expectations for students was a useful exercise — they realized they had some expectations they never fully articulated to students. Others found that identifying the central lessons in the PIFS curriculum was a good way to set their priorities. Useful though it is, this process of reflection and revision of practice is as slow and difficult for teachers as it is for students. For example, one teacher learned that although her students did, in fact, know the material quite well, the format of several test questions had puzzled them — leading to generally poor performance on the test. For instance, many could not read a graph that she had included with one question. Others were stumped by unfamiliar phrases or words.

This teacher came to appreciate that it was the lack of PIFS skills — knowing how to take a test — and not lack of content knowledge that was keeping her students from performing well. Before the next test, she spent more time talking with her class about the types of questions and format of the test. The students did remarkably better. It was a rewarding experience, but a time-consuming one. Real progress came only after this teacher had taken considerable personal and class time to explore the problem, reassess and adjust her own practice, and then guide the students in readjusting their approaches to studying and test taking.

PIFS offered teachers the chance to reflect on the assumptions and values implicit in their practice. Two teachers commented that after many years of teaching in an environment that defines intelligence primarily through test scores and academic achievement (strictly defined), they had tacitly adopted this belief themselves. Because of this emphasis on strictly "academic" intelligence, the teachers believed that they sometimes failed to recognize children's abilities

when these abilities were in the practical or creative domains. Thus, PIFS exercises and lessons designed to help students and classmates focus on abilities in other domains had a fortunate side effect: These lessons also encouraged teachers to broaden their focus and their assessment of students to include a broader definition of intelligence. All students benefit when teachers expand their views, since teachers are then better able to prepare students for the world outside of school, where many types of intelligence are important.

These teachers also pointed out that PIFS had provided them with valuable lessons for personal development. Despite being veteran teachers, they reported that they often forgot just how different from their own were their colleagues' expectations of students. These teachers noted that the PIFS emphasis on articulating teacher expectations and explicit criteria for good student performance forced them to examine critically aspects of their own views on teaching they generally took for granted. The PIFS group meetings, where teachers discussed their experiences with the curriculum, provided a forum for teachers to discuss their implicit expectations and views, which were now made explicit. Realizing how different they were from one another was instructional and beneficial in that it alerted teachers to the types of problems students face as they move from teacher to teacher, both as they advance in school and as they move within a grade from class to class. How confusing it must be to an 11-year-old, one PIFS teacher remarked, to recalibrate her understanding, expectations, and internal rules to fit different teacher expectations.

Because of this experience, these teachers vowed to examine their implicit assumptions more often and revise them if warranted. They also committed to making these assumptions explicit for students by discussing them early in the term and repeatedly at appropriate points. During sessions devoted to articulating their expectations of students, these teachers solicited students' input about these expectations: Did anything seem unfair to the students? How could the teacher improve his or her outlook? The value of this evolving process was in promoting teacher reflection and improvement, as well as in improving student understanding of the expectations by which they would be judged. Students were also given input into the system, which consequently came to be perceived by them as more nearly fair. With so many aspects of school life usually left tacit now out in the open, everyone could be assessed fairly. PIFS was the catalyst for this improvement.

3. *The stand-alone and infused lessons play complementary roles.* The stand-alone (mostly introductory) lessons were welcomed by teachers far more readily than we expected them to be, at least at first. Although these lessons took time away from the standard curriculum, they seemed to offer teachers a simple and straightforward way to introduce new skills and ideas to their students. However, as the year progressed, teachers seemed more interested in those lessons that would fit well with their students' traditional school work.

This fact did not make infusion easy. Although teachers could plan to incor-

porate PIFS skills into particular lessons or projects, many did not do so spontaneously as opportunities arose in class to reinforce skills that already had been taught. This is partly because all the PIFS teachers were experienced and quite effective in their classroom practices: They already had a repertoire of responses to particular situations and questions. It is also a function of the length of the PIFS curriculum. There are so many skills and ideas covered that the ones introduced at the beginning of the year may lie dormant for several weeks while new ones are being introduced and reinforced.

4. *PIFS taught in a middle-school cluster situation has both advantages and disadvantages when compared to the elementary-school self-contained classrooms.* With more and more sixth grades being included in middle school rather than elementary school settings, it is useful to consider the implementation issues that arise in the team-taught cluster situation (the most typical middle-school organization). In this setting, four teachers (one for each of the core participants) teach four classes of students who rotate among the teachers while keeping the same classmates. From a PIFS perspective, such a situation offered several advantages.

First, since the team of teachers shared common planning time, there was ample opportunity for them to talk with each other about the progress and implementation of the PIFS program. This collaboration proved especially helpful when particular lessons seemed ineffective or when students did not respond to an apparently engaging topic. The teachers supported and encouraged one another in a way that researchers who are not "in the trenches" often cannot. And since they all knew and taught the same students, they were able to make suggestions to each other about how to handle particular children or situations. Second, because the students move to different classes for different subjects, the issue of identifying differences among the various subjects becomes an easier one to address.

However, the organizational difficulties of a team-taught cluster of classes prevented us from taking more advantage of this opportunity. The PIFS curriculum was built around four focal areas — reading, writing, homework, and test-taking — skills that are used in all core subject areas. Deciding how to divide the PIFS lessons effectively among four different teachers and then orchestrating follow-up activities in the other three classes proved difficult at best. The teachers of the self-contained classrooms were able to do follow-up lessons more effectively and consistently.

In conclusion, the combined quantitative and qualitative results of this set of studies provide empirical evidence supporting the importance of practical intelligence for school in students' school success.

General Discussion

The Practical Intelligence for School project showed that practical skills

essential to school success can be defined and taught. This is welcome news for students who fall short of their potential because they lack basic practical insights into their teachers' expectations and how to fulfill these expectations. For teachers, the possibility of training practical intelligence for school may mean less frustration with students who do not perform satisfactorily due to an array of factors not related to lack of analytical ability. The PIFS program provides one method for and approach to training these essential practical skills: Students exposed to PIFS become better able to make optimal use of their gifts and abilities within the context of the school environment, while learning practical skills they can use throughout their lives.

Over the past several decades there have been various trends favoring different types of curricular approaches and interventions. Sometimes, curricula are developed and implemented on the basis of "anecdata," or anecdotal reports, instead of carefully controlled studies. For example, a recent trend in educational intervention has focused on building emotional and moral intelligence (Coles, 1996). Lately it is often said that we must educate our youth for character and moral values (e.g., the "Character Counts" curriculum in use across the United States). Certainly we are not suggesting that there is anything wrong with wanting children to develop character and morality! However, the questions of exactly how to accomplish this goal (or any other educational goal), and of how to know if we are succeeding with any individual program, can only be answered in a scientifically adequate way through empirical research.

Unlike the latest fad that sometimes becomes a curricular intervention in the absence of a solid theoretical foundation and rigorous supporting data, the PIFS program is rooted in theory and based on hard empirical evidence. Our data show something meaningful and promising by providing evidence that it is possible to improve broad-based intellectual skills. By focusing on reading, writing, homework, and test-taking ability, we cast our net wide in an attempt to create meaningful changes in broad areas of students' intellectual performances.

As already discussed, there has been widespread disagreement regarding the degree to which children's intellectual capabilities can be modified. For example, Herrnstein and Murray (1994) essentially dismissed intervention effects, arguing that short of adoption, there are no meaningful ways to raise the intellectual performance of children. Although future research is needed to assess the long-term durability of training in practical intelligence, our data show reasonable increases over the school year. Thus, on the topic of the controversy regarding potential intervention effects, we weigh in on the side of cautious optimism. In the very least, our data suggest that further research on increasing practical intelligence is warranted.

A broader issue raised by this research concerns the definition of intelligence itself and how one's definition affects one's viewpoint regarding the modifiability of intelligence and how best to enhance it. On the theoretical side a growing literature suggests that traditional *g*-based psychometric conceptions of intelli-

gence are incomplete. Interest in the type of intelligence people use to solve real-world problems, referred to here as practical intelligence, has led to a broad cross-section of studies identifying practical intelligence in different domains (Ceci, 1996; Rogoff & Lave, 1984; Scribner & Cole, 1981; Sternberg & Wagner, 1986, 1994; Sternberg, Wagner, & Okagaki, 1993; Voss, Perkins, & Segal, 1991) and even studies showing that practical intelligence can be assessed and taught (see Sternberg, Forsythe, Hedlund, Horvath, Wagner, Williams, Snook, & Grigorenko, 2000; and Sternberg, Wagner, Williams, & Horvath, 1995, for reviews).

Thus, theoretically speaking, a psychometrically based measure of intelligence such as *g* may not be the whole story when it comes to understanding children's intelligence, even though it accounts for far more of the variance in grades and scores than do any of the specific factors subsumed under *g*. Practically speaking, psychometrically based *g* may not be sufficient if we wish to be fair and accurate in our assessment of children's capabilities in the classroom and beyond. Just as Renzulli (1986) has argued for a broad-based approach to identifying giftedness — in which children are identified based not only on above-average ability, but also on high levels of motivation and creativity — we argue for a broad-based approach to identifying school-based competence. We believe that practical intelligence should be seen as an essential component of children's competence, worthy of assessment and instruction in its own right. Our research suggests that training in practical intelligence can help children remediate areas of weakness, as well as build on existing skills, to improve their performance in many academic areas.

The evidence regarding the modifiability of *g*-based intelligence is mixed. But if we can accept that intelligence is more than *g* (and the attendant specific factors revealed by factor analysis of test scores, such as visualization ability, memory, and perceptual speed), there is hope that meaningful increases in intelligence can be achieved, even if (for example) these increases do not focus on *g*-based abilities. Thus, putting aside one's point of view regarding whether we can affect measures of *g* through training, our research shows that we can affect measures of practical intelligence through training. Whether one wishes to call practical intelligence a type of intelligence, a type of knowledge, a set of skills, or whatever, the point is that it can be delineated and taught successfully, and it has real consequences for school success.

Theoretical arguments aside, then, we would advocate further research on training of practical intelligence, regardless of how the concept fits into one's particular view of intelligence, because of the increases in school performance that result from such training. The training of practical intelligence may be particularly useful in challenged populations because students in these populations may have had little opportunity to acquire school-relevant practical intelligence on their own or at home. Training challenged students may help them to over-

come a deficit in school-related knowledge and skills that could otherwise have derailed them. Many of these students may have latent capabilities that they have not harnessed or profited from because of a lack of fit between the student and the school environment. By helping students understand what is expected of them and why, and by demystifying the process of succeeding in school, training in practical intelligence may help to reach students who have previously opted out of the school experience.

In conclusion, we have shown that practical intelligence can be identified, assessed, and taught in order to achieve meaningful increases in real-world success in the classroom. We advocate further research on practical intelligence for school to broaden our appreciation and understanding of this construct and its applications in the classroom.

References

Beale, C. R. (1996). The role of comprehension monitoring in children's revision. *Educational Psychology Review*, **8**, 219-238.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool.* Englewood Cliffs, NJ: Prentice Hall.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. Reese (Ed.), *Advances in child development and behavior*, Vol. 10. New York: Academic Press.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development* (expanded edition). Cambridge, MA: Harvard University Press.

Coles, R. (1996). *The moral intelligence of children.* Cambridge, MA: Harvard University Press.

Consortium for Longitudinal Studies (1983). *As the twig is bent: Lasting effects of preschool programs.* Hillsdale, NJ: Erlbaum.

Detterman, D. K., & Sternberg, R. J. (Eds.). (1982). *How and how much can intelligence be increased?* Norwood, NJ: Ablex.

Feuerstein, R. (1980). *Instrumental enrichment.* Baltimore, MD: University Park Press.

Finn, J. P., & Achilles, C. H. (1990). Answers and questions about class size: A statewide experiment. *American Education Research Journal*, **27**, 557-577.

Finn, J. P., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, and misconceptions. *Educational Evaluation and Policy Analysis*, **21**, 97-110.

Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk.* Washington, DC: American Association on Mental Retardation.

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

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H. (1999). Are there additional intelligences? In J. Kane (Ed.), *Education, information, and transformation.* Englewood Cliffs, NJ: Prentice Hall.

Gardner, H., Krechevsky, M., Sternberg, R. J., & Okagaki, L. (1994). Intelligence in context: Enhancing students' practical intelligence for school. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 105-127). Cambridge, MA: MIT Press.

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

Herrnstein, R. J., Nickerson, R. S., DeSanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist*, **41**, 1279-1289.

Honig, A. S. (1994). Intervention, infant and preschool. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 599-607). New York: Macmillan.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, **39** (1), 1-123.

Jensen, A. R. (1989). Raising IQ without increasing g?: A review of "The Milwaukee Project: Preventing mental retardation in children at risk." *Developmental Review*, **9**, 234-258.

Lazar, I., & Darlington, R. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development*, **47**(2-3), 1-141.

McLaughlin, D. H. (1977). *Title 1, 1965-1975: A synthesis of the findings of federal studies*. Palo Alto, CA: American Institutes for Research.

Metcalfe, J., & Shimamura, A. (1996). *Metacognition*. Cambridge: MIT Press.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, **5** (2), 113-127.

