NRC G/T

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED

University of Connecticut
University of Virginia
Yale University

Dynamic Instruction for and Assessment of Developing Expertise in Four Ethnic Groups

Robert J. Sternberg
Elena L. Grigorenko
Damian P. Birney
Nancy Fredine
Linda Jarvin
Ida Jeltova
Yale University
New Haven, Connecticut

May 2007
RM07226

Dynamic Instruction for and Assessment of Developing Expertise in Four Ethnic Groups

Robert J. Sternberg
Elena L. Grigorenko
Damian P. Birney
Nancy Fredine
Linda Jarvin
Ida Jeltova
Yale University
New Haven, Connecticut

May 2007
RM07226

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED

The National Research Center on the Gifted and Talented (NRC/GT) is funded under the Jacob K. Javits Gifted and Talented Students Education Act, Institute of Education Sciences, United States Department of Education.

The Directorate of the NRC/GT serves as an administrative and a research unit and is located at the University of Connecticut.

The participating universities include the University of Virginia and Yale University, as well as a research unit at the University of Connecticut.

Copies of this report are available from:
NRC/GT
University of Connecticut
2131 Hillside Road Unit 3007
Storrs, CT 06269-3007

Visit us on the web at:
www.gifted.uconn.edu

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R206R000001-05, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Institute of Education Sciences or the U.S. Department of Education.

Note to Readers...

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R206R000001-05, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Institute of Education Sciences or the U.S. Department of Education.

The National Research Center on the Gifted and Talented
University of Connecticut
2131 Hillside Road Unit 3007
Storrs, CT 06269-3007

Acknowledgments

The authors would like to thank the collaborations of the following individuals with this research project: Laima Broderick, Derrick Carpenter, Carol Gordon, Rachel Levine, and Heidi Soxman. A special thanks goes to Rachel Levine and other volunteer data collectors.

Dynamic Instruction for and Assessment of Developing Expertise in Four Ethnic Groups

Robert J. Sternberg
Elena L. Grigorenko
Damian P. Birney
Nancy Fredine
Linda Jarvin
Ida Jeltova
Yale University
New Haven, Connecticut

ABSTRACT

The goal of this research project was to investigate the use of dynamic assessment to increase equity, fairness, and accuracy in the testing of abilities and achievement. Dynamic tests have been found to reveal developing expertise in underrepresented minorities around the world that is not revealed by conventional static tests. The gender-balanced fourth grade participants were divided into three main groups: Experimental, Irrelevant Treatment Control and No Treatment Control group. We sampled students from four ethnic groups: European American, Asian American, African American, and Hispanic American. All students were given instruction and/or dynamic assessments (either individually or group administered) nurturing (instruction) and measuring (assessment) their developing expertise in mathematics. The data collected from participating students and teachers show that (a) it is possible to develop dynamic assessments that can be used to asses groups of and individual students in a regular classroom setting, (b) such dynamic assessments with a process oriented (rather than a filler) activity between post tests tends to lead to higher student achievement, and (c) dynamic instruction tends to reduce the achievement gap between minority and non-minority students.

Dynamic Instruction for and Assessment of Developing Expertise in Four Ethnic Groups

Robert J. Sternberg
Elena L. Grigorenko
Damian P. Birney
Nancy Fredine
Linda Jarvin
Ida Jeltova
Yale University
New Haven, Connecticut

EXECUTIVE SUMMARY

The goal of this research project was to investigate the use of dynamic assessment to increase equity, fairness, and accuracy in the testing of abilities and achievement. Dynamic tests have been found to reveal developing expertise in underrepresented minorities around the world that is not revealed by conventional static tests. In particular, it was proposed that use of dynamic tests will decrease or eliminate the differences typically obtained between ethnic groups on conventional static tests. Dynamic assessment was proposed as an alternative to static assessment because it provides measurement not only of *developed* skills, but also of *developing* skills.

The theoretical framework underlying this research project is one of abilities, like achievement, as aspects of developing expertise. In other words, abilities and achievements cannot be measured as qualitatively distinct entities, but rather are measured by similar kinds of tests assessing developing expertise at different levels.

Initially, it was planned that the 1500 students would be recruited would be divided equally between 2 experimental and 2 control groups. However, due to a revision of the design to include 7 distinct conditions, the students were distributed somewhat differently. The gender-balanced fourth grade participants were divided into 3 main groups: approximately 450 in the experimental group (with 3 subgroups), 600 in the Irrelevant Treatment Control group (with 2 subgroups), and 450 in the No Treatment Control group (with 2 subgroups). We sampled students from 4 ethnic groups: European American, Asian American, African American, and Hispanic American. Schools in the districts of Danbury, CT, Hamden, CT, New London, CT, Stamford, CT, Vernon, CT, and New York, NY were enrolled in the study. The schools were assigned to 1 of the 7 conditions within the experimental and control groups: 2 groups with dynamic instruction and group-administered dynamic assessments, 1 group with dynamic instruction and individually administered dynamic assessment, 2 groups with triarchic instruction and group-administered dynamic assessments, and 2 groups with standard instruction and group-administered dynamic assessments.

As a function of the geographical location of the study, there were a majority of European American students (n=687), and the breakdown was fairly equivalent among the other 3 ethnicities, with approximately 300 students in each group. All students were given instruction and/or dynamic assessments (either individually or group-administered) nurturing (instruction) and measuring (assessment) their developing expertise in mathematics. Control participants were divided into no-treatment and irrelevant-treatment instructional groups.

Data were analyzed so that learning gains in the 4 groups in the 7 conditions were compared. The main hypothesis was that, whereas learning gains in the experimental conditions would exceed those in the control conditions across the 4 ethnic groups, the difference would be especially pronounced in the ethnic minority groups. Thus, it was hypothesized that dynamic tests would reduce or eliminate differences among groups, while at the same time providing more equitable, fair, and comprehensive assessments of skills.

The data collected from participating students and teachers show that (a) it is possible to develop dynamic assessments that can be used to asses groups of and individual students in a regular classroom setting, (b) such dynamic assessments with a process oriented (rather than a filler) activity between posttests tends to lead to higher student achievement, and (c) dynamic instruction tends to reduce the achievement gap between minority and non-minority students.

Table of Contents

Table of Contents (continued)

List of Tables

List of Figures

Dynamic Instruction for and Assessment of Developing Expertise in Four Ethnic Groups

Robert J. Sternberg
Elena L. Grigorenko
Damian P. Birney
Nancy Fredine
Linda Jarvin
Ida Jeltova
Yale University
New Haven, Connecticut

Background and Theory

Rationale

Many members of underrepresented minority groups tend to show reduced school learning, lower ability test scores, and lower achievement test scores than do members of other groups. The fact that both their ability test scores and their achievement test scores are lower is taken to indicate the validity of the ability tests as predictors of school achievement, and of the achievement test scores as criteria against which to evaluate ability tests. The contention of this research project is that this reasoning is incorrect, simply because ability tests and achievement tests largely measure the same constructs, and that those constructs inadequately represent both abilities and achievement. Even the way instruction is done in schools often draws upon the same narrow range of abilities. Put another way, educational leaders in the United States have bought into a system that leads to mistaken conclusions, but worse, leads to inadequate education for members of underrepresented minority groups as well as many others. The research project suggests a different way in which ability testing, instruction, and achievement testing can be done that may more adequately and equitably represents students' abilities and achievements.

Theory: Abilities and Achievement as Developing Expertise

The conventional view of abilities is that they represent relatively stable attributes of individuals that develop as an interaction between heredity and environment. Factor analysis and related techniques then can be used on psychometric tests of intelligence to determine the structure of intellectual abilities, as illustrated by the massive analysis by Carroll (1993).

The argument of this research project, advancing that of Sternberg (1998, 1999), is that this view of what abilities are and of what ability tests measure may be incorrect. An alternative view is that of abilities and achievement both as forms of developing expertise. In this view, ability tests, like achievement tests, measure an aspect—typically a limited aspect—of developing expertise. Developing expertise is defined here as the ongoing process of the acquisition and consolidation of a set of skills needed for a high level of mastery in one or more domains of life performance. Good performance on ability tests requires a certain kind of expertise, and to the extent this expertise overlaps with the expertise required for learning and performance in schooling or in the workplace, there will be a correlation between the tests and performance in school or in the workplace. But such correlations represent no intrinsic relation between abilities and other kinds of performance, but instead, overlaps in the kinds of expertise needed to perform well under somewhat different kinds of circumstances.

There is nothing privileged about ability tests. One could as easily use, say, academic achievement to predict intelligence-related scores. For example, it is as simple to use the SAT-II (a measure of achievement) to predict the SAT-I (a measure formerly called the Scholastic Assessment Test and before that called the Scholastic Aptitude Test) as vice versa, and of course, the levels of prediction will be the same. Both tests measure achievement, although the kinds of achievements they measure are different.

According to this view, although ability tests may have temporal priority relative to various criteria in their administration (i.e., ability tests are usually administered first, and later, criterion indices of performance, such as grades or achievement test scores, are collected), they have no psychological priority. All of the various kinds of assessments are of the same kind psychologically. What distinguishes ability tests from other kinds of assessments is how the ability tests are used (usually predictively) rather than what they measure. There is no qualitative distinction among the various kinds of assessments. All tests measure various kinds of developing expertise.

Conventional tests of intelligence and related abilities measure achievement that individuals should have accomplished several years back (see Anastasi & Urbina, 1997). Tests such as vocabulary, reading comprehension, verbal analogies, arithmetic problem solving, and the like, are all, in part, tests of achievement. Even abstract-reasoning tests measure achievement in dealing with geometric symbols, skills taught in Western schools (Laboratory of Comparative Human Cognition, 1982). One might as well use academic performance to predict ability-test scores. The problem regarding the traditional model is not in its statement of a correlation between ability tests and other forms of achievement but in its proposal of a causal relation whereby the tests reflect a construct that is somehow causal of, rather than merely temporally antecedent to, later success. The developing-expertise view in no way rules out the contribution of genetic factors as sources of individual differences in who will be able to develop a given amount of expertise. Many human attributes, including intelligence, reflect the covariation and interaction of genetic and environmental factors. But the contribution of genes to an individual's intelligence cannot be directly measured or even directly estimated. Rather, what is measured is a portion of what is expressed, namely, manifestations of developing expertise, the kind of expertise that potentially leads to reflective practitioners in a variety of fields (Schön, 1983). This approach to measurement has been used explicitly by Royer, Carlo, Dufresne, and Mestre (1996), who have shown that it is possible to develop measurements of reading skill reflecting varying levels of developing expertise. In such assessments, outcome measures reflect not simply quantitative assessments of skill, but qualitative differences in the types of developing expertise that have emerged (e.g., ability to understand technical text material, ability to draw inferences from this material, or ability to draw about "big ideas" in technical text).

The same arguments that apply to assessment apply to instruction as well. Present instruction always builds upon an already existing knowledge base. The extent to which learning takes place in a given instructional setting will thus depend on the quality of the teaching and the student's learning skills, of course. But it also will depend upon the extent to which the instruction is properly scaffolded, given the student's prior developmental level and current zone in which learning maximally can take place. Construction of such scaffolding is a process that requires a skilled teacher, much as the building of a scaffold for a house requires a skilled carpenter. Easier than building appropriate scaffolding is to take one of two routes that essentially enable the teacher to opt out. One is to expose the child to weak, watered-down instruction that will be easy enough for almost any student, but that will provide little in the way of enrichment. Such instruction tends to build cumulative deficits, so that children who start off behind in their work become successively further behind. This sort of instruction is commonly given to children in resource-poor environments who are labeled as having learning disabilities (Sternberg & Grigorenko, 1999), but it also is given to other children who are not viewed as performing well. The second opt-out route is to expose children to instruction that is over their heads, and thus for which they are not ready. Many people, even of the middle class, feel they have experienced such instruction, often in mathematics. The result is that the child gets further and further behind simply because the instruction is too much ahead of his or her developmental level.

According to this view, measures of abilities *should* be correlated with later success, because both measures of abilities and various measures of success require developing expertise of related types. For example, both typically require what are sometimes referred to as *metacomponents* of thinking: recognition of problems, definition of problems, formulation of strategies to solve problems, representation of information, allocation of resources, and monitoring and evaluation of problem solutions (Sternberg, 1985). These skills develop as results of gene-environment covariation and interaction. If we wish to call them *intelligence* that is certainly fine, so long as we recognize that what we are calling intelligence is a form of developing expertise.

A major goal of work under the point of view presented here is to integrate the study of intelligence and related abilities (see reviews in Cianciolo & Sternberg, 2004, Sternberg, 1990, 1994a, 2000) with the study of expertise (Chi, Glaser, & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991; Hoffman, 1992). These literatures, typically viewed as distinct, are here viewed as ultimately involved with the same psychological mechanisms.

The Specifics of the Developing-Expertise Model

The specifics of the developing-expertise model are described below. At the heart of the model is the notion of *developing expertise*—which individuals are constantly in a process of developing expertise when they work within a given domain. They may and do, of course, differ in rate and asymptote of development. The main constraint in achieving expertise is not some fixed prior level of capacity, but purposeful engagement involving direct instruction, active participation, role modeling, and reward. Instruction and assessment should take into account, in the ideal, all of the elements of the model.

Elements of the Model

The model of developing expertise has five key elements (although certainly they do not constitute an exhaustive list of elements in the development of expertise): metacognitive skills, learning skills, thinking skills, knowledge, and motivation. Although it is convenient to separate these five elements, they are fully interactive. They influence each other, both directly and indirectly. For example, learning leads to knowledge, but knowledge facilitates further learning.

These elements are, to some extent, domain specific. The development of expertise in one area does not necessarily lead to the development of expertise in another area, although there may be some transfer, depending upon the relationship of the areas, a point that has been made with regard to intelligence by others as well (e.g., Gardner, 1983, 1999).