Neisser, U. (1979). The concept of intelligence. *Intelligence*, **3**, 217-227.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, W. A., 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.

Nickerson, R. S. (1994). Interventions, later. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 607-613). New York: Macmillan.

Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.

Nisbett, R. F., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, **238**, 625-631.

O'Sullivan, J. T., & Pressley, M. (1984). Completeness of instruction and strategy transfer. *Journal of Experimental Child Psychology*, **55**, 396-414.

Piaget, J. (1959). *The language and thought of the child* (3rd rev. ed.). London: Routledge & Kegan Paul (original French edition 1923).

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, **25**, 19-34.

Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, **23**, 270-288.

Ramey, C. T. (1994). Abecedarian project. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. I, pp. 1-3). New York: Macmillan.

Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, **88**, 515-523.

Ramey, C. T., & Campbell, F. A. (1992). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A. Huston (Ed.), *Children in poverty* (pp. 190-221). New York: Cambridge University Press.

Ramey, C. T., Campbell, F. A., Bryant, D. M., Burchinal, M., Sparling, J., & Wasik, B. H. (1992). Early intervention and long-term predictors of school status. In F. Parker et al. (Eds.), *New directions in child and family research* (pp. 170-174). Washington, DC: Administration on Children, Youth, and Families.

Ramey, C. T., & Ramey, S. L. (1998). Prevention of intellectual disabilities: Early interventions to improve cognitive development. *Preventative Medicine*, **27** (1998), 224-232.

Ree, M. J., & Earles, J. A. (1993). *g* is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Science*, **1**, 11-12.

Renzulli, J. S. (1986). The three ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 53-92). New York: Cambridge University Press.

Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: Its development in social context*. Cambridge, MA: Harvard University Press.

Scarr, S. (1997). Behavior-genetic and socialization theories of intelligence: Truth and reconciliation. In R. J. Sternberg & E. Grigorenko (Eds.), *Intelligence, heredity, and environment* (pp. 3-41). Cambridge, UK: Cambridge University Press.

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.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool Study through age 27* (Monograph Number 10). Ypsilanti, MI: High/Scope Press.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University. Press.

Snow, R. E., & Yalow, R. (1982). Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493-585). New York: Cambridge University Press.

Spitz , H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence*. Hillsdale, NJ: Erlbaum.

Spitz, H. H. (1992). Does the Carolina Abecedarian early intervention project prevent sociocultural mental retardation? *Intelligence*, **16**, 225-237.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1996). *Successful intelligence*. New York: Simon & Schuster.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology*, **3**, 292-316.
Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, **41**, 37-55.
Sternberg, R. J., Ferrari, M., Clinkenbeard, P. C., & 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. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge Univ. Press.
Sternberg, R. J., Okagaki, L., & Jackson, A. (1990). Practical intelligence for success in school. *Educational Leadership*, **48**, 35-39.
Sternberg, R. J., & Spear-Swerling, L. (1996). *Teaching for thinking*. Washington, DC: American Psychological Association.
Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge Univ. Press.
Sternberg, R. J., & Wagner, R. K. (Eds.). (1994). *Mind in context*. New York: Cambridge Univ. Press.
Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacic knowledge in work and at school. In H. Reese & J. Puckett (Eds.), *Advances in lifespan development* (pp. 205-227). Hillsdale, NJ: Erlbaum.
Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist*, **50** (11), 912-927.
Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.), (1991). *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.
Williams, W. M., Blythe, T., White, N., Li, J., Sternberg, R. J., & Gardner, H. (1996). *Practical intelligence for school*. New York: HarperCollins.
Zigler, E., & Berman, W. (1983). Discerning the future of early childhood intervention. *American Psychologist*, **38**, 894-906.

Emotional Intelligence Meets Traditional Standards for an Intelligence

John D. Mayer, University of New Hampshire; David R. Caruso, Work Life Strategies; and Peter Salovey, Yale University

An intelligence must meet several standard criteria before it can be considered scientifically legitimate. First, it should be capable of being operationalized as a set of abilities. Second, it should meet certain correlational criteria: the abilities defined by the intelligence should form a related set (i.e., be intercorrelated), and be related to pre-existing intelligences, while also showing some unique variance. Third, the abilities of the intelligence should develop with age and experience. In two studies, adults (N = 503) and adolescents (N = 229) took a new, 12-subscale ability test of emotional intelligence: the Multifactor Emotional Intelligence Scale (MEIS). The present studies show that emotional intelligence, as measured by the MEIS, meets the above three classical criteria of a standard intelligence.

Emotions are internal events that coordinate many psychological subsystems including physiological responses, cognitions, and conscious awareness. Emotions typically arise in response to a person's changing relationships. When a person's relationship to a memory, to his family, or to all of humanity changes, that person's emotions will change as well. For example, a person who recalls a happy childhood memory may find that the world appears brighter and more joyous (e.g., Bower, 1981). Because emotions track relationships in this sense, they convey meaning about relationships (Schwarz & Clore, 1983). Emotional intelligence refers to an ability to recognize the meanings of emotions and their relationships, and to reason and problem-solve on the basis of them. Emotional intelligence is involved in the capacity to perceive emotions, assimilate emotion-related feelings, understand the information of those emotions, and manage them (Mayer & Salovey,1997; Salovey & Mayer, 1990).

Emotional intelligence can be assessed most directly by asking a person to solve emotional problems, such as identifying the emotion in a story or painting, and then evaluating the person's answer against criteria of accuracy (Mayer, DiPaolo, & Salovey, 1990; Mayer & Geher, 1996). It is worth noting, however, that emotional intelligence, as an ability, is often measured in other ways. Some approaches have asked people their personal, self-reported beliefs about their emotional intelligence. Test items such as, "I'm in touch with my emotions," or "I am a sensitive person," assess such self-understanding (e.g., Mayer & Stevens, 1994; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995). Self-reports of ability and actual ability, however, are only minimally correlated in the realm

Reprinted from Intelligence 27 (4), 2000

of intelligence research (e.g., $r = 0.20$; Paulhus, Lysy, & Yik, 1998) and that appears to hold in the area of emotional intelligence as well (Davies, Stankov, & Roberts, 1998).[1] Self-concept is important, of course, because people often act on their beliefs about their abilities as opposed to their actual abilities (Bandura, 1977). Emotional intelligence as a domain of human performance, however, is best studied with ability measures.

Emotional intelligence has often been conceptualized (particularly in popular literature) as involving much more than ability at perceiving, assimilating, understanding, and managing emotions. These alternative conceptions include not only emotion and intelligence per se, but also motivation, non-ability dispositions and traits, and global personal and social functioning (e.g., Bar-On, 1997; Goleman, 1995). Such broadening seems to undercut the utility of the terms under consideration. We call these *mixed* conceptions because they combine together so many diverse ideas. For example, the Bar-On Emotional Quotient Inventory (EQ_i) includes 15 self-report scales that measure a person's self regard, independence, problem solving, reality-testing, and other attributes (Bar-On, 1997). Such qualities as problem solving and reality testing seem more closely related to ego strength or social competence than to emotional intelligence. Mixed models must be analyzed carefully so as to distinguish the concepts that are a part of emotional intelligence from the concepts that are mixed in, or confounded, with it.

General intelligence serves as an umbrella concept that includes dozens of related groups of mental abilities. Most of the smaller subskills studied in this century are related to verbal, spatial, and related logical information processing (see Carroll, 1993, for an authoritative review). Such processing is sometimes referred to as "cold" to denote that its ego- or self-involvement is minimal (Abelson, 1963; Mayer & Mitchell, 1998; Zajonc, 1980). Information processing, however, also deals with "hot," self-related, emotional processing. Emotional intelligence is a hot intelligence. It can be thought of as one member of an emerging group of potential hot intelligences that include social intelligence (Sternberg & Smith, 1985; Thorndike, 1920), practical intelligence (Sternberg & Caruso, 1985; Wagner & Sternberg, 1985); personal intelligence (Gardner, 1993), non-verbal perception skills (Buck, 1984; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), and emotional creativity (Averill & Nunley, 1992). Each of these forgoing concepts forms coherent domains that partly overlap with emotional intelligence, but that divide human abilities in somewhat different ways.

---

[1] The Davies et al. (1998) article critiques early scales in the emotional intelligence literature. The present article was essentially completed before the Davies et al. work was published, and so we do not comment specifically on those authors' criticisms of emotional intelligence scales developed before this one. Nonetheless, it is our hope that the test results here will put to rest certain of the Davies et al. concerns, such as those related to the reliability of emotional intelligence tests.

The ability conception of emotional intelligence was developed in a series of articles in the early 1990s (Mayer et al., 1990; Mayer & Salovey, 1993; Salovey & Mayer, 1990). For example, the first empirical study in the area demonstrated that people's abilities to identify emotion in three types of stimuli — colors, faces, and designs — could be accounted for by a single ability factor which we supposed was emotional intelligence (Mayer et al., 1990). Another study examined the understanding of emotion in stories (Mayer & Geher, 1996); this latter study provided further indications that the underlying factor "looked like" an intelligence. Simultaneous with this empirical work, we have honed our definition of emotional intelligence and the abilities involved (e.g., Mayer & Salovey, 1997). The present article represents a culmination of this work, testing our most highly developed conception of emotional intelligence by operationalizing it according to 12 ability tests of emotional intelligence. The present study can help answer important questions about emotional intelligence, among them whether emotional intelligence is a single ability or many, and how it relates to traditional measures of general intelligence and other criteria.

Standard Criteria for an Intelligence

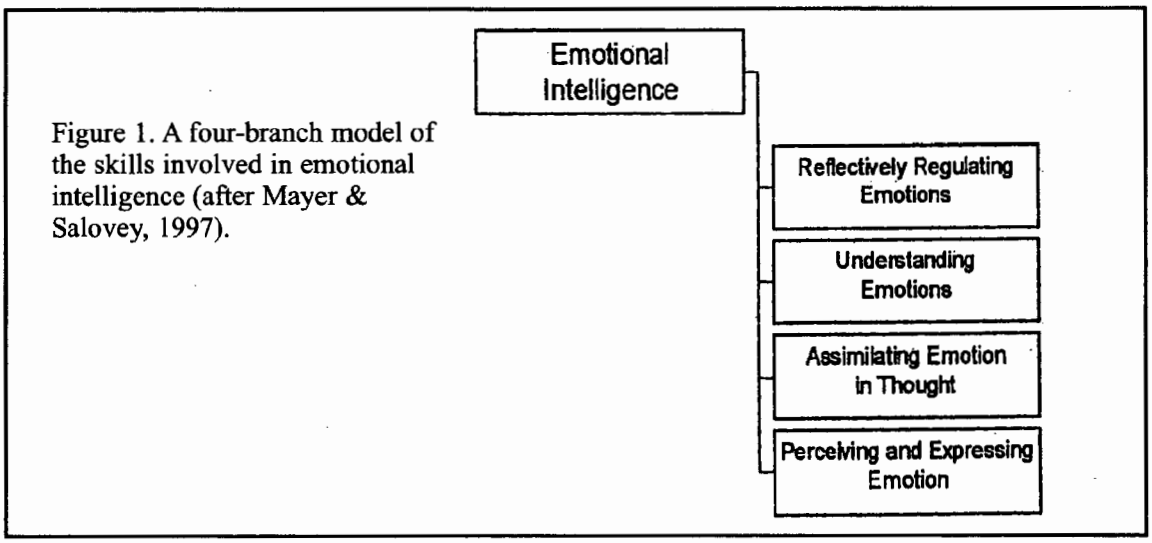


Figure 1. A four-branch model of the skills involved in emotional intelligence (after Mayer & Salovey, 1997).

Three Criteria for an Intelligence

An intelligence such as emotional intelligence must meet stringent criteria in order to be judged as a true intelligence. For the purposes here, these criteria can be divided into three fairly distinct groups: conceptual, correlational, and developmental. The first, conceptual criteria, includes that intelligence must reflect mental performance rather than simply preferred ways of behaving, or a person's self-esteem, or non-intellectual attainments (Carroll, 1993; Mayer & Salovey, 1993; Scarr, 1989); moreover, mental performance should plainly measure the concept in question, i.e., emotion-related abilities. The second, correlational criteria, describe empirical standards: specifically, that an intelligence

should describe a set of closely related abilities that are similar to, but distinct from, mental abilities described by already-established intelligences (Carroll, 1993; Neisser et al., 1996).[2] The third, developmental criterion, states that intelligence develops with age and experience, and is based on the groundbreaking work by Binet and Simon at the beginning of century (as reviewed in Fancher, 1985, p. 71; see also, Brown, 1997). These three criteria will be next examined in greater detail.

Conceptual Criteria for an Intelligence

We have argued elsewhere that emotional intelligence does indeed describe actual abilities rather than preferred courses of behavior. These four broad classes of abilities can be arranged from lower, more molecular, skills to higher, more molar, skills, as is done in Fig. 1 (Mayer & Salovey, 1993, 1997). The lowest level skills involve the perception and appraisal of emotion, e.g., in a facial expression or artwork. The next level up involves assimilating basic emotional experiences into mental life, including weighing emotions against one another and against other sensations and thoughts, and allowing emotions to direct attention. An example includes holding an emotional state in consciousness long enough to compare its correspondences to similar sensations in sound, color, and taste. The third level involves understanding and reasoning about emotions. Each emotion — happiness, anger, fear, and the like — follows its own specific rules. Anger rises when justice is denied; fear often changes to relief; sadness separates us from others. Each emotion moves according to its own characteristic rules, like the different pieces on a chessboard. Emotional intelligence involves the ability to see the pieces, know how they move, and reason about emotions accordingly. The fourth, highest level, involves the management and regulation of emotion, such as knowing how to calm down after feeling angry or being able to alleviate the anxiety of another person. Tasks defining these four levels or branches are described in greater detail in the section concerning scale development below.

In considering tasks for an emotional intelligence test, how are we to discriminate right from wrong answers? One common approach drawn from emotions research has been to look for group *consensus* as to the emotional content of stimuli (e.g., Mayer et al., 1990; Wagner, MacDonald, & Manstead, 1986). If the group agrees that a face is happy, say, then that becomes the correct answer. A second possibility is to use *expert* criteria for emotional meanings. An expert

---

[2] One very different set of criteria, suggested by Howard Gardner, includes requirements that an intelligence be identified with a specific brain region or structure and be a culturally valued mental characteristic. Intelligences that are valid according to Gardner's criteria alone are definitely worth studying and may provide information for a next generation of intelligence tests. Still, intelligences that fit his criteria but that are indistinguishable from general intelligence at a behavioral level plainly cannot assist in predicting criteria such as academic success. For that reason, correlational approaches remain of the greatest pragmatic concerns for now.

could bring a history of philosophy and empirical psychology to bear on judgments about emotional meanings (e.g., Darwin, 1872/1965; Ortony, Clore, & Collins, 1988; Plutchik, 1984; Spinoza, 1675/1984), and this might provide answers similar to, or different from, a consensus criterion. On the other hand, it has been argued that experts simply provide estimates of group consensus, and those estimates are fallible (Legree, 1995). Finally, a *target* criterion is applicable in selected circumstances in which a target individual's emotions or emotional creations are being judged. In such cases, the target can report the emotion he or she was feeling or expressing at the time. The group's consensus, the expert, and the target criteria represent somewhat different perspectives, and it is therefore unlikely that they would be in complete agreement. For example, target individuals sometimes report pleasant feelings, perhaps to be socially conforming, when in fact they are perceived by a group as experiencing less pleasant feelings (Mayer & Geher, 1996). Such differences in perspective do not necessarily rule out a general convergence toward a criterion. Such a rough convergence would substantiate the view that emotions convey information, and that emotional intelligence is, in fact, an intelligence.

Correlational Criteria for an Intelligence

The Logic of Correlational Criteria for Intelligence

Emotional intelligence should define a set of abilities that are moderately intercorrelated with one another. There are many excellent overviews of mental abilities and the criteria for defining their class (e.g., Carroll, 1993; Flanagan, Genshaft, & Harrison, 1997). This logic can be illustrated with an example drawn from the clinical assessment of intelligence. The original Wechsler Adult Intelligence Scales (i.e., WAIS, WAIS-R, WAIS-III; Wechsler, 1958; see Anastasi & Urbina, 1997, for a review of later tests) contained a set of verbal intelligence scales. These consisted of many related mental tests including identifying similarities among concepts, recognizing word meanings (vocabulary), general information, comprehension, and arithmetic. The abilities measured, e.g., vocabulary and information, are moderately intercorrelated — they rise and fall across people at about the $r = 0.40$ level. The tasks can be summarized by a verbal IQ, where the IQ is based on a person's overall performance on those tasks compared to the performance of other people their age (because ability levels change with age).

The Wechsler tests from mid-century to 1998 typically paired verbal intelligence with performance intelligence. Performance abilities, such as assembling puzzles, identifying missing elements in visual depictions, and ordering picture sequences, also correlate highly with each other. These can be summarized by a performance IQ, similarly based on the person's overall performance on the tasks. The verbal and performance tasks correlate less highly with each other;

i.e., the verbal and performance tasks are related to each other, but not quite as closely as skills within each group.[3] They are also related, however, and can be combined to form an overall IQ, which represents the individual's average performance on a broader range of mental tasks.

The Establishment of New Intelligences

The possibility that there exists one or more additional classes of intelligence, beyond verbal and performance intelligence, has long intrigued researchers. The identification of a new class of intelligence would broaden our contemporary concepts of intelligences. Moreover, adding missing intelligences to an omnibus IQ test can increase the test's fairness by more accurately representing individuals whose abilities were higher on unknowingly omitted tests than on the tests that were present.

The identification of a class of intelligence, such as verbal or performance, however, does not occur all at once. Usually, there proceeds a painstaking process of developing candidate tasks for the intelligence, finding a rationale for correct answers (if not obvious), and then examining their intercorrelations with existing measures of intelligences. For example, social intelligence was proposed as a third member of the verbal/performance grouping earlier in the century; it was defined as "the ability to understand men and women, boys and girls, to act wisely in human relations" (Thorndike, 1920). Measures of verbal intelligence, however, already incorporate much social thinking; in fact, normal verbal communication is so social that it is difficult to come up with vocabulary ("What is democracy?") or general knowledge questions ("Who was John F. Kennedy?") that do not contain social information. In part, for such reasons, Cronbach (1960) concluded that social intelligence could not be distinguished from verbal intelligence. The search for a third broad intelligence abated for the next several decades, although a number of alternative intelligences have been discussed as possible candidates.

Research on social intelligence has continued, with important work by Sternberg and Smith (1985), Cantor and Kihlstrom (1987), Legree (1995) and others. Much of that work represented important conceptual development of social intelligence; little of that work, however, concerned itself with actual ability measurement in relation to other intelligences (some exceptions are Legree, 1995; Wagner & Sternberg, 1985). In addition, other intelligences have been proposed, e.g., the multiple intelligences of Gardner (1993) which included personal, musical, and other intelligences. Here, too, research on individual differ-

---

[3] If two intelligences are entirely unrelated, however, we may want to raise the question as to whether one of them is a real intelligence, because mental abilities are generally related to one another. In fact, the "First Law of Intelligence" of Guttman and Levy (1991) states that all mental ability measures are positively correlated.

ences and their relations to already-existing intelligences was de-emphasized (Sternberg, 1994).

Emotional intelligence represents an alternative grouping of tasks to social intelligence. On one hand, emotional intelligence is broader than social intelligence, including not only reasoning about the emotions in social relationships, but also reasoning about internal emotions that are important for personal (as opposed to social) growth. On the other hand, emotional intelligence is more focused than social intelligence in that it pertains primarily to the emotional (but not necessarily verbal) problems embedded in personal and social problems. For example, reasoning about a sequence of internal feelings, or about the feelings in a relationship, can be readily distinguished from general questions about democracy or John F. Kennedy, as described above. This increased focus means that emotional intelligence may be more distinct from traditional verbal intelligence than is social intelligence.

The Developmental Criterion for an Intelligence

There remains a third criterion an intelligence must meet: that it develops with age and experience, from childhood to adulthood. That third criterion will be discussed at the outset of Study 2, which is focused on studying developmental issues in emotional intelligence.

Introduction to the Present Studies

Widely accepted intelligences share certain features in common: they are abilities, they manifest specific correlational patterns among themselves and in relation to other intelligences, and they develop with age and experience. The two studies described here operationalize emotional intelligence as a set of abilities, study the intercorrelational pattern among those abilities, and examine evidence for their growth between adolescence and early adulthood.

In Study 1, we constructed a set of 12 ability measures drawn from each of the four defined areas of emotional intelligence including perceiving, assimilating, understanding, and managing emotion (Mayer, Salovey, & Caruso, 1997). The test was administered to a large group of adults. We predicted moderate correlations among the 12 tasks, and that a group factor, i.e., one that loads all 12 tasks, can be derived. As in our earlier work, we predicted that a combination of these tasks correlates with traditional forms of intelligence such as verbal intelligence at such a level as to be distinct from such traditional intelligences. Study 1 also examines evidence of whether this emotional intelligence predicts empathy, parental warmth, and cultural pursuits.

Study 2 focused more specifically on whether emotional intelligence meets the developmental criterion for an intelligence. An adolescent sample was given

a reduced set of the same group of tasks. The adolescent data are then compared to a subset of the adult data from Study 1 so as to test the hypothesis that adults outperform adolescents on the tasks.

Study I
Method

Participants

Participants were 503 adults (164 men and 333 women, six unreported) with a mean age of 23 years (range: 17-70), drawn from several sources. One group of individuals (47%, *N*= 235) was comprised of full-time college students who participated to fulfill an introductory psychology course research requirement, or who were paid (US$15) for their participation. The remainder (53%, 268) were part-time college students, corporate employees, career workshop attendees, and executives in an outplacement setting who volunteered. The full sample was roughly representative of the ethnic composition of the United States census (Self-identified ethnicity/race: African-American, 12% (58); Asian or Asian-American, 6% (31); Hispanic, 6% (32); Native American, 1% (4); White: 68% (340); Other/Not Reported: 7% (38)). The sample was above-average in education: less than 1 % (2) had no college; 80% (401) was in college or had been; 12% (59) was college graduates; 7% (34) had advanced degrees; information on the remainder 1 % (7) was unreported.

The Multifactor Emotional Intelligence Scale (MEIS)

Overview of Test Organization. The MEIS consists of 12 tasks, divided into four classes or "branches" of abilities including (a) perceiving, (b) assimilating, (c) understanding, and (d) managing emotion (Mayer & Salovey, 1997; Mayer et al., 1997). Branch 1's four tests measured emotional perception in Faces, Music, Designs, and Stories. Branch 2's two tests measured Synesthesia Judgments and Feeling Biases. Branch 3's four tests examined the understanding of emotion, including Blends, Progressions, and Transitions between and among emotions, and Relativity in emotional perception. Branch 4's two tests examined Emotion Management in Self and Others. The content of the subtests and their scoring is described below, as are the three scoring methods employed: consensus, expert, and target.

Branch 1: Perceiving Emotion

Branch 1 tasks concerned the ability to perceive and identify the emotional content of a variety of stimuli.

Faces (Eight Stimuli; 48 Items). The first Branch 1 task, Faces, used as stimuli eight faces from a CD-ROM photographic library and from personal photos, chosen to represent a variety of emotions, and for their authenticity in

representing those emotions. Each face was followed by six emotions: *happiness*, *anger*, *fear*, *sadness*, *disgust*, and *surprise*. The test-taker was to answer on a five-point scale whether a given emotion (e.g., anger) was "Definitely Not Present" (1) or "Definitely Present" (5). The responses were scored according to two criteria: consensus and expert.

Consensus Scoring. The group consensus served as the criterion for this scoring approach. Each participant response was scored according to its agreement with the proportion of the participant group who endorsed the same alternative. For example, if 0.51 of the participant group reported that anger was somewhat present ("4" on the scale), then a participant who chose "4" would receive 0.51 for the item. If the participant believed anger was definitely not present ("1" on the scale), and only 0.06 of the sample agreed, then the individual would receive a 0.06 for the item.

Expert Scoring. The first two authors served as experts for the tasks, and went through the test answering questions by bringing to bear, as much as possible, their reading of Western philosophical treatments of emotion, and their reading of contemporary psychological models of emotion. For example, in deciding questions about emotional blends, reference was made to the theory of emotional blends by Plutchik (1984). For each item, the authors identified the best alternative (from 1 to 5) for each response; general agreement with this best response (choosing the selected value, or the integer on either side of it) was scored "1"; otherwise, the individual received a "0."[4]

Music (Eight Stimuli; 48 Items). The second Branch 1 task, Music, was similar to the Faces task. The stimuli consisted of eight brief (5-10 seconds) original pieces of music composed for this project. Participants heard each piece of music and then rated each one as to its emotional content on a series of mood adjective scales. Each mood adjective was rated from 1 ("Definitely Not Present") to 5 ("Definitely Present"). The same six mood adjectives were employed as in Faces.

Target Scoring. The music test was scored according to the consensus and expert methods used above. In addition, the target scoring method was employed here. Target scoring made use of an additional data set. As the composer-musician worked, he was requested to think about his feelings and the feelings his music conveyed, which he then recorded on a mood scale. Target scoring was scored for agreement with the target's feelings (in this case, the composer-musician). It was scored as the expert scoring was, with a "1" for a match (give or take 1) and "0" for a non-match. Indeed, the target can be thought of as a second type of expert.

Designs (Eight Stimuli; 48 Items). The third Branch 1 task, Designs, was

---

[4] If the expert-selected value was "3" on the five-point scale, responses from 2 to 4 were assigned the value of 1 (correct). If the expert-selected value was "1," then 1-2 would be correct; if the expert value was "5," then 4-5 would be correct, etc.

identical to the above except that eight original computer-generated graphic designs served as the stimuli. The designer was requested to create graphics that portrayed a variety of feelings. As the designer worked, he recorded his feelings on the six-adjective mood scale about what he expressed in the design. Consensus, expert, and target scoring were employed for this task.