In the theory of successful intelligence (Sternberg, 1985, 1997, 1999, 2005), intelligence is viewed as having three aspects: analytical, creative, and practical. Our research suggests that the development of expertise in one creative domain (Sternberg & Lubart, 1995) or in one practical domain (Sternberg, Wagner, Williams, & Horvath, 1995) shows modest correlations with the development of expertise in other such domains. Psychometric research suggests more domain generality for the analytical domain (Jensen, 1998, see essays in Sternberg & Grigorenko, 2002a). Moreover, people can show analytical, creative, or practical expertise in one domain without showing all three of these kinds of expertise, or even two of the three.

1. *Metacognitive skills.* Metacognitive skills (or metacomponents—Sternberg, 1985) refer to people's understanding and control of their own cognition. For example, such skills would encompass what an individual knows about writing papers or solving arithmetic word problems, both with regard to the steps that are involved and with regard to how these steps can be executed effectively. Seven metacognitive skills are particularly important: problem recognition, problem definition, problem representation, strategy formulation, resource allocation, monitoring of problem solving, and evaluation of problem solving (Sternberg, 1985, 1986). All of these skills are modifiable (Sternberg, 1986, 1988, 2003; Sternberg & Grigorenko, 2000; Sternberg & Spear-Swerling, 1996).

2. *Learning skills.* Learning skills (knowledge-acquisition components) are essential to the model (Sternberg, 1985, 1986), although they are certainly not the only learning skills that individuals use. Learning skills are sometimes divided into explicit and implicit ones. Explicit learning is what occurs when we make an effort to learn; implicit learning is what occurs when we pick up information incidentally, without any systematic effort. Examples of learning skills are selective encoding, which involves

distinguishing relevant from irrelevant information; selective combination, which involves putting together the relevant information; and selective comparison, which involves relating new information to information already stored in memory (Sternberg, 1985).

3. *Thinking skills*. There are three main kinds of thinking skills (or performance components) that individuals need to master (Sternberg, 1985, 1986, 1994b). It is important to note that these are sets of, rather than individual, thinking skills. Critical (analytical) thinking skills include analyzing, critiquing, judging, evaluating, comparing and contrasting, and assessing. Creative thinking skills include creating, discovering, inventing, imagining, supposing, and hypothesizing. Practical thinking skills include applying, using, utilizing, and practicing (Sternberg, 1997; Sternberg & Grigorenko, 2003). They are the first step in the translation of thought into real-world action.

4. *Knowledge*. There are two main kinds of knowledge that are relevant in academic situations. Declarative knowledge is of facts, concepts, principles, laws, and the like. It is "knowing that." Procedural knowledge is of procedures and strategies. It is "knowing how." Of particular importance is procedural tacit knowledge, which involves knowing how the system functions in which one is operating (Sternberg et al., 2000; Sternberg & Horvath, 1999; Sternberg, Wagner et al., 1995).

5. *Motivation.* One can distinguish among several different kinds of motivation. A first kind of motivation is achievement motivation (McClelland, 1985; McClelland, Atkinson, Clark, & Lowell, 1976). People who are high in achievement motivation seek moderate challenges and risks. They are attracted to tasks that are neither very easy nor very hard. They are strivers—constantly trying to better themselves and their accomplishments. A second kind of motivation is competence (self-efficacy) motivation, which refers to persons' beliefs in their own ability to solve the problem at hand (Bandura, 1977, 1996). Experts need to develop a sense of their own efficacy to solve difficult tasks in their domain of expertise. This kind of self-efficacy can result both from intrinsic and extrinsic rewards (Amabile, 1996; Sternberg & Lubart, 1996). Of course, other kinds of motivation are important too. Indeed, motivation is perhaps the indispensable element needed for school success. Without it, the student never even tries to learn.

6. *Context*. All of the elements discussed above are characteristics of the learner. Returning to the issues raised at the beginning of this monograph, a problem with conventional tests is that they assume that individuals operate in a more or less decontextualized environment. A test score is interpreted largely in terms of the individual's internal attributes. But a test measures much more, and the assumption of a fixed or uniform context across test-takers is not realistic. Contextual factors that can affect test performance include native language, emphasis of test on speedy performance, importance to the test taker of success on the test, and familiarity with the kinds of material on the test.

Interactions of Elements

The novice works toward expertise through deliberate practice. But this practice requires an interaction of all 5 of the key elements. At the center, driving the elements of the model, is motivation. Without it, the elements remain inert. Eventually, one reaches a kind of expertise, at which one becomes a reflective practitioner of a certain set of skills. But expertise occurs at many levels. The expert first-year graduate or law student, for example, is still a far cry from the expert professional. People thus cycle through many times, on the way to successively higher levels of expertise.

Motivation drives metacognitive skills, which in turn activate learning and thinking skills, which then provide feedback to the metacognitive skills, enabling one's level of expertise to increase (see Sternberg, 1985). The declarative and procedural knowledge acquired through the extension of the thinking and learning skills also results in these skills being used more effectively in the future.

All of these processes are affected by, and can in turn affect, the context in which they operate. For example, if a learning experience is in English but the learner has only limited English proficiency, his or her learning will be inferior to that of someone with more advanced English-language skills. Or if material is presented orally to someone who is a better visual learner, that individual's performance will be reduced.

How does this model of developing expertise relate to the construct of intelligence?

The g-Factor and the Structure of Abilities

Some intelligence theorists point to the stability of the alleged general factor of human intelligence as evidence for the existence of some kind of stable and overriding structure of human intelligence (see essays in Sternberg & Grigorenko, 2002a). But the existence of a *g* factor may reflect little more than an interaction between whatever latent (and not directly measurable) abilities individuals may have and the kinds of expertise that are developed in school. With different forms of schooling, *g* could be made either stronger or weaker. In effect, Western forms and related forms of schooling may, in part, create the *g* phenomenon by providing a kind of schooling that teaches in conjunction the various kinds of skills measured by tests of intellectual abilities.

Suppose, for example, that children were selected from an early age to be schooled for a certain trade. Throughout most of human history, this is in fact the way most children were schooled. Boys, at least, were apprenticed at an early age to a master who would teach them a trade. There was no point in their learning skills that would be irrelevant to their lives.

To bring the example into the present, imagine that we decided, from an early age, that certain students would study English (or some other native language) to develop language expertise; other students would study mathematics to develop their mathematical expertise. Still other students might specialize in developing spatial expertise to be used in flying airplanes or doing shop work or whatever. Instead of specialization beginning at the university level, it would begin from the age of first schooling.

This point of view is related to, but different from, that typically associated with the theory of crystallized and fluid intelligence (Cattell, 1971; Horn, 1994). In that theory, fluid ability is viewed as an ability to acquire and reason with information whereas crystallized ability is viewed as the information so acquired. According to this view, schooling primarily develops crystallized ability, based in part upon the fluid ability the individual brings to bear upon school-like tasks. In the theory proposed here, however, both fluid and crystallized ability are roughly equally susceptible to development through schooling or other means societies create for developing expertise. One could argue that the greater validity of the position presented here is shown by the near-ubiquitous Flynn effect (Flynn, 1987, 1998; Neisser, 1998), which documents massive gains in IQ around the world throughout most of the 20th century. The effect must be due to environment, because large genetic changes worldwide in such a short time frame are virtually impossible. Interestingly, gains are substantially larger in fluid abilities than in crystallized abilities, suggesting that fluid abilities are likely to be as susceptible as or probably more susceptible than crystallized abilities to environmental influences. Clearly, the notion of fluid abilities as some basic genetic potential one brings into the world, whose development is expressed in crystallized abilities, does not work.

These students then would be given an omnibus test of intelligence or any broad-ranging measure of intelligence. There would be no general factor because people schooled in one form of expertise would not have been schooled in others. One can imagine even negative correlations between subscores on the so-called intelligence test. The reason for the negative correlations would be that developing expertise in one area might preclude developing expertise in another because of the form of schooling.

Lest this tale sound far-fetched, we hasten to add that it is a true tale of what is happening now in some places. In the United States and most of the developed world, of course, schooling takes a fairly standard course. But this standard course and the value placed upon it are not uniform across the world. And we should not fall into the ethnocentric trap of believing that the way Western schooling works is the way all schooling should work (e.g., Serpell, 1993).

In a collaborative study among children near Kisumu, Kenya, (Sternberg, Nokes, et al., 2001; see also Sternberg & Grigorenko, 1997), we devised a test of practical intelligence that measures informal knowledge for an important aspect of adaptation to the environment in rural Kenya, namely, knowledge of the identities and use of natural herbal medicines that could be used to combat illnesses. The children use this informal knowledge an average of once a week in treating themselves or suggesting treatments to other children, so this knowledge is a routine part of their everyday existence. By *informal knowledge*, we are referring to kinds of knowledge not taught in schools and not assessed on tests given in the schools. It is essentially the same as tacit knowledge.

The idea of our research was that children who knew what these medicines were, what they were used for, and how they should be dosed would be in a position better to adapt to their environments than would children without this informal knowledge. We do not know how many, if any, of these medicines actually work, but from the standpoint of measuring practical intelligence in a given culture, the important thing is that the people in Kenya believe that the medicines work. For that matter, it is not always clear how effective are the medicines used in the Western world.

We found substantial individual differences in the tacit knowledge of like-aged and schooled children about these natural herbal medicines. More important, however, was the correlation between scores on this test and scores on an English-language vocabulary test (the Mill Hill), a Dholuo equivalent (Dholuo is the community and home language), and the Raven Coloured Progressive Matrices. We found significantly *negative* correlations between our test and the English-language vocabulary test. Correlations of our test with the other tests were trivial. The better children did on the test of indigenous tacit knowledge, the worse they did on the test of vocabulary used in school, and vice versa. Why might we have obtained such a finding?

Based on ethnographic observation, we believe a possible reason is that parents in the village may emphasize either a more indigenous or a more Western education. Some parents (and their children) see little value to school. They do not see how success in school connects with the future of children who will spend their whole lives in a village, where they do not believe they need the expertise the school teaches. Other parents and children seem to see Western schooling as of value in itself or potentially as a ticket out of the confines of the village. The parents thus tend to emphasize one type of education or the other for their children, with corresponding results. The kinds of developing expertise the families value differ, and so therefore do scores on the tests. From this point of view, the intercorrelational structure of tests tells us nothing intrinsic about the structure of intelligence per se, but rather, something about the way abilities as developing forms of expertise structure themselves in interaction with the demands of the environment.

Nuñes (1994) has reported related findings based on a series of studies she conducted in Brazil (see also Ceci & Roazzi, 1994). Street children's adaptive intelligence is tested to the limit by their ability to form and successfully run a street business. If they fail to run such a business successfully, they risk either starvation or death at the hands of death squads should they resort to stealing. Nuñes and her collaborators have found that the same children who are doing the mathematics needed for running a successful street business cannot well do the same types of mathematics problems presented in an abstract, paper-and-pencil format.

From a conventional-abilities standpoint, this result is puzzling. From a developing-expertise standpoint, it is not. Street children grow up in an environment that fosters the development of practical but not academic mathematical skills. We know that even conventional academic kinds of expertise often fail to show transfer (e.g., Gick & Holyoak, 1980). It is scarcely surprising, then, that there would be little transfer here. The street children have developed the kinds of practical arithmetical expertise they need for survival and even success, but they will get no credit for these skills when they take a conventional abilities test.

It also seems likely that if the scales were reversed, and privileged children who do well on conventional ability tests or in school were forced out on the street, many of them would not survive long. Indeed, in the ghettoes of urban America, many children and adults who, for one reason or another end up on the street, in fact barely survive or do not make it at all.

Jean Lave (1989) has reported similar findings with Berkeley housewives shopping in supermarkets. There just is no correlation between their ability to do the mathematics needed for comparison-shopping and their scores on conventional paper-and-pencil tests of comparable mathematical skills. And Ceci and Liker (1986) found, similarly, that expert handicappers at racetracks generally had only average IQs. There was no correlation between the complexity of the mathematical model they used in handicapping and their scores on conventional tests. In each case, important kinds of developing expertise for life were not adequately reflected by the kinds of developing expertise measured by the conventional ability tests.

One could argue that these results merely reflect the fact that the problem that these studies raise is not with conventional theories of abilities, but with the tests that are loosely based on these theories: These tests do not measure street math, but more abstracted forms of mathematical thinking. But psychometric theories, we would argue, deal with a similarly abstracted general factor. The abstracted tests follow largely from the abstracted theoretical constructs. In fact, our research has shown that tests of practical intelligence correlate minimally, if at all, with scores on these abstracted tests (e.g., Sternberg et al., 2000; Sternberg & The Rainbow Project Collaborators, 2006; Sternberg, Wagner et al., 1995).

The problem with the conventional model of abilities does not just apply in what to us are exotic cultures or exotic occupations. In a collaborative study with Michel Ferrari, Pamela Clinkenbeard, and Elena Grigorenko (Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999), high school students were tested for their analytical, creative, and practical abilities via multiple-choice and essay items. The multiple-choice items were divided into 3 content domains: verbal, quantitative, and figural pictures. Students' scores were factor analyzed and then later correlated with their performance in a college-level introductory-psychology course.

We found that when students were tested not only for analytical abilities, but for creative and practical abilities too (as follows from the model of successful intelligence, Sternberg, 1985, 1997), the strong general factor that tends to result from multiple-ability tests becomes much weaker. Of course, there is always some general factor when one factor analyzes but does not rotate the factor solution, but the general factor was weak, and of course disappeared with a varimax rotation. We also found that all of analytical, creative, and practical abilities predicted performance in the introductory-psychology course (which itself was taught analytically, creatively, or practically, with assessments to match). Moreover, although the students who were identified as high analytical were the traditional population—primarily White, middle- to upper middle-class, and well educated, the students who were identified as high creative or high practical were much more diverse in all of these attributes. Most importantly, students whose instruction better matched their triarchic pattern of abilities outperformed those students whose instruction more poorly matched their triarchic pattern of abilities.

In a more recent study with high school and college students (Sternberg & The Rainbow Project Collaborators, 2006), we found that when tests of creative and practical abilities supplement a test of analytical abilities, creative and practical factors are found in addition to an omnibus multiple-choice factor, which appears to be similar to what is usually extracted as "*g*." We also found that these supplemental tests roughly doubled prediction of college freshman grade-point average, and substantially reduced ethnic-group differences in comparison with the analytical tests.