Stories (Six Stimuli; 42 Items). The fourth Branch 1 task, Stories, was identical to the above tasks except that six stories were employed. The stories were obtained as in Mayer and Geher (1996). Fifteen adult acquaintances of the authors were asked to report on situations or thoughts affecting their moods, including (a) "What led up to the situation?"; (b) "What is the situation, or what you are thinking about?"; and (c) "What happened in this situation which made you feel the way you do?" Immediately thereafter, these 15 supplemental participants recorded their moods on a 30-item mood-adjective checklist, using the five-point rating scale described above (see Faces). The passages were then edited lightly. The six passages were then presented to participants in the main study. An example was as follows:

> This story comes front a middle-aged man. Everything has been piling up at work and I am falling behind. I have been working late many nights and as a result, my wife and daughter are feeling left out. My relationship with them is being stressed. I feel that I am letting them down emotionally. I feel guilty not spending time with them. At the same time, a close family member moved in with us after his divorce and job loss. We have no privacy and I finally told him he has to move out. It was very difficult for me, especially since in the way I was raised, you don't treat a guest this way.

Each story was followed by a seven-adjective mood scale; the adjectives varied from story to story. They were selected so as to balance adjectives that were applicable to the story and those that were not, as well as to balance positive- and negative-toned adjectives. For the above story, the seven adjectives were, "depressed, frustrated, guilty, energetic, liking, joyous, and happy." The participant's job was to identify the emotion in the story. The responses were scored by consensus, expert, and target criteria.

Branch 2: Assimilating Emotions

Branch 2 tasks concerned the ability to assimilate emotions into perceptual and cognitive processes.

Synesthesia (Six Stimuli, 60 Items). The first Branch 2 task, Emotional Synesthesia, measured people's ability to describe emotional sensations and their parallels to other sensory modalities. The analysis of emotions often involves describing their composition in regard to other sense modalities, including movement, touch, pace, and color (Clynes, 1977; de Rivera, 1977). In this task, people imagined an event that could make them feel a particular feeling, which they then described on 10 semantic differential scales. For example,

one item asked, "Imagine an event that could make you feel both somewhat surprised and somewhat displeased . . . Now describe your feelings on," each of 10 five-point semantic differential scales, including "warm 1 2 3 4 5 cold," and other scales involving color (yellow or purple), touch (sharp or dull) and so forth; the scales were invariant across stimuli. This task was scored by consensus and expert criteria.

Feeling Biases (Four Stimuli; 28 Items). The second Branch 2 task, Feeling Biases, asked people to assimilate their present mood into their judgments of *how they felt toward a* [fictional] *person at the moment.* Thus, one task instructed participants to:

> Imagine that Jonathan is one of your relatives. He is a tall, muscular person. Jonathan said something to you that made you feel both guilty and afraid. Feeling both guilty and afraid about Jonathan, how does he seem?

The seven traits following each passage varied so as to be relevant to each passage; in the above example, traits included "sad, trusting, tense, cynical, aggressive, controlling, and hasty." The traits were rated on a five-point scale ("Definitely Does Not Describe" (1) to "Definitely Does Describe" (5)). The rationale for this task was that people who use their emotions in thinking do so, in part, by analyzing judgmental transformations that occur with mood. This task was scored according to consensus and expert criteria.

Branch 3: Understanding Emotions

Branch 3's tasks concerned reasoning about and understanding emotions.

Blends (Eight Stimuli, Eight Items). The first Branch 3 task, Blends, concerned the ability to analyze blended or complex emotions. Items were of the following form.

> Optimism most closely combines which two emotions?
> (a) pleasure and anticipation
> (b) acceptance and joy
> (c) surprise and joy
> (d) pleasure and joy.

Participants were instructed to select the single best answer. The eight items covered blends of two emotions (four items), blends of three emotions (two items), and blends of four emotions (two items). Scoring was by consensus and expert criteria.

Progressions (Eight Stimuli, Eight Items). The second Branch 3 task, Progressions, concerned people's understanding of how emotional reactions proceed over time, with a special focus on the intensification of feelings. A sample item read:

> If you feel angrier and angrier toward someone so that you are losing control, it would result in (choose one):
> (a) gloating
> (b) resentment
> (c) hate
> (d) rage.

Participants were instructed to identify the single best answer. Items were scored according to consensus and expert criteria.

Transitions (Four Stimuli; 24 Items). The third Branch 3 task, Transitions, concerned people's understanding of how emotions (and implicitly, the situations eliciting them) follow upon one another. Items were of the following form:

> A person is afraid and later is calm. In between, what are the likely ways the person might feel?

Each item was followed by six alternative feelings. Alternatives for the above item were acceptance, fear, anger, anticipation, surprise, and disappointment. The participant rated each item as "Extremely Unlikely" (1) to have occurred, or as "Extremely Likely" (5). The remaining three items followed the same form.

Relativity (Four Stimuli; 40 Items). The fourth Branch 3 task, Relativity, was composed of items depicting conflictual social encounters between two characters. The participant's task was to estimate the feelings of both those characters. One item read:

> A dog is chasing sticks outside when he runs out in the street and gets hit by a car. The driver stops when the dog's owner dashes over to check on the dog.

The first items concern the dog-owner's feelings. Participants must decide to what extent the dog owner feels each of five ways, including, "ashamed about not being able to have better trained the dog," or "challenged to protect other dogs from mishaps." Each alternative was rated according to how likely a feeling-reaction was, from "Extremely Unlikely" (1) to "Extremely Likely" (5). Next, the participant made similar judgments as to the second character (the driver, above). In the above example, participants judged whether the driver felt "relief that it was only a dog," or "guilty for not being a more cautious driver," and so on, on the same response scale.

Branch 4: Managing Emotions

Branch 4 concerns the ability to manage emotions.

Managing Feelings of Others (Six Stimuli; 24 Items). The first Branch 4 task, Managing Feelings of Others, examines how participants manage the emotions of others. Participants were asked to evaluate plans of action in response to

fictional people, described in brief vignettes, who needed assistance. The task consisted of six vignettes, each followed by four possible courses of action. For example:

> One of your colleagues at work looks upset and asks if you will eat lunch with him. At the cafeteria, he motions for you to sit away from the other diners. After a few minutes of slow conversation, he says that he wants to talk to you about what's on his mind. He tells you that he lied on his resume about having a college degree. Without the degree, he wouldn't have gotten the job.

Participants were to rate alternatives such as (for the above vignette):

> Ask him how he feels about it so you can understand what's going on. Offer to help him, but don't push yourself on him if he really doesn't want any of your help.

Participants rated responses from "Extremely Ineffective (1)" to "Extremely Effective (5)." Tasks were scored according to consensus and expert criteria.

Managing Feelings of the Self (Six Stimuli; 24 Items). The second Branch 4 task, Managing Feelings of the Self, concerns how a person would regulate his own emotions. This task consisted of six vignettes, each one describing a particular emotional problem. For example:

> You have been dating the same person for several months and feel very comfortable. Lately, you are thinking that this relationship may be the one and although marriage hasn't been discussed, you are assuming that it is a real possibility. The last thing you expected was the phone call you received saying that the relationship is over. You have lost the love of your life.

Participants were instructed that not every situation is equally applicable to everyone but to imagine, if in that situation, the effectiveness of given responses. One such response to the above situation was:

> The best way to cope with this terrible blow is to do whatever you can to block it out and not let it get to you any more than it has. You would throw yourself into your work or some activity and then try to put it behind you.

Participants rated each response from "Extremely Ineffective (1)" to "Extremely Effective (5)." Tasks were scored according to consensus and expert criteria.

Criterion Scales

Two classes of criterion scales were employed along with the MEIS. Primary criteria included measures of intelligence and self-reported empathic feeling, both of which have been predicted to correlate with emotional intelligence in

the past (Mayer et al., 1990; Salovey & Mayer, 1990). Secondary criteria included measures of several areas in which emotionally intelligent individuals are thought to differ from others. These include higher life satisfaction, a family environment that encourages learning about feelings, and aesthetic perception and participation (Mayer & Salovey, 1990; 1995; Salovey & Mayer, 1990).

Primary Criteria

Intelligence Measure. The intelligence criterion was adapted from the Army Alpha test of intelligence (Yerkes, 1921). The Army Alpha was employed because its validity is well-established and its form is ideal for group testing of the sort carried out here. The vocabulary scale was used because that subtest is the strongest component of verbal intelligence (e.g., Wechsler, 1958; Morrison, 1976, pp. 318-325). Thirty of the more difficult vocabulary items from the 50 were selected; more difficult items were favored so as to tailor the test to the participant population, which included mostly college-educated individuals. The Army Alpha vocabulary scale employs four response options for each word to be defined (e.g., "Reply: (1) make, (2) do, (3) answer, (4) come"). Participants were instructed to select the alternative from the list that most nearly meant the same as the target word. The scale had an alpha reliability of $\alpha = 0.88$ in this data set.

Empathy Measure. A 30-item empathy scale (Caruso & Mayer, 1999) was developed with content coverage similar to the Epstein-Mehrabian scale (Mehrabian & Epstein, 1972), but with identifiable factor-based subscales. This newer scale was employed so that overall self-reported empathy and also its subcomponents could be compared to emotional intelligence. The scale's overall self-reported empathy score has an alpha reliability of $\alpha = 0.86$. Because of content overlap with the Epstein-Mehrabian scale, it is likely to perform similarly to it. In contrast to the Epstein-Mehrabian, however, the present scale can be divided into five more specific factor-based scales. The five subscales, their reliabilities, and a sample item from each are: (a) Empathic Suffering, $\alpha = 0.79$, "The suffering of others deeply disturbs me"; (b) Positive Sharing, $\alpha = 0.72$, "Seeing other people smile makes me smile"; (c) Responsive Crying, $\alpha = 0.74$, "I cry easily when seeing a sad movie"; (d) Avoidance [*reversed*], $\alpha = 0.72$, "I find it annoying when other people cry in public"; and (e) Feeling for Others, $\alpha = 0.61$, "If someone is upset, I get upset too."

Secondary Criteria

Life Satisfaction. Each person was asked about his satisfaction with his Relationships, Academic Status, and Career and Work Situation, to be reported on a five-point rating scale (from "Not at All Satisfied" to "Extremely Satisfied"). A factor analysis indicated the items were unifactorial although they were only moderately intercorrelated. A single life satisfaction score ($\alpha = 0.59$)

was employed, representing the sum of each person's responses.

Artistic Skills. Participants also reported their degree of artistic skill in eight areas (from "1" no or little talent to "3" very talented). A unifactorial Artistic Skill score ($\alpha = 0.71$) indicated overall self-reported artistic skill in those areas, which included sculpture, music, and writing.

Parental Warmth. Participants also described their parents' behaviors on a seven-item scale. A unifactorial Parental Warmth factor ($a = 0.81$) included items reporting that parents were warm, listened, were non-abusive, and (reversed) yelled and were strict.

Psychotherapy. Psychotherapy was scored as the number of months a person had psychotherapy (which for some people was zero) multiplied by the number of sessions of psychotherapy per month.

Life Space Leisure. Life space scales consist of items that record a person's environment in terms of discrete, externally verifiable, responses (e.g., "How many pairs of shoes do you own?" "How many times have you attended the theater in the last year?"; Mayer, 1998; Mayer, Carlsmith, & Chabot, 1998). Certain life activities, particularly those involving aesthetic appreciation, have been predicted to involve more emotional intelligence than others (Mayer et al., 1990). Thirty-three items concerning leisure activities were administered to participants covering books read, television watched, and cultural events observed. These items yielded three factor-based scales of leisure activities, based on an unrotated, principal components analysis, using all items loading on a given factor $r > \pm 0.45$. The first, Culture-Seeking, factor scale ($a = 0.78$) loaded the following items: listening to classical music, attending concerts, listening to soul, listening to gospel, listening to country, listening to rap, listening to new-age music, listening to bluegrass, listening to rock, and attending museums. The second, Improvement-Seeking, scale ($a = 0.60$), loaded reading self-help books, how-to books, medical books, business books, and short stories. The third, Entertainment-Seeking, factor scale ($a = 0.67$), loaded watching action television programs, watching comedy programs, listening to punk music, listening to blues music, and watching televised sports.

Procedure

Participants completed the study in small groups or individually. Each participant received an item and answer booklet that contained all necessary instructions, test items and responses. The test was not timed and the test materials were self-administered, with the exception of the music task, for which a tape of instructions and music was played by the experimenter in group settings.

Results

This section is divided into three parts. First, scoring methods for emotional intelligence are compared. Second, the emotional intelligence tasks are intercor-

related and factor analyzed. Third, emotional intelligence is correlated with various external criteria.