Thus, conventional tests may unduly favor a small segment of the population by virtue of the narrow kind of developing expertise they measure. When one measures a broader range of developing expertise, the results look quite different (Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001). Moreover, the broader range of expertise includes kinds of skills that will be important in the world of work and in the world of the family.

Analytical, creative, and practical abilities, as measured by our tests or anyone else's, are simply forms of developing expertise. All are useful in various kinds of life tasks. But conventional tests may unfairly disadvantage those students who do not do well in a fairly narrow range of kinds of expertise. By expanding the range of developing expertise we measure, we discover that many children not now identified as able have, in fact, developed important kinds of expertise. The abilities conventional tests are important for school and life performance, but they are not the only abilities that matter.

Teaching in a way that departs from notions of abilities based on a general factor also pays dividends. In a recent set of studies, we have shown that generally lower socioeconomic class third grade and generally middle-class eighth grade students who are taught social studies (a unit in communities) or science (a unit on psychology) for successful intelligence (analytically, creative, and practically, as well as for memory) outperform students who are taught just for analytical (critical) thinking or just for memory (Sternberg, Torff, & Grigorenko, 1998a, 1998b). The students taught "triarchically" outperform the other students not only on performance assessments that look at analytical, creative, and practical kinds of achievements, but even on tests that measure straight memory (multiple-choice tests already being used in the courses). None of this is to say that analytical abilities are not important in school and life—obviously, they are. Rather, what our data suggest is that other types of abilities—creative and practical ones—are important as well and that students need to learn how to use all 3 kinds of abilities together.

Thus, teaching students in a way that takes into account their more highly developed expertise and that also enables them to develop other kinds of expertise results in superior learning outcomes, regardless of how these learning outcomes are measured. The children taught in a way that enables them to use kinds of expertise other than memory actually remember better, on average, than do children taught for memory.

We have also done studies in which we have measured informal procedural knowledge in children and adults. We have done such studies with business managers, college professors, elementary school students, sales people, college students, and general populations. This important aspect of practical intelligence, in study after study, has been found to be uncorrelated with academic intelligence as measured by conventional tests, in a variety of populations, occupations, and at a variety of age levels (Hedlund et al., 2003; Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006; Sternberg et al., 2000; Sternberg & Hedlund, 2002; Sternberg, Wagner et al., 1995). Moreover, the tests predict job performance as well as or better than do tests of IQ. The lack of correlation of the two kinds of ability tests suggests that the best prediction of job performance will result when both academic and practical intelligence tests are used as predictors. Most recently, we have developed a test of common sense for the work place—for example, how to handle oneself in a job interview—that predicts self-ratings of common sense but not self-ratings of various kinds of academic abilities (Sternberg et al., 2000).

It is important to note that practical, informal procedural knowledge can not only be assessed in children, but also, taught as well (Sternberg, Okagaki, & Jackson, 1990; Williams et al., 2002). We devised a program, *Practical Intelligence for School,* in which children were taught homework, test taking, reading, and writing skills of the kinds that typically are not explicitly taught in schools but that students somehow are expected to learn. Examples would be knowing how to read materials of different levels of difficulty in different ways, or to read such materials for multiple-choice versus essay tests. We found that it is possible to teach these skills to improve students' achievement (Williams et al., 2002).

Although the kinds of informal procedural expertise we measure in these tests do not correlate with academic expertise, they do correlate across work domains. For example, we found that subscores (for managing oneself, managing others, and managing tasks) on measures of informal procedural knowledge are correlated with each other and that scores on the test for academic psychology are moderately correlated with scores on the test for business managers (Sternberg et al., 2000; Sternberg, Wagner et al., 1995). So the kinds of developing expertise that matter in the world of work may show certain correlations with each other that are not shown with the kinds of developing expertise that matter in the world of the school.

It is even possible to use these kinds of tests to predict effectiveness in leadership. Studies of military leaders showed that tests of informal knowledge for military leaders predicted the effectiveness of these leaders, whereas conventional tests of intelligence did not. We also found that although the test for managers was significantly correlated with the test for military leaders, only the latter test predicted superiors' ratings of leadership effectiveness (Hedlund et al., 2003; Sternberg et al., 2000).

Both conventional academic tests and our tests of practical intelligence measure forms of developing expertise that matter in school and on the job. The 2 kinds of tests are not qualitatively distinct in that they measure "formed," developed knowledge and skills. The reason the correlations are essentially null is that the kinds of developing expertise they measure are quite different. The people who are good at abstract, academic kinds of expertise are often people who have not emphasized learning practical, everyday kinds of expertise, and vice versa, as we found in our Kenya study. Indeed, children who grow up in challenging environments such as the inner city may need to develop practical over academic expertise as a matter of survival. As in Kenya, this practical expertise may better predict their survival than do academic kinds of expertise. The same applies in business, where tacit knowledge about how to perform on the job is as likely or more likely to lead to job success than is the academic expertise that in school seems so important.

The practical kinds of expertise matter in school too. In a study at Yale, Wendy Williams and Robert Sternberg (cited in Sternberg, Wagner, & Okagaki, 1993) found that a test of tacit knowledge for college predicted grade-point average as well as did an academic-ability test. But a test of tacit knowledge for college life better predicted adjustment to the college environment than did the academic test.

Taking Tests

Developing expertise applies not only to the constructs measured by conventional intelligence tests, but also to the very act of taking the tests.

Sometimes the expertise children learn that is relevant for in-school tests may actually hurt them on conventional ability tests. In one example, we studied the development of children's analogical reasoning in a country day school where teachers taught in English in the morning and in Hebrew in the afternoon (Sternberg & Rifkin, 1979). We found a number of second grade students who got no problems right on our test. They would have seemed, on the surface, to be rather stupid. We discovered the reason why, however. We had tested in the afternoon, and in the afternoon, the children always read in Hebrew. So they read our problems from right to left, and got them all wrong. The expertise that served them so well in their normal environment utterly failed them on the test.

Our sample was of upper middle-class children who, in a year or two, would know better. But imagine what happens with other children in less supportive environments who develop kinds of expertise that may serve them well in their family or community lives or even school life, but not on the tests. They will appear to be stupid rather than lacking the kinds of expertise the tests measure.

Patricia Greenfield (1997) has done a number of studies in a variety of cultures and found that the kinds of test-taking expertise assumed to be universal in the U.S. and other Western countries are by no means universal. She found, for example, that children in Mayan cultures (and probably in other highly collectivist cultures as well) were puzzled when they were not allowed to collaborate with parents or others on test questions. In the U.S., of course, such collaboration would be viewed as cheating. But in a collectivist culture, someone who had not developed this kind of collaborative expertise, and moreover, someone who did not use it, would be perceived as lacking important adaptive skills (see also Laboratory of Comparative Human Cognition, 1982).

Renzulli, Reis, Hébert, and Diaz (1995) researched the performance of African American students compared to their majority counterparts. They found that a number of factors contribute to their relatively low group performance on academic achievement tests, including: reduced opportunities to acquire academic skills, limited parental support and expectancies for educational attainment, and disengagement from or distrust of majority-cultural values for education. They suggested that this performance differential would decrease with more equivalent learning opportunities or with an alternative form of assessment.

Delpit (1995), Gordon and Yowell (1994), and Taylor (1991) hypothesized that academic risk is associated with the potential discontinuity or "lack of fit" between the behavioral patterns and values socialized in the context of low income and minority families and communities, and those expected in the mainstream classroom and school context. Borman and Overman (2004) found that in their sample of participants of African American, Latino, and White students from relatively homogeneous low-SES backgrounds, minority students have lower academic self-efficacy and are exposed to school environments that are less conducive to academic resilience. These differences between minority and White children, and differences in their schools, could in part explain the frequently noted achievement gaps that separate minority and majority students.

Fagundes, Haynes, Haak, and Moran (1998) investigated claims that standardized testing presents a significant threat to the fair assessment of children from diverse language groups (Kamhi, Pollock, & Harris, 1996; Taylor & Payne, 1983; Vaughn-Cooke, 1986). They found that typical types of bias on standardized tests that can have a negative effect on culturally diverse children are: situational bias (examination format is threatening to student), direction bias (directions for test can be misinterpreted by student), value bias (asking student to give moral/ethical judgments that may differ culturally from the examiner's), linguistic bias, format bias (test procedures are inconsistent with student's cognitive style), cultural misinterpretation (negative interpretation of student's behavior when it is culturally appropriate), and stimulus bias (test is highly object-/picture oriented when child is socially oriented).

In Sum: The Need for a View of Abilities as Developing Expertise

Thus, we have argued in this section that ability tests, like achievement tests, measure developing expertise. Tests can be created that favor the kinds of developing expertise formed in any kind of cultural or subcultural milieu. Those who have created conventional tests of abilities have tended to value the kinds of skills most valued by Western schools. This system of valuing is understandable, given that Binet and Simon (1905) first developed intelligence tests for the purpose of predicting school performance. Moreover, these skills are important in school and in life. But in the modern world, the conception of abilities as fixed or even as predetermined is an anachronism. Moreover, our research and that of others (reviewed more extensively in Sternberg, 1997) shows that the set of abilities assessed by conventional tests measures only a small portion of the kinds of developing expertise relevant for life success. It is for this reason that conventional tests predict only about 10% of individual-difference variation in various measures of success in adult life (Herrnstein & Murray, 1994).

Not all cultures value equally the kinds of expertise measured by these tests. In a study comparing Latino, Asian, and Anglo subcultures in California, for example, we found that Latino parents valued social kinds of expertise as more important to intelligence than did Asian and Anglo parents, who more valued cognitive kinds of expertise (Okagaki & Sternberg, 1993). Predictably, teachers also more valued cognitive kinds of expertise, with the result that the Anglo and Asian children would be expected to do better in school, and did. Of course, cognitive expertise matters in school and in life, but so does social expertise. Both need to be taught in the school and the home to all children. This latter kind of expertise may become even more important in the work place. Until we expand our notions of abilities, and recognize that when we measure them, we are measuring developing forms of expertise, we will risk consigning many potentially excellent contributors to our society to bleak futures. We will also be potentially overvaluing students with expertise for success in a certain kind of schooling, but not necessarily with equal expertise for success later in life.

Students undervalued by the present system may have developed unusual resilience and ability to negotiate their own environment, and a set of attributes that enables them to defy negative expectations for success (see Cogan, Sternberg, & Subotnik, 2006; Gordon & Armour-Thomas, 1991; Gordon & Meroe, 1991; Gordon & Song, 1994; Gordon & Wilkerson, 1996; Sternberg, 2006; see essays in Sternberg & Subotnik, 2006). These skills could enable them to succeed—perhaps quite admirably in school—if only the school took advantage of the skills the children have developed. But if the school either fails to reward these skills or actively discourages their display, children with the ability to succeed may actually be unsuccessful.

The best way to measure developing expertise, we believe, is not through static, but rather through dynamic tests. Indeed, dynamic tests were created explicitly to measure developmental potential. We consider the nature of dynamic tests in the next part.

Operationalization of Theory: Dynamic Instruction and Assessment

During the course of their lives, many people will, at some time or another, have taken a conventional test of cognitive skills or achievements. Such tests include IQ tests, as well as other tests that measure some mix of abilities and achievements, which often cannot be distinguished clearly, in any case. Such tests would include A-level tests (created in the United Kingdom), or SATs, ACTs, GREs, LSATs, and many other tests (created in the United States).

Latent Capacities and Developed Abilities

Conventional tests of cognitive skills attempt to quantify developed abilities. If, as we argue above, abilities are always forms of developing expertise (and thus never fully developed), then a measure of developed abilities must be incomplete. For example, these tests might measure a person's ability to retrieve meanings of words. A typical test item of this type might be to define the word *absolution*. Or the tests might measure the person's ability to complete series of numbers, given what the person (a) knows about numbers, (b) the person's ability to infer relations between these numbers, and (c) the person's ability to hold the numbers in working memory. A typical test item might be to say what number comes next in the following series: 1, 4, 9, 16, ?.

Thus, conventional tests measure latent capacity only as it is realized in performance, which, in turn, is affected by many variables, such as amount of education, test-wiseness skills, parental support, and so on. For example, someone with more education will be at an advantage in knowing the meaning of a word or in recognizing a series of perfect squares. Someone with more test-wiseness skills will have techniques available for increasing the probability of responding correctly. For example, such a person may not know what *absolution* is, but know that *to absolve* means to clear of blame, and thereby infer what *absolution* might mean.

These tests measure some unknown mix of abilities that have fully developed and abilities that are not yet fully developed. The extent to which the abilities can develop will depend, in turn, on both latent capacity and the kind of instruction one receives that will help one to develop this latent capacity. Sometimes, the term *ability* is used to refer to a developed latent *capacity*. For example, children brought up in upper middle class households in pricey suburbs are likely to have the educational opportunities that will allow them to make the most—or almost the most—of the latent capacities they have. They are thus likely to score relatively higher on tests of developed abilities. In contrast, children brought up in lower class households in urban slums are much less likely to have the educational opportunities that will allow them to capitalize fully on their latent capacities. They are therefore likely to score relatively lower on tests of developed abilities.

Often, we may wish we could know the extent to which developed abilities reflect latent capacities and the extent to which they reflect developed abilities. In other words, to what extent does a score on a test reflect what a person can do, given the opportunities they have had, and to what extent does it reflect what the person could do, given ideal or nearly ideal opportunities in life. We also may wish to know the difference between the developed abilities and the latent capacities—to what extent do the developed abilities fully reflect the latent capacities? In other words, we may wish to understand the difference between latent capacities and developed abilities.

Consider an example. Alberto and Javier, hypothetical children, have both grown up in Caracas, Venezuela. Suppose, for the sake of argument, they were born with nearly identical latent capacities. Differences between them will begin to emerge very quickly as a result of their different social classes.