Scoring for Emotional Intelligence

Consensus, Expert, and Target Criteria for Correct Answers

Emotional intelligence depends on the idea that certain emotional problems have answers that can be judged correct and incorrect. Convergence among different scoring criteria provides a foundation for such assumptions. The data analysis began by comparing the three different methods for identifying a correct answer: according to (a) the group consensus, (b) expert's identification, and (c) a target's assessment (for three tasks only). We began by examining the degree to which these three methods converged toward a correct response. To the extent that the group consensus (as identified by the modal response) and experts agree as to the best answer, their selections should intercorrelate over the items of a given test. For example, if both the group consensus and experts agree that anger is high in one story ("4" or "5"), but low in another ("1" or "2") then the correlations should be high.

To test the relation between consensus and expert ratings, we selected four tasks, each drawn randomly from one of the four branches, and calculated the intercorrelation between ratings. The four tasks collectively contain 127 items, representing each branch, and provide a good estimate of the test's overall pattern. For each item, we paired the modal consensus choice with the specific

Table 1. Means, Standard Deviations, and Reliabilities (Coefficient Alpha) of the Agreement with Consensus, Expert, and Target Criteria

	Scoring Method								
	Consensus			*Expert*			*Target*		
Branch and Task	*M*	*S*	α	*M*	*S*	α	*M*	*S*	α
Emotional Identification									
1: Faces	0.40	0.08	0.89	0.64	0.11	0.74	–	–	–
1: Music	0.44	0.11	0.94	0.73	0.10	0.86	0.75	0.12	0.88
1: Designs	0.36	0.08	0.90	0.69	0.11	0.74	0.65	0.14	0.81
1: Stories	0.38	0.07	0.85	0.72	0.11	0.72	0.66	0.10	0.61
Assimilating Emotions									
2: Synesthesia	0.31	0.04	0.86	0.69	0.09	0.66	–	–	–
2: Feeling biases	0.30	0.05	0.70	0.72	0.12	0.60	–	–	–
Understanding Emotions									
3: Blends	0.49	0.10	0.49	0.60	0.19	0.35	–	–	–
3: Progressions	0.58	0.10	0.51	0.83	0.16	0.50	–	–	–
3: Transitions	0.30	0.04	0.94	0.56	0.11	0.85	–	–	–
3: Relativity	0.30	0.04	0.78	0.56	0.11	0.63	–	–	–
Managing Emotions									
4: Managing others	0.28	0.04	0.72	0.60	0.12	0.42	–	–	–
4: Managing self	0.27	0.04	0.70	0.55	0.12	0.40	–	–	–

Table 1. Means, Standard Deviations, and Reliabilities (Coefficient Alpha) of the Agreement with Consensus, Expert, and Target Criteria

expert selection. In fact, the consensus and expert ratings were fairly highly intercorrelated across tasks: Stories (Branch 1) r = 0.70; Feeling Bias (Branch 2) r = 0.64; Relativity (Branch 3) r = 0.61; and Managing Feelings of Others (Branch 4) r = 0.80. (All rs were significant, $p < 0.0001$ level). This suggests that the two criteria are closely related.

The third, Target criterion, was available only for three Branch 1 tasks. This criterion involves reports by the Targets of their actual feelings as they were creating their artistry (Music and Designs) or telling how they felt (Stories). In general, Consensus correlated a bit more highly with the Target criteria than did the Expert criterion (Music task: r = 0.61 for consensus, r = 0.52 for expert; Designs: r = 0.70 and 0.60; Stories: r = 0.80 and 0.69).

Generally speaking, the three criteria appeared to correlate moderately highly, indicating that some answers were "more correct" than others, according to any and all of the scoring methods used: consensus, expert, and target.

Agreement Scoring for Consensus, Expert, and Target Criteria

A given participant's performance can be assessed in relation to each of the above three scores: Consensus, Expert, and Target. These agreement scores represent the degree to which a given participant's responses coincided with those of the criteria. The means, standard deviations, and reliabilities of the participants' performance, broken down by the three scoring methods, can be seen in Table 1. The means are not directly comparable across consensus, expert, and target because of their substantially different scoring approaches (see above). The figures do indicate, however, the average performance level of the sample, and also that there were no problems of floor or ceiling effects in any of the three scoring methods.

The reliabilities of the agreement scores are also promising, with individual tasks having reliabilities most often between $\alpha = 0.70$ and 0.94 for consensus, and a bit lower for expert agreement. The first two tasks of Branch 3, which were also the shortest, had lower reliabilities, $\alpha = 0.35$-0.51; the Branch 4 management tasks were also low, but for expert scoring only. All the reliabilities are satisfactory for this exploratory study concerning the factorial structure of emotional intelligence and what it predicts.

Sex Differences in Performance

Women performed somewhat higher than men on the 12 tasks, according to all the scoring procedures. The difference was 0.5 standard deviation for consensus agreement ($M_{women} = 0.376$; $SD_{women} = 0.029$; $M_{men} = 0.358$; $SD_{men} = 0.036$; Hotelling's $F(12, 409) = 4.0$, $p < 0.001$.), and about 0.1 standard deviation for each of expert agreement ($M_{women} = 0.664$; $SD_{women} = 0.048$; $M_{men} = 0.657$; $SD_{men} = 0.061$; Hotelling's $F(12, 408) = 4.7$, $p < 0.001$), and target agreement ($M_{women} = 0.689$; $SD_{women} = 0.093$; $M_{men} = 0.676$; $SD_{men} = 0.079$; Hotelling's $F(3,482) = 1.34$, n.s.). This replicates earlier similar

findings (cf. Buck, 1984; Mayer & Geher, 1996). We endeavored to understand more about this difference by focusing on the Story task in particular, which was representative of the full test according to subsequent factor analyses (see below). The Story task also showed the greatest sex differences, and contained all three scoring criteria. It is possible that women outperformed men using consensus scoring because the women were using a women's criterion which was different than the men, and the larger number of women in the sample ($N = 333$ vs. 164) meant that the women's choices were scored with higher values than the men's. This, however, did not account for the women's slightly better performance. Women and men seemed to be employing close to the same criterion. The correlation between women's and men's choices for the emotional content across the 42 story items (six stories, seven items each) was $r(42) = 0.993$, indicating a high level of agreement (nor was there any difference in the average emotion-level perceived on an item: $M_{women} = 2.67$; $SD_{women} = 1.19$; $M_{men} = 2.69$; $SD_{men} = 1.04$; $t(41) = 0.77$, n.s.).

We further examined women's and men's performance by employing a two (male participant/female participant) by two (male story character/female story character) by two (male-selected consensus/female-selected consensus) MANOVA on the story data. The MANOVA yielded a main effect representing the women's better consensual accuracy ($F(1,495) = 20.08$, $p < 0.001$). Women outperformed men under all conditions, even using male-chosen consensus across stories ($M_{women} = 0.37$; $SD_{women} = 0.056$; $M_{men} = 0.35$; $SD_{men} = 0.059$; $t(495) = 3.68$, $p < 0.001$). A second main effect indicated that the participants, as a whole, were more accurate when using women's consensus criteria over men's ($F(1,495) = 928.7$, $p < 0.001$). A sex by consensus interaction indicated that women did slightly better using their own consensus criterion ($F(1,495) = 67.3$, $p < 0.001$). There was also a sex-of-target effect that favored judgments concerning male targets ($F(1,495) = 696.3$, $p < 0.001$). Collectively, these results indicate that women generally do better than men on these tasks and that the results are not caused by any simple bias in the test materials or how they are scored. Moreover, if one judges by the consensus scoring (which may be fairest, as the expert's criteria were developed by the male authors), the difference between women and men's performance is a moderate 0.5 standard deviation in size.

Correlations among Consensus, Expert, and Target-scored Tasks

Consensus and Expert Scoring Considered Individually. One correlational standard for an intelligence is that it defines a cluster of interrelated abilities (Guttman & Levy, 1991). The following analyses examine the intercorrelations among the 12 tasks to see if they show a "positive manifold"; i.e., a correlation matrix in which most tasks correlate positively with one another. Correlations among the 12 emotional intelligence tasks were calculated using all three scoring methods. In each case, a positive manifold was evident. Scored by the con-

sensus method, the tasks mostly correlated with one another between r = 0.20 and 0.50, with the full range spanning r = 0.07-0.68. Scored by the expert method, the tasks mostly correlated with one another r = 0.10-0.40, with a full range from r = 0.00 to 0.54. In either case, the matrix possesses a positive manifold; almost all the tasks are positively intercorrelated, as expected in regard to a unified intelligence. The consensus-scored tasks (with alpha reliabilities on the diagonal) can be seen in Table 2.

Table 2. Intercorrelations of the Consensus Scored Tasks with Reliabilities (Coefficient Alpha) on the Diagonal*

	1				*2*		*3*				*4*	
Branch and Task	*Fa*	*Mu*	*De*	*St*	*Sy*	*Fe*	*Bl*	*Pr*	*Tr*	*Re*	*Mo*	*Ms*
Emotional Identification												
1: Faces	**0.89**											
1: Music	0.61	**0.94**										
1: Designs	0.68	0.60	**0.90**									
1: Stories	0.54	0.47	0.54	**0.85**								
Assimilating Emotions												
2: Synesthesia	0.24	0.24	0.26	0.38	**0.86**							
2: Feeling biases	0.30	0.24	0.35	0.47	0.39	**0.71**						
Understanding Emotions												
3: Blends	0.07	0.13	0.09	0.24	0.22	0.26	**0.49**					
3: Progressions	0.10	0.15	0.14	0.25	0.34	0.35	0.41	**0.51**				
3: Transitions	0.25	0.29	0.29	0.37	0.26	0.34	0.19	0.17	**0.94**			
3: Relativity	0.30	0.35	0.32	0.41	0.32	0.38	0.30	0.34	0.43	**0.78**		
Managing Emotions												
4: Managing others	0.20	0.21	0.20	0.28	0.25	0.24	0.16	0.22	0.18	0.37	**0.72**	
4: Managing self	0.19	0.15	0.14	0.30	0.27	0.22	0.20	0.23	0.17	0.25	0.54	**0.70**

*N = 500. Note that correlations above $r \approx 0.08$ are significant at beyond the $p < 0.01$ level.

Comparisons among Consensus, Expert, and Target Scoring. To further compare scoring methods, we examined participants' performance on each of the 12 tasks, scored according to a consensus, expert-scoring, or target criterion. For consensus and expert scoring (which were available for all 12 tasks), participants' performance, scored each way, correlated between r = -0.16 and 0.95, with half the tasks above r = 0.52. The only negative correlation (r = -0.16), which occurred for Faces, and other low correlations for the Designs task (r = 0.24) may have been a consequence of different color photocopying employed to reproduce the stimuli for the groups and the experts.

The convergence for participant's consensus and target scores for the Music, Designs, and Stories tasks (where target scoring was available) were r = 0.81, 0.22, and 0.43, respectively; the same values for expert and target scoring were r = 0.67, 0.46, and 0.16. Subsequent analyses indicated the general superiority of the consensus scoring method in relation to the other alternatives. It yielded higher alpha test reliabilities for every task without exception, clearer factor results (which were, nonetheless, highly similar to expert scoring), and higher correlations with criteria. The superiority of consensus scoring has been argued persuasively elsewhere (e.g., Legree, 1995). For that reason, the subsequent analyses focus on the consensus scoring.

Factor Structure of the MEIS

Our next question was whether emotional intelligence is best characterized as one or many abilities. Although a highly developed theory of emotional intelligence motivates this article, this represented our own first empirical examination of so many tasks. For that reason, we employed exploratory factor analysis at the outset. We therefore applied principal axis factoring (with communalities on the matrix diagonal) to scores on the 12 MEIS subscales. A joint scree/meaningfulness criterion yielded a three-factor solution (first six eigenvalues: 4.3, 1.6, 1.1, 0.9, 0.8, 0.6). Table 3 (left columns) shows the three-factor, unrotated solution for the 12 consensus-scored subscales. We then further analyzed this three-factor solution by rotating it according to an oblique criterion (using an oblimin procedure). The right columns of Table 3 show this result.

In the unrotated solution, the first factor may be interpreted as a general emotional intelligence (g_{ei}) because it loads all the tasks without exception. This g_{ei} apparently represents a group factor of emotional intelligence tasks, suggesting their interrelatedness (below, we explore this question further). The second factor, Managing vs. Perceiving Emotions, discriminates tasks high in reasoning from those high in simple emotional perception. And the third factor, Managing Emotions, describes the two Branch 4 tasks concerning regulating emotions in oneself and others.

The rotated version of this three-factor solution tells the same story from a different angle. The first factor, Emotional Understanding, loads most of the tasks on Branch 3 (Understanding), along with tasks on Branch 2 (Assimilation). The second factor, Emotional Perception, loads most of the tasks on Branch 1 (Perception). The third, Managing Emotion, factor loads the two Branch 4 (Regulation) tasks, as in the unrotated solution. In this analysis, oblimin factors 2 and 3 had uniformly negative loadings. We reversed the loadings in sign so that a higher score indicated a higher level of ability across tasks. We similarly changed the sign of factor scores and scales based on these two factors. This procedure simplifies the presentation and discussion of results while remaining consistent with the substantive findings.