Alberto is a child of the upper class and has had extensive and intensive educational opportunities since his birth. He went to an expensive preschool that provided him with basic literacy skills, and then he went to a series of exclusive private schools that spared no expense in educating him. As a first-year university student, he is fully literate and now has a vocabulary well in excess of 100,000 words. He is knowledgeable about mathematics through calculus, and is also knowledgeable about many other subjects, such as science and history. He speaks English as well as Spanish, both of which help him get along in the world and prepare for an intended career in international finance.

Javier is a child of the lower class who has grown up in a rancho in the slums surrounding Caracas. A rancho is a usually illegally constructed dwelling of stone, metal, and whatever elements happen to be lying around that is placed on the ground with no real foundation. A severe storm or even a fairly mild earthquake can be enough to demolish it. There is no running water in the rancho, and the electricity is illegally stolen from power lines. The electricity supports only a few electric lights and a television. Javier had no preschooling and his parents, who are semi-literate, cannot help him develop literacy skills. Javier went to an elementary school on the outskirts of Caracas, but the school had few books or even desks. The schooling was uninteresting and unmotivating. Because Javier was on the streets trying to earn money any way he could from an early age, he did not spend much time at school, and by grade 5 he had dropped out. He was underage for dropping out, but no one was going to make any fuss.

It would make very little difference what conventional test of abilities or achievements Alberto and Javier might take. Alberto would outscore Javier by a substantial margin. There would be no way of knowing that the two children were born with nearly equally capacities. It certainly would be of theoretical interest to have a test that would show that Alberto and Javier had roughly equal capacities. It also would be of practical use: Javier is someone who, with proper educational interventions, might develop into a citizen with useful literacy skills who could make as much of a difference to the society as might Alberto. But how might we obtain information regarding the 2 boys' underlying capacities, and how these capacities differ from their developed abilities?

Defining Dynamic Instruction and Assessment and Comparing Them to Static Instruction and Assessment

Dynamic assessment has been proposed as a way of uncovering this information. What is dynamic assessment? Dynamic assessment is testing plus an instructional intervention. In other words, the instructional and assessment functions, instead of being separated, are integrated. In a conventional assessment, sometimes called a *static assessment,* individuals receive a set of test items, and solve these items with little or no feedback. Often, giving feedback is viewed as a source of error of measurement, and therefore as something to be avoided at all costs. In a *dynamic assessment*, individuals receive a set of test items with explicit instruction (Grigorenko & Sternberg, 1998; Lidz, 1987, 1997; Sternberg & Grigorenko, 2002b; Wiedl, Guthke, & Wingenfeld, 1995). Most importantly from the standpoint of the present research project, *dynamic assessments have been found to reveal developing expertise in members of underrepresented minority groups around the world that is not revealed by conventional static tests* (see e.g., Feuerstein, Rand, & Hoffman, 1979; Lidz & Elliott, 2000; Sternberg & Grigorenko, 2002b, Sternberg et al., 2002).

Why should dynamic instruction and assessment tend to benefit members of underrepresented minority groups in particular? There are at least 5 reasons.

1. Members of such groups may have less tacit knowledge about how to manage themselves in schools, which often reflect middle-class values. Moreover, they may have less knowledge of how to take tests (test-wiseness), due to lesser experience with tests. Dynamic instruction and assessment help make this tacit knowledge explicit.
2. The coldness and interpersonal distance characteristic of static-learning and assessment situations may be more threatening to members of underrepresented minority groups than to others.
3. Members of underrepresented minority groups may have less cognitive scaffolding than do members of other groups. Dynamic instruction and assessment help provide this missing scaffolding.
4. Members of underrepresented minority groups who might disidentify with a static assessment situation may identify with the situation when they are given an opportunity not only to show what they have learned in the assessment situation, but also to learn in this situation.
5. Member of underrepresented minority groups may actually have less developed expertise than do members of others groups. But they may have as great or greater developing expertise, or at least, capacity to develop expertise. Dynamic instruction and assessment help elucidate this developing expertise and capacity to acquire developing expertise.

Two Common Formats for Dynamic Assessment

There are 2 common formats for dynamic assessments. The first format is that the instruction may be sandwiched between a pretest and a posttest. The second format is that the instruction may be in response to the examinee's solution to each test item. Note that they are not the only possible formats, just the two most commonly used ones. We shall use 2 terms of our own invention to describe these 2 formats: the *sandwich format* and the *cake format*.

In the first format, examinees take a pretest, which is essentially equivalent to a static test. After they complete the pretest, they are given instruction in the skills measured by the pretest. The instruction may be given in an individual or a group setting. If it is in an individual setting, it may or may not be individualized to reflect a particular examinee's strengths and weaknesses. If it is individualized, then the amount as well as the type of feedback can be individualized. If it is in a group setting, then the instruction typically is the same for all examinees. After instruction, the examinees are tested again on a posttest. The posttest is typically an alternate form of the pretest, although less commonly, it may be exactly the same test. For convenience, this format will be referred to as the *sandwich format*. In individual-testing settings, the exact contents of the sandwich (type of instruction) as well as its thickness (amount of instruction) can be varied to suit the individual. In group-testing settings, the contents and thickness of the sandwich are typically uniform.

In the second format, which is always done individually, examinees are given instruction item by item. An examinee is given an item to solve. If he or she solves it correctly, then the next item will be presented. But if the examinee does not solve the item correctly, he or she is given a graded series of hints. The hints are designed to make the solution successively more nearly apparent. The examiner then determines how many and what kinds of hints the examinee needs to solve the item correctly. Instruction continues until the examinee is successful, at which time the next item is presented. The successive hints are presented like successive layers of icing on a cake. For convenience, this format will be referred to as the *cake format*. In the cake format, the number of layers of the cake is almost always varied (i.e., the amount of feedback depends on how quickly the examinee is able to use the format to reach a correct solution). The contents of the layers, however (i.e., the type of feedback), may or may not be constant. Most often, they are constant: The number of hints varies across examinees, but not the content of them.

Differences Between Static and Dynamic Assessment

There are 3 major differences between the static and dynamic paradigms. The differences are best viewed as ones of emphasis rather than of dichotomous differences. A static test can have dynamic elements, just as a dynamic test can have static elements.

The first difference regards the respective roles of static states versus dynamic processes. *Static assessment* emphasizes products formed as a result of preexisting skills, whereas *dynamic assessment* emphasizes quantification of the psychological processes involved in learning and change. In other words, static testing taps more into a developed state, whereas dynamic testing taps more into a developing process. In both of the formats of dynamic testing described above, the examiner is able to assess how the problem-solving process develops as a result of instruction. In the sandwich format of dynamic testing, the instruction is given all at once between the pretest and the posttest. In the cake format of dynamic testing, the instruction is given in graded bits after each test item, as needed. Static testing typically does not allow the examiner to draw such inferences.

The second difference regards the role of feedback. In *static assessment*, an examiner presents a graded sequence of problems and the test-taker responds to each of the problems. There is no feedback from examiner to test-taker regarding quality of performance. In *dynamic assessment*, feedback is given, either explicitly or implicitly.