As a pattern matrix should do, this solution "turns up the contrast" on the loadings, separating the test into three portions: perception, understanding, and managing. This is done, in part, by transferring the common variance shared among the individual tasks to the three factors underlying them. As a consequence, the three factors intercorrelate fairly substantially. Perception correlated $r = 0.39$ with Understanding and $r = 0.49$ with Management; the latter two intercorrelated, $r = 0.33$.

Table 3. Three-factor Solutions for the Emotional Intelligence Test Scored According to Consensus and According to Expert Criteria, in Unrotated and Rotated Solutions: Principal Components Factoring[a]

		Unrotated			*Oblique Rotated (Pattern Matrix)*[b]		
Solution	*Branch/Task*	*I*	*II*	*III*	*I*	*II*	*III*
Unrotated Solution							
Emotional Identification							
	1: Faces	**0.67**	**−0.48**	**−0.11**	**−0.10**	**0.86**	**0.04**
	1: Music	**0.63**	**−0.34**	**−0.04**	**0.02**	**0.70**	**0.02**
	1: Design	**0.69**	**−0.44**	**−0.02**	**0.01**	**0.82**	**−0.03**
	1: Stories	**0.73**	**−0.09**	**0.05**	**0.30**	**0.52**	**0.08**
Assimilating Emotions							
	2: Synesthesia	**0.51**	**0.19**	**0.10**	**0.43**	**−0.12**	**0.10**
	2: Feeling biases	**0.59**	**0.13**	**0.21**	**0.53**	**−0.20**	**−0.00**
Understanding Emotions							
	3: Blends	**0.35**	**0.32**	**0.24**	**0.57**	**−0.10**	**−0.01**
	3: Progressions	**0.43**	**0.38**	**0.25**	**0.64**	**−0.11**	**0.02**
	3: Transitions	**0.48**	**0.04**	**0.12**	**0.35**	**0.23**	**0.00**
	3: Relativity	**0.61**	**0.18**	**0.09**	**0.45**	**0.20**	**0.14**
Managing Emotions							
	4: Managing others	**0.49**	**0.36**	**−0.49**	**−0.05**	**0.00**	**0.81**
	4: Managing self	**0.44**	**0.36**	**−0.38**	**0.03**	**−0.03**	**0.68**

[a]Loadings above ±0.25 are in bold typeface for clarity.

[b]Loadings indicated that all three factors were unipolar (i.e., loadings on a factor above ±0.25 all shared the same sign). Rotated factors II and III, however, were negative. To clarify results and facilitate discussion, loadings on rotated Factors II and III were reversed in sign here and in subsequent analyses.

The above results provide strong empirical support for a three-factor model of the MEIS. Recall that our theoretical model involves a four-branch model. We wondered whether there was also evidence for a four-factor model. To fully investigate this possibility, we modeled the data as a four-factor solution using covariance structural modeling. We used a stringent model in which each task was forced to load only on its hypothesized factor and no other (e.g., Arbuckle, 1997, p. 396). The factors themselves, however, were allowed to intercorrelate as above. The model fit was sufficiently good to be informative, with a Root Mean Square Error of Approximation (RMSEA) of 0.09 with no relaxation of parameters[5] (one rule of thumb is that a RMSEA 0.05 indicates a close fit; Browne & Cudeck, 1993). As appealing as this four-factor model is to us, the drawback is that the model estimates two of the factors, Assimilation and Understanding, to intercorrelate $r = 0.87$, which makes them difficult to distinguish from one another. For that reason, we continue to focus on the three-factor model in our analyses, while acknowledging that the four-factor model remains viable.

[5] Simply allowing the Stories task to load on the Understanding (as well as Perception) moves the RMSEA index to 0.077.

Hierarchical Relations among Factors and the Creation of MEIS Scales

The first unrotated factor of the MEIS was earlier said to represent a g_{ei}, or general factor of the test. Such general factors sometimes can arise spuriously due to the nature of principal axis factoring. For that reason, it is often recommended that a hierarchical factor analysis be employed as a secondary check of the existence of a hierarchical factor (e.g., Carroll, 1993; Jensen & Weng, 1994). Obtaining a hierarchical (second-order) factor that loads all the primary factors is generally considered stronger evidence for a general factor because it is based solely on the covariances among the primary factors. A new factor analysis was therefore conducted on the Perception, Understanding, and Managing factor scores. A single hierarchical factor was extracted that loaded Perception, Understanding, and Management at substantial levels (r = 0.50, 0.86, and 0.75, respectively).[6] This hierarchical factor correlated with the unrotated first factor of the principal axis factoring at r = 0.94. This final result indicates that general emotional intelligence can be reasonably represented by the first unrotated principal axis factor, and that it loads all the scales studied here.

The Construction of Factor-based Scales

For our further analyses, we first constructed factor scales for Perception, Understanding, and Managing Emotions factors (and for General Emotional Intelligence, g_{ei}, based on the first unrotated factor). The scales were constructed by summing z-scored subscale scores from tasks that loaded on the factors above $r = \pm 0.35$. The resulting factor-based scales were then correlated with the original factor scales (based on a weighted sum of all 12 scales) to ensure that they represented the original scales adequately. The three factor-based scales representing perception, understanding, and managing, correlated very highly with their respective factor scales (r = 0.98, 0.97, and 0.98, respectively), and were highly reliable (α = 0.96, 0.92, and 0.81). The three factor-based scales were moderately intercorrelated (Perception with Understanding, r = 0.44; Perception with Managing; r = 0.29; Understanding with Managing, r = 0.43). The overall General Emotional Intelligence factor-based scale also correlated with its original factor scale r = 0.97; and had a reliability of α = 0.96. These are the scales reported in the rest of the article.

Researchers wishing to retain the four-branch theoretical model (modestly supported by covariance structural modeling above) may wish to employ four, rather than three, factor-based scales. The above three scales may be transformed into four by (a) retaining the Perception and Managing scales as calcu-

---

[6] As with Oblique factors II and III, we reversed the sign of the hierarchical factor so that a higher score reflected better ability.

lated above, and (b) splitting the Understanding scale, above, into two scales. The first of these two scales, the revised Understanding scale, is calculated as the sum of the z-scores of the Blends, Progressions, Transitions, and Relativity tasks. The second of these two scales, the new Assimilation scale, is calculated as the sum of the z-scores of the Synesthesia and Feeling Biases tasks.[7] The a reliabilities of the (unchanged) Perception, (new) Assimilation, (revised) Understanding, and (unchanged) Management factor-based scales are, respectively, a = 0.96, 0.86, 0.89, and 0.81. The Understanding factor-based scale still correlates with the original factor scale r = 0.89. The new Assimilation factor-based scale correlates with Understanding, r = 0.65.

Relation of the Emotional Intelligence Factors to Criterion Measures

The final correlational criterion for an intelligence is that it correlates moderately with intelligences in other domains. The correlation should be high enough to indicate that the new skill is an intelligence, but low enough to illustrate that it says something new about human abilities. Aside from emotional intelligence's correlation with verbal intelligence, emotional intelligence will be important to the degree that it predicts other criteria as well.

Table 4 shows the correlation of the emotional intelligence factors with various criteria. The central correlations to examine are those with the gei factor (first column). General emotional intelligence is then divided into subfactors of perception, understanding, and management; correlations with those subfactors are shown in the next three columns.

The correlation between the General Emotional Intelligence factor-based scale and verbal intelligence is $r = 0.36$, $p < 0.001$. This is the moderate level at which one would hope that a new domain of intelligence would be correlated with existing domains. In addition, emotional intelligence has a number of interesting correlations with other variables. The g_{ei} factor-based scale correlates $r = 0.33$, $p < 0.001$ with overall empathy, also as expected (Salovey & Mayer, 1990; Mayer et al., 1990; Mayer & Geher, 1996), and possesses a number of significant correlations with subtypes of empathy as well, correlating positively and at similar levels with Suffering and Positive Sharing, and negatively with Avoidance. Emotional intelligence had a positive correlation with parental warmth, $r = 0.23$, $p < 0.01$, and a negative correlation with pragmatic attempts at self-improvement $r = -0.16$, $p < 0.01$, including reading self-help books, books on business methods, and the like. The subfactor scales further qualify the relations, suggesting that Understanding is most closely related to verbal intelligence among the three subfactors, and that Management most accounts for empathy; all three subfactors are related to Parental Warmth.

---

[7] These two tasks had estimated loadings of 0.51 and 0.59, respectively, on the assimilation factor, of the oblique four-factor model.

Table 4. Correlations Between Individual Tasks and Selected Criterion Variables

	Overall Score	*Subfactor Scores*		
Criterion Variables	g_{ei}	*Perception*	*Understanding*	*Management*
Primary Criteria				
Ability				
Verbal IQ	0.36**	0.16**	0.40**	0.20**
Empathy				
Overall	0.33**	0.20**	0.25**	0.34**
Suffering	0.35*	0.18**	0.28**	0.37**
Positive Sharing	0.26**	0.12**	0.16*	0.36**
Crying	0.14**	0.10*	0.10*	0.13**
Avoidance	−0.26**	−0.20**	−0.23**	−0.15**
Feeling for others	0.16**	0.08	0.09*	0.24**
Secondary Criteria				
Life satisfaction	0.11*	0.01	0.11*	0.13**
Artistic skills	0.05	0.03	0.07	0.00
Parental warmth	0.23**	0.20**	0.18**	0.15**
Psychotherapy	0.03	0.04	0.14*	0.02
Leisure (life space)				
Culture-seeking	0.00	−0.07	0.01	0.03
Self-improvement	−0.16**	−0.07	−0.22**	−0.05
Entertainment	−0.02	0.09*	−0.04	0.05

*$p < 0.05$ level.
**$p < 0.01$ level, two-tailed tests.

An extremely stringent test would partial verbal IQ and self-reported empathy out of the correlation between emotional intelligence and the six secondary criteria. Doing this may remove variance that legitimately belongs to emotional intelligence, but it also ensures that emotional intelligence contributes unique variance in predicting criteria. Partialing out the influence of intelligence and empathy yielded a g_{ei} that maintained its significant negative correlation with attempted self-improvement ($r = -0.10$, $p < 0.05$), and added a negative relation to culture-seeking ($r = -0.09$, $p < 0.05$), although it no longer correlated with life satisfaction or parenting.

Although the statistical relations between emotional intelligence and the life space criteria may seem low, two things are worth noting about them. First, our central focus has been on understanding the structure of emotional intelligence. The few secondary criterion scales included here were exploratory and brief. Even these crude measures, however demonstrate that emotional intelligence predicts criteria independent of the influence of both verbal intelligence and empathy. Second, it is worth recalling that personality relations tend to be small but consistent over the years. This small but consistent influence can substantially change a person's position in life, just like a slow but steady current can move a boat a considerable distance across a lake over time.

Summary and Discussion of Study 1

The results from Study 1 indicate that emotional intelligence shows a pattern that is consistent with a new domain of intelligence. Emotional intelligence can be operationalized as sets of abilities, and better answers can be distinguished from worse answers, as indicated by the convergence of three scoring methods. The 12 tasks also intercorrelate with one another, independent of which scoring method is employed. The scale yields four scores: A first, superordinate factor of general emotional intelligence that provides one excellent and economical method for representing the concept. The General Emotional Intelligence factor can be divided in turn into three subscales: Perception, Understanding and Managing (thus reducing our four-branch model to a three-branch model). Finally emotional intelligence correlates moderately with a measure of verbal intelligence indicating that it is related to other intelligences without being the same as them. Emotional intelligence shows promise as a predictor of other qualities such as empathy, (retrospective) parenting style, and life activities.

Study 2

Thus far, emotional intelligence has met two of three important criteria of a traditional intelligence. First, it has been operationalized as a set of abilities. Second, it has shown a pattern of correlations consistent with the existence of such an intelligence. The third criterion is that intellectual capacities grow with age and experience from childhood to early adulthood (Brown, 1997; Fancher, 1985). The importance of age to intelligence was first recognized by Binet. As Fancher (1985, p. 71) describes it:

> Gradually . . . a key insight developed — one which seemed perfectly obvious once recognized, but which nevertheless had previously eluded Binet and other investigators of intelligence. Age was a crucial factor to be considered: both subnormal and normal children might learn to pass the same tests, but normal children did so at a younger age.

Fancher attributes Binet's success in measuring intelligence, in comparison to the failures of his contemporaries, to the realization that mental abilities grow with age and experience.

For emotional intelligence to behave as does a standard intelligence, it should be shown to increase with age. To test whether this actually occurs, several portions of the scale employed in Study 1 were administered to a young adolescent sample (ages 12-16) in Study 2. The performance of the adolescents was then compared to the performance of an adult subsample drawn from Study 1. The use of two samples close in age ensures that the same test items can be used and understood by both groups. It also provides a challenging test of the

developmental hypothesis because proximity in age should yield only small differences in performance between the two groups. We hypothesized that the adult sample would significantly outperform the adolescents on the scale.

Method

Adolescent Sample

Participants were 229 adolescents (125 young men, 101 young women; 3 unidentified) with a mean age of 13.4 (range 12-16) who were recruited from two independent secondary schools and a religious youth group. These were split among 35% (81) 7th graders, 36% (83) 8th graders, 9% (20) 9th graders, 12% (27) 10th graders, 6% (13) 11th graders, and 1% (2) 12th graders; (percentages add to 99% due to rounding error). The sample deviated somewhat from the ethnic composition of the United States census in under-representing minority groups (Self-identified ethnicity/race: African-American, 5% (12); Asian or Asian-American, 3% (6); Hispanic, 3% (7); Native American, 0% (0); White: 79% (177); Other/Not Reported: 9% (27)).

Adult Sample

The adult sample from Study 1 was again used in Study 2. Here, however, the adult sample was divided on the basis of subject number into two equal-sized samples: the "Independent Adult Sample" and the "Consensus Sample." The Independent Adult Sample served as the comparison group for the adolescent group. The Consensus Sample was used to calculate a consensus score to which the first, "Independent Sample," had not contributed.

Materials

For reasons of time and age-appropriateness, only a subset of the scales administered to adults was administered to the developmental sample. These included Faces, Music, Designs, and (age-appropriate portions of) Stories from Branch 1,[8] Synesthesia from Branch 2, and Blends and Relativity from Branch 3. In addition, the Army Alpha Vocabulary scale and the Empathy scale were administered as criteria.[9]

---

[8] Adults and adolescents were compared on two of the eight stories as six stories were deemed potentially unsuitable to adolescents, using extremely cautious criteria, due to their content.

[9] Several additional scales were administered that had been rewritten for a younger age group. Reports on the downward extension of the emotional intelligence test can be found elsewhere (see Caruso, Van Buren, Mayer & Salovey 1998). Only those tests that were identical across groups are examined here because only those are relevant to the developmental hypothesis examined here.

Procedure

Parental consent was first obtained for each participant in the adolescent group, and then informed consent obtained from each subject. All data were collected anonymously; no names were requested. Furthermore, subjects were explicitly instructed not to answer any questions that made them uncomfortable.

Participants in the developmental sample were tested in a similar manner to the adults. They completed the materials in small groups. Each participant received an item and answer booklet that contained all necessary instructions, test items and responses. For the music task, a researcher (or a classroom teacher) played the cassette tape that included all necessary instructions as well as the musical selections. Students required 45-75 minute to complete the test booklet.

Results

Scoring

Three scoring procedures were employed as in Study 1: agreement with consensus, expert ratings, and target reports. Some modifications in the consensus scoring were necessary for this study. Using the adult consensus as in Study 1 would plainly favor adults because each adult's score contributed to the consensus. To control for this, the adult sample was divided in half (on the basis of odd/even subject number). Next, new adult consensus scores were calculated for the even half of the sample only (the consensus sample). This left the odd half of the adult sample with responses that were independent of the adult consensus (the independent sample). It was this "independent" adult sample whose consensus scores were compared to the adolescent's consensus scores. Expert-scoring and target-scoring were the same as in Study 1.

Adult-Adolescent Comparisons

The central purpose of Study 2 was to examine whether adults functioned at a higher level of emotional intelligence than adolescents. This hypothesis was tested via a two (Age-Group) by seven (Task) ANOVA, where the seven tasks were within-subjects variables. As the developmental hypothesis predicted, scores were higher for adults than for adolescents for consensus agreement (Grand Mean = 0.38 vs. 0.36; $F(1,713) = 23.8, p < 0.001$), for expert agreement (Grand Mean = 0.66 vs. 0.64; $F(1,709) = 22.3, p < 0.001$) and for target agreement (Grand Mean = 0.69 vs. 0.67; Hotelling's $F(1,718) = 8.0, p < 0.01$). Significant Task and Age x Task effects were also present for all three scoring methods. Focusing on consensus scoring, there was a significant Age-Group x Task interaction for consensus ($F(6,708) = 12.5, p < 0.01$). Table 5 shows a more detailed comparison between the adult and adolescent groups for consensus scoring on the individual tasks they both received, and which tasks showed significant differences in the predicted direction on their own. We did not exam-

ine age-performance correlations beyond demonstrating these average differences. First, the two samples were not strictly comparable, as the adolescents were of slightly higher social class (and therefore would be expected to perform more highly than average). More seriously, the restriction of age-range in both samples would render the correlation impossible to assess. We did, however, determine that the adolescents' scores showed the same pattern of correlations with verbal intelligence ($r(220) = 0.45$; $p < 0.001$) and empathy ($r(227) = 0.37$, $p < 0.001$) as did the adults.

Table 5. A Test of the Developmental Hypothesis: Means (and Standard Deviations) of Adult vs. Adolescent Performance on Selected Consensus-scored Scales of Emotional Intelligence

	Adult		*Adolescent*		
Branch/Task	*M*	*SD*	*M*	*SD*	*F*
Emotional Identification					
1: Faces	0.400	0.078	0.384	0.075	5.2*
1: Music	0.445	0.092	0.438	0.074	0.5
1: Design	0.359	0.086	0.353	0.077	0.7
1: Stories	0.328	0.069	0.323	0.061	1.5
Assimilating Emotions					
2: Synesthesia	0.306	0.045	0.295	0.047	7.8**
Understanding Emotions					
3: Blends	0.491	0.087	0.424	0.105	52.2**
3: Relativity	0.307	0.053	0.304	0.058	0.2
Combined tests	0.378	0.046	0.359	0.048	25.6**

*$p < 0.05$.
**$p < 0.01$.

Discussion of Study 2

Study 2 tested whether emotional intelligence met the third of three criteria for a standard, conventional intelligence: whether ability levels increase with age. As predicted, adults performed at higher ability levels than do adolescents. In addition, emotional intelligence in adolescence shows the same relations to verbal intelligence and empathy as with adults.

General Discussion

Three major criteria for a standard intelligence are that it consists of mental abilities, that those abilities meet certain correlational criteria, and that the abilities develop with age. In the tests conducted here, emotional intelligence met all three criteria. First, emotional intelligence could be operationalized as a set of ability tests. Second, performance on those ability tests was intercorrelated and partly distinct from verbal intelligence, against which they were compared. Third, emotional intelligence was shown to grow from early adolescence to young adulthood. Collectively, these findings bring us a major step forward toward demonstrating a plausible case for the existence of this intelligence. The

data also tell us about the structure of emotional intelligence, and what it might predict.

The Nature of Emotional Intelligence and Its Measurement

Our factor analyses of the 12 MEIS tasks suggest that one can best conceptualize emotional intelligence as involving three primary factors, and a higher order, General Emotional Intelligence factor that combines the three. The three primary factors involve Perception, Understanding, and Managing of emotion. Perception skills include those drawn from the first branch of the model, including recognizing emotions in Faces, Music, Designs, and Stories. Understanding skills include those drawn from the Assimilation and Understanding branches of the model, including Synesthesia, Feeling Biases, Blends, Progressions, Transitions, and Relativity. Finally, Managing emotions represents skills drawn from the fourth branch of the model, including Managing Others and Managing the Self. The three primary factors, in other words, could be said to capture the four branches of our most recent model of emotional intelligence (Mayer & Salovey, 1997): The Perception branch was captured by the Perception factor scale, the Assimilation and Understanding branches were combined into a single Understanding factor scale, and the Managing Branch was captured by the Managing factor scale. This three-branch measurement approach is also broader than our original 1990 model (which omitted the Understanding branch). The three primary facets of Perception, Understanding, and Management clearly emerged from the data, and although one still might possibly develop a four-branch measure, it would apparently require developing substantially different Branch 2 tasks than the ones employed here, so as to better distinguish them from the Understanding branch.

The three primary abilities appear to be differentially related to traditional intelligence, with Perception least related ($r = 0.16$), Management moderately related, and Understanding most related ($r = 0.40$). This is consistent with other findings that scales of nonverbal perception, such as the PONS (Profile of Nonverbal States; Rosenthal et al., 1979) which appear to be loaded on emotional perception, are relatively unrelated to intelligence, whereas problem solving of the sort covered on the Understanding branch plainly resembles traditional test items for intelligence more closely. Thus, skills representing emotional intelligence can be ordered along a continuum from those least to those most related to general intelligence. At the same time, the three tasks apparently share a common core of emotionally intelligent processing, as indicated by a more general, overall emotional intelligence factor.

As just noted, a single factor of emotional intelligence incorporates all the tasks studied here. This factor arises as a hierarchical factor obtained from factor-analyzing the three primary factor-based scales of Perception, Understanding, and Management (which are obtained by an oblimin rotation of the 12 tasks). Those three primary scales are fairly intercorrelated and factor-analyzing

them yields a single overall factor that summarizes performance across them all. This hierarchical factor is essentially identical to the first, unrotated principal factor of the 12 scales (their r = 0.94). The global factor indicates that it makes sense to talk about a single, unified emotional intelligence and a single emotional intelligence score. Such a score provides a reasonable first approximation of a person's ability level in the domain of emotional intelligence. As with any generalization, however, this overall score neglects variations in three subsidiary aspects of emotional intelligence, which can provide further clarification of any overall score.

Alternative Representations

It should be noted that the 12 tasks employed here do not exhaust the universe of emotionally intelligent abilities. As other tasks are developed, it is possible that more factors will be identified. One sort of task, in particular, that does not lend itself to group testing but that may form a separate factor, is ability at expressing emotion (Branch 1 skills). It may also be that a factor better encompassing assimilating emotions (Branch 2 of our model, which merged into Branch 3) might emerge as a more independent factor were it operationalized in tasks different than the ones used here.

Sex Differences

The identification of a new intelligence should increase the fairness of mental ability tests on average. That is because measures of the new intelligence help assess more of the total domain of intelligence, thereby giving any previously neglected capacities their fair consideration. Women and men appear to perform about the same on most intelligence-related mental tests, with most mean differences between 0.15 and 0.30 of an estimated population standard deviation (Hedges & Nowell, 1995). There are, however, some regular differences in the profiles of the two groups. Women are somewhat better on tests of reading comprehension, perceptual speed, associative memory, and composition. Men are somewhat better in mathematics, social studies, and scientific knowledge.

To the list of tasks at which women are somewhat better may be added emotional intelligence. Women performed about 0.5 standard deviation higher than men in the present study. The fact that women are slightly superior to men in perceiving emotion has been known for some time, through tests of nonverbal perception (that include emotion) such as the PONS (Rosenthal et. al., 1979), as well as through earlier-developed tests of emotional intelligence (Mayer & Geher, 1996). One possible explanation for this is that women must read emotions more carefully because they possess less power in society than do men (LaFrance & Hecht, 2001). It is women in more powerful positions rather than less, however, who exhibit the greater emotional accuracy (Hall & Halberstadt, 1994). Such findings suggest that emotional intelligence operates like other

areas of intelligence, potentially raising the occupational status of an individual. Issues of power and status aside, women may be socialized to pay more attention to emotions, or they may be better biologically prepared to perform at such tasks; our research does not address the relative contributions of the two (cf., LaFrance & Banaji, 1992).

Emotional Intelligence, Intelligence; and Empathy

The findings here also concern what emotional intelligence predicts. From the outset, emotional intelligence has been hypothesized to correlate with both intelligence and self-reported empathy (Mayer et al., 1990; Salovey & Mayer, 1990). Overall emotional intelligence, g_{ei}, correlated with verbal intelligence at a low-to-moderate level, as predicted. This replicates some of our earlier work as well (Mayer & Geher, 1996).

Overall emotional intelligence, gei, also correlates with self-reported empathy. This, too, replicates earlier studies (Mayer & Geher, 1996; Mayer et al., 1990). Emotional intelligence appears to correlate reliably with self-report empathy scales that share content overlap with the Epstein-Mehrabian scale (Mehrabian & Epstein, 1972). Such scales, including the one used here and the Davis (1983) empathy subscales of "empathic concern" and "emotion-related fantasy," involve a view of oneself as emotionally responsive and concerned about the feelings of others. The new scale employed here divides that same content domain into a variety of subfactors including Empathic Suffering, Positive Sharing, Responsive Crying, (reversed) Avoidance, and Feeling For Others. Emotional Intelligence correlated with each of these criteria in the expected direction.

Emotional Intelligence and Other Intelligences

The above demonstrations indicate that Emotional Intelligence, as measured by the MEIS, meets the most essential criteria for a standard intelligence. Our results illustrated that emotional intelligence does relate to general intelligence (via its proxy, verbal intelligence). The results, however, provide only the roughest idea of the relation between emotional intelligence and other intelligences. For example, traditional, academic intelligences can be divided into fluid and crystallized intelligences, or verbal and performance intelligences, or divided in many other ways (e.g., Carroll, 1993; Flanagan et al., 1997; Horn & Noll, 1994). Intelligence researchers will want to examine emotional intelligence and those various breakdowns in greater detail. The relation between emotional intelligence and other potentially similar intelligences such as social intelligence and personal intelligence, and the like, are similarly yet-to-be explored. As stated at the outset, any final choice between emotional intelligence and such alternatives as social intelligence, will depend upon the relative clarity of their operationalizations, their relative relations to general intelligence, and what criteria they predict. It is too early to make this comparison as of yet.

Only one of the competing intelligences (social intelligence) has been operationalized well enough (e.g., a minimum of three or four ability tasks) to compare to emotional intelligence. Other alternative intelligences, however, such as personal intelligence, could move in that direction in the future. As alternative intelligences become operationalized, it will be of interest to see how they compare. Finally, emotional creativity (Averill & Nunley, 1992) emphasizes generative, divergent thinking rather than the reasoning and problem solving of emotional intelligence. Emotional intelligence can be thought of as bearing the same relation to emotional creativity as general intelligence bears to general creativity. The intelligence-creativity relations are likely to be complex, but the retention of both concepts likely will be useful.

Emotional Intelligence and Other Criteria

A crucial job of the field is to relate internal characteristics of personality — including abilities such as emotional intelligence — to other psychological tests, and ultimately, to criteria in the life space (Mayer, 1998; Mayer et al., 1998). The test developed here has not yet been correlated with other personality scales such as the Big Five (McCrae & Costa; 1997), and that would be a desirable future direction. Instead, we moved modestly into comparing these internal abilities with actual life criteria: parental warmth, life satisfaction, psychotherapy, artistic ability, and leisure activities related to culture. The findings indicate that emotional intelligence is related to (self-reported) parental warmth and support, and, to a lesser extent, to life satisfaction. Emotional intelligence was also related to leisure pursuits including, negatively, to reading a large number of self-help books. Although these findings are preliminary, and better criteria are desirable, they are suggestive of the fact that emotional intelligence will be of use in predicting particular life criteria.

Future Research

We are presently at the beginning of the learning curve about emotional intelligence. Many questions remain unanswered. Some still concern the factorial structure of emotional intelligence: With the development of more tasks, will there be an additional factor of emotional expressiveness, or of assimilating emotion? Are there nonverbal tests that should be developed? Other questions concern the relation between emotional intelligence and other intelligences: How highly does emotional intelligence correlate with social intelligence, or with performance intelligence, or with spatial intelligence? More generally, how will it relate to the multitude of traditional cognitive abilities reviewed by Carroll (1993) and Horn & Noll (1994)?

Many of the questions of greatest interest to people, however, are those raised (as claims, rather than questions) by members of the press (e.g., Gibbs, 1995; Goleman, 1995). Specifically, these claims included that emotional intelli-

gence accounts in some large part for an individual's success, perhaps more so than conventional analytic intelligence (IQ). Despite the fact that certain among these claims appeared in reputable magazines and newspapers, there has been little or no direct evidence to support them (Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2001). Until the present article, in fact there has been no widespread systematic attempt to understand the measurement of emotional intelligence as an ability, although self-report mixed-model scales are proliferating (e.g., Bar-On, 1997). The present results indicate that emotional intelligence does play some role in everyday life. It is our hope that the field can move forward employing measures such as the MEIS. Measures such as the MEIS can provide serious answers to the questions above, as well as those that will arise in the future.

There are some matters that are clearly important about emotional intelligence already. Although emotions often have been regarded with respect in the West, there also exists a widespread negative view of people who think emotionally (Payne, 1986). Emotional thinkers have been referred to over the centuries variously as "overly emotional," romantics (or hopeless romantics), people who think with their hearts (instead of their heads), people swayed by emotions, or "biased" by emotions. Such labeling does accurately capture a kind of person who is overwrought with unthinking emotionality. What the existence of emotional intelligence tells us, however, is that there exists another type as well: the emotional, romantic, thinker-with-a-heart, who is engaged in sophisticated information processing, and who, in such a manner, contributes importantly to our lives and culture.

Conclusion

Measures of intelligence focused on verbal and performance intelligence have been developed over the century. Although verbal, performance, and other similar intelligences have taken us far (cf., Ree & Earles, 1992), there has also been a dissatisfaction with such limited conceptions of mental abilities. Over the century, many have sought out broader sets of mental capacities (e.g., Gardner, 1993; Guilford, 1967; Sternberg, 1988; Thorndike, 1920), or depicted a system of mental abilities (Detterman, 1986). Emotional intelligence represents, to us, an important candidate to enlarge the group on which general intelligence is based. Perhaps a general intelligence that includes emotional intelligence will be a more powerful predictor of important life outcomes than one that does not.

References

Abelson, R. P. (1963). Computer simulation of "hot cognitions." In S. Tomkins & S. Messick (Eds.), *Computer simulation of personality* (pp. 277-298). New York: Wiley.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. Upper Saddle River, NJ: Prentice-Hall.

Arbuckle, J. L. (1997). *Amos Users' Guide (Version 3.6)*. Chicago, IL: SmallWaters Corporation.

Averill, J. R., & Nunley, E. P. (1992). *Voyages of the heart: Living an emotionally creative life*. New York: Free Press.

Bandura, A. (1977). Self-efficacy: Toward a unified theory of behavioral change. *Psychological Review, 84* 191-215.

Bar-On, R., (1997). *Bar-On emotional quotient inventory: A measure of emotional intelligence*. Toronto, ON: Multi-Health Systems Inc.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*, 129-148.

Brown, B. (1997). Raw scores of cognitive ability are real psychological variables: IQ is a hyperspace variable. In V C. Shipman (Chair). *IQ or cognitive ability?* Symposium presented to the 9th Annual Convention of the American Psychological Society. Washington, DC.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Buck, R. (1984). *The communication of emotion*. New York: The Guilford Press.

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

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

Caruso, D. R., & Mayer, J. D. (1999). *A quick scale of empathy*. Manuscript in preparation.

Caruso, D. R., Van Buren, A., Mayer, J. D. & Salovey (1998). *An extension of the emotional intelligence test to an adolescent sample*. Manuscript in preparation.

Clynes, M. (1977). *Sentics: The touch of emotions*. Garden City, NY: Anchor Press/ Doubleday.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd edn.). New York: Harper and Row.

Darwin, C. (1965). *The expression of emotions in man and animals*. Chicago: University of Chicago Press. (Original work published 1872).

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989-1015.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.

de Rivera, J. (1977). *A structural theory of the emotions*. New York: International Universities Press.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 57-61). Norwood, NJ: Ablex.

Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy*. New York: W. W. Norton.

Flanagan, D. P., Genshaft, J. L., & Harrison, P. L. (1997). *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford Press.

Gardner, H. (1993). *Frames of mind. The theory of multiple intelligences*, (10th Anniversary edn.). New York: Basic Books.

Gibbs, N. (October 1995). The EQ factor. *Time*, 60-68.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Guttman L., & Levy, S. (1991). Two structural laws for intelligence. *Intelligence, 15*, 79-103.

Hall, J. A., & Halberstadt, A. G. (1994). "Subordination" and sensitivity to nonverbal cues: A study of married working women. *Sex Roles, 31*, 149-165.

Hedges L. V., & Nowell A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41-45.

Horn, J., & Noll, J. (1994). A system for understanding cognitive capacities: A theory and the evidence on which it is based. In D. K. Detterman (Ed.), *Current topics in human intelligence: Theories, tests, and issues* (Vol. 4). New York: Guilford Press.

Jensen, A. R., & Weng, L. (1994). What is good *g*? *Intelligence, 18*, 231-258.

LaFrance, M., & Banaji, M. (1992). Towards a reconsideration of the gender-emotion relationship. *Review of Personality and Social Psychology, 14*, 178-201.

LaFrance, M., & Hecht, M. A. (2001). Option or obligation to smile: The effects of power and gender on facial expression. In P. Philippot, R. Feldman, & E. Coates (Eds.), *The social context of non-verbal behavior*. Cambridge: Cambridge University Press.

Legree, P. L. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence, 21*, 247-266.

Mayer, J. D. (1998). A systems framework for the field of personality. *Psychological Inquiry, 9*, 118-144.

Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*, 89-113.

Mayer, J. D., & Mitchell, D. C. (1998). Intelligence as a subsystem of personality: From Spearman's *g* to contemporary models of hot-processing. In W. Tomic & J. Kingma (Eds.), *Advances in cognition and educational practice: Conceptual issues in research on intelligence* (Vol. 5, pp. 43-75). Greenwich, CT: JAI Press.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17*, (4), 433-442.

Mayer, J. D., & Salovey, P. (1995). Emotional intelligence and the construction and regulation of feelings. *Applied and Preventive Psychology, 4*, 197-208.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3-31) New York: Basic Books.

Mayer, J. D., & Stevens, A. (1994). An emerging understanding of the reflective (meta-) experience of mood. *Journal of Research in Personality, 28*, 351-373.

Mayer, J. D., DiPaolo, M. T., & Salovey P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment, 54*, 772-781.

Mayer, J. D., Salovey P., & Caruso, D. R. (1997). *Emotional intelligence test*. Needham, MA: Virtual Knowledge [Producer and Distributor] [CD-ROM].

Mayer, J. D., Carlsmith, K. M., & Chabot, H. F. (1998). Describing the person's external environment: Conceptualizing and measuring the life space. *Journal of Research in Personality, 32*, 253-296.

Mayer, J. D., Salovey P., & Caruso, D. R. (2001). Competing models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (2nd edn.). New York: Cambridge.

McCrae, R. R. & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509-516.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.

Morrison, D. F. (1976). *Multivariate statistical methods* (2nd edn.). New York: McGraw-Hill.

Neisser, U., Boodoo, G., Bouchard, T. J., 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.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-554.

Payne, W. L. (1986). A study of emotion: Developing emotional intelligence; self-integration; relating to fear, pain and desire. *Dissertation Abstracts International* (University Microfilms No. AAD9-5947), 47, (1-A), 203A.

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 197-219) Hillsdale, NJ: Erlbaum.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science,.1*, 86-89.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P., & Archer, D. (1979). *Sensitivity to nonverbal communication: A profile approach to the measurement of individual differences*. Baltimore: Johns Hopkins University Press.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*, 185-211.

Salovey, P., Mayer, J. D., Goldman, S., Turvey, C., & Palfai, T. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125-154) Washington, DC: American Psychological Association.

Scarr, S. (1989). Protecting general intelligence: Constructs and consequences for interventions. In R. L. Linn (Ed.), *Intelligence: Measurement, theory and public policy*. Urbana: University of Illinois Press.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.

Spinoza, B. (1984). "Ethics (Part III): On the origin and nature of emotions." Reprinted in C. Calhoun & R. C. Solomon (Eds.), *What is emotion?* Oxford: Oxford University Press.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Penguin Books.

Sternberg, R. J. (1994). Commentary: Reforming school reform: Comments on "Multiple intelligences: The theory in practice." *Teachers College Record, 95*, 561-569.

Sternberg, R. J., & Caruso, D. R. (1985). Practical modes of knowing. In E. Eisner (Ed.), *Learning and teaching the ways of knowing: 84th Yearbook of the National Society for the Study of Education* (Part II; pp. 133--158). Chicago: University of Chicago Press.

Sternberg, R. J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication.

Social Cognition, 3, 168-192.
Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine, 140,* 227-235.
Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.
Wagner, H. L., MacDonald, C. J., & Manstead, A. S. R. (1986). Communication of individual emotions by spontaneous facial expression. *Journal of Personality and Social Psychology, 50*, 737-743.
Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th edn.). Baltimore, MD: Williams and Wilkins.
Yerkes, R. M. (1921). Psychological examining in the United States. *Memoirs of the National Academy of Sciences* (15: 890 pp. entire).
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151-175.

The Mensa Awards for Excellence in Research

Mensa Awards for Excellence in Research are given each year to scientists who have published outstanding research papers in peer-reviewed professional journals about the nature of human intelligence or giftedness, education for the intellectually gifted, etc. This worldwide competition is sponsored jointly by the Mensa Education and Research Foundation and Mensa International, Ltd.

Typically, half the awards are given to established, senior scientists and half to researchers who have entered into research in the past five years. Eligible fields of research have included psychology, education, sociology, neurology, physiology, biochemistry, and psychometrics.

Each award consists of $500 and a certificate. Many of the winning articles are reprinted in the Mensa Research Journal.

Judging is done by the joint American Mensa, Ltd./Mensa Education and Research Foundation Research Review Committee.

For additional information about how to enter a paper into this competition, write to MERF, Awards for Excellence in Research, 1229 Corporate Drive West, Arlington, TX 76006, USA.

Mensa Education and Research Foundation Board of Trustees

Greg Timmers, President
Tom Ehrhorn, Vice President
Pat Merk, Treasurer
Ed Pirtle, Secretary
Jean K. Becker, Ph.D., AMC Chairman
Mary Dwyer Wolfe, Ph.D., National Scholarship Chair

TRUSTEES
Francis Cartier, Ph.D.
Dave Felt
Michael H. Jacobson, Ph.D., J.D.
Shirley Baker Loges, Ph.D.
Marie Mayer
Phyllis Miller
Dave Remine
Ember A Skidmore, Pharm.D.

MENSA
Research Journal

Mensa Education & Research Foundation
1229 Corporate Drive West
Arlington, TX 76006-6103

NON-PROFIT
ORGANIZATION
U.S. POSTAGE
PAID
Arlington, TX
Permit No. 922

100

**********3-DIGIT 222
G#100058523 G
ERIC CLEARINGHOUSE S5 P2
1110 N GLEBE RD STE 300
ARLINGTON VA 22201-4795

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").