The type of feedback depends on which kind of dynamic assessment is used. In the sandwich format described above, the feedback may be explicit if the testing is individual, but will probably be implicit if the testing is in a group. The instruction sandwiched between the pretest and the posttest gives each examinee an opportunity to see which skills he or she has mastered and which skills he or she has not mastered. But in a group-testing situation, the examiner will not be able explicitly to tell each examinee about these skills. In an individual-testing situation with the sandwich format, it will be possible to provide explicit feedback, should the examiner decide to give it.

In the cake format, the examiner presents a sequence of progressively more challenging tasks; but after the presentation of each task, the examiner gives the test-taker feedback, continuing with this feedback in successive iterations until the examinee either solves the problem or gives up. Testing thus joins with instruction, and the test-taker's ability to learn is quantified while she or he learns.

The third difference between static and dynamic assessment pertains to the quality of the examiner-examinee relationship. In static testing, the examiner attempts to be as neutral and as uninvolved as possible toward the examinee. The examiner wants to have good rapport, but nothing more. Involvement beyond good rapport risks the introduction of error of measurement. In dynamic assessment, the assessment situation and the type of examiner-examinee relationship are modified from the one-way traditional setting of the conventional psychometric approach to form a two-way-interactive relationship between the examiner and the examinee.

In individual dynamic assessment, this tester—testee interaction is individualized for each child: The conventional attitude of neutrality is thus replaced by an atmosphere of teaching and helping. In group dynamic assessment using the sandwich format, the examiner is still helpful, although at a group rather than an individual level. The examiner is giving instruction in order to help the examinees improve on the posttest. As in the individual-assessment format, he or she is anything but neutral.

Thus, dynamic assessment is based on the link between testing and intervention and examines the processes of learning as well as its products. Dynamic assessment is multidimensional in nature (e.g., cognitive, motivational, and metacognitive dimensions). Due to its adaptive nature, dynamic assessment is instrumental in achieving a good fit between the learner and the teacher along any one of the dimensions in instruction and assessment. One of the commonly observed mismatches between students' and schooling is in the area of preferred thinking modalities. While most schools stress verbal and analytical skills in their instructional and assessment practices, many minority children have strongly developed visual and spatial skills (e.g., Tharp, 1989). Consequently, these children's preferential mode of learning is mismatched with the school's preferred mode of instruction. This mismatch has been linked to negative consequences in educational (e.g., school drop-out), social (e.g., peer rejection), and emotional (e.g., low self-esteem) domains of functioning (Alves-Martins, Peixoto, Gouveia-Pereira, Amaral, & Pedro, 2002; Dweck, 1999; Hattie & Marsh, 1996; Lane, Lane, & Kyprianou, 2004).

By embedding learning in evaluation, dynamic assessment assumes that the examinee can start at the "zero (or almost zero) point" of having certain developed skills to be assessed, and that teaching will provide all the necessary information for mastery of the assessed skills. In other words, what is assessed is not just previously acquired skills, but the capacity to master, apply, and reapply skills taught in the dynamic-assessment situation. This view of the testing procedure underlies the use of the term, *test of learning potential*, which is often applied to dynamic assessment.

Specifics of Individual Dynamic Assessment

The individual curriculum-based dynamic assessment (IDA) used in this study incorporates elements from a variety of existing dynamic approaches, with a particular focus on standardized prompts, adaptive testing, and reliance on learners' profiles rather than an overall score, from learning test approach (e.g., Beckman & Guthke, 1995, 1999; Guthke & Stein, 1996); hierarchically developed prompts that afford the least intrusive prompting procedure, from the graduated prompt approach (e.g., Campione, Brown, & Bryant, 1985; Resing, 1993), and student–assessment match from the testing-the-limits approach (e.g., Carlson & Wiedl, 1976, 1978, 1979, 1992; Sternberg & Grigorenko, 2002). The individual curriculum-based dynamic assessment entails a strength-based assessment method where the examiner provides the student with a series of prompts and helps the student to think about various approaches to solving the problem. The IDA can be employed to assess any skill, competence, or ability that can be analyzed in terms of factual, conditional, and procedural knowledge (e.g., Piaget, 1978; Rittle-Johnson, Siegler, & Alibali, 2001).

Why Don't We Hear More About Dynamic Assessment?

The scientific community, especially in the fields of psychology and education, has paid insufficient attention to dynamic testing. There are several reasons behind this lack of attention.

The first reason is the relative lack of published empirical data on the reliability and validity of dynamic testing. Without an adequate database, scholars and educators find themselves unable adequately to evaluate a procedure, and thus may be inclined not to pay much attention to it.

The second reason is, for some approaches, insufficient detail in the presentation of methods—which has made replication difficult. There have been only a handful of reviews of dynamic-assessment studies published in peer-reviewed journals (e.g., Day, Engelhardt, Maxwell, & Bolig, 1997; Elliott, 1993; Grigorenko & Sternberg, 1997; Jitendra & Kameenui, 1993; Laughon, 1990; Missiuna & Samuels, 1988, Sternberg et al., 2002). Most of these studies focus on the educational and clinical applicability of dynamic assessment, rather than on the underlying psychological models and hard empirical data yielded by such assessment.

The third reason is the novelty of dynamic assessment. The constructs are not familiar and do not fit well with what psychologists and educators learn about assessment during the years they are in training. These professionals may therefore be inclined to ignore dynamic assessment because it does not fit at all into their prototype of what assessment is and should be about.

In Sum: The Need for Dynamic Instruction and Assessment

Russian psychologist Sergei Rubinstein (1946) wrote that, in order for an educator to evaluate students' ability to learn, the educator needs to teach students something and then to observe their learning. People draw conclusions about other people's ability to learn—their learning potential—all the time (see also Davydov, 1988). Experts in different fields are able to predict the future performance of novices by first giving the novices a chance to participate in professional activities and then by evaluating their performance while they are learning. When a professor starts working with a student on research, the first step is usually some kind of informal pretest on the student's understanding of the problem to be solved. Usually the student, who has just started working on the problem, does not know much. Therefore, the professor suggests ideas, appropriate readings, and issues on which to concentrate. After a series of subsequent visits and discussions based on the learned material, the professor has enough information to make a preliminary judgment about the learning potential of the student. Similarly, an experienced car mechanic, trying to train apprentices in the garage, gradually involves them in the operation and lets them handle more and more difficult tasks, observing and correcting the novices' performance. In this way, the expert evaluates the ability of the novice to learn.

This kind of implicit prediction of a novice's future achievement, based on learning during an apprenticeship, occurs frequently in everyday life. Now picture a test that measures the ability to learn something new. For example, a person is about to make a career decision. He takes two tests in different fields, let us say, in biology and psychology. The tests are designed in such a way that initially they assess his unassisted performance. Then they measure his performance while working on problems with experts in each field. Each expert is equally effective as a teacher. Finally, the experts measure his individual performance when he is retested. He does equally well on the pretests, but he has learned much more successfully while working with the biology expert than with the psychology expert. Thus, his posttest biology performance is significantly better than his posttest psychology performance.

These results might be interpreted as suggesting that the field of biology is going to be more promising for the person—that he can better realize his potential in this field. Thus, the test provides results that predict to some degree his future performance in the field. Or picture a child whose parents are recent immigrants to a new culture. Irrespective of his or her knowledge of English, if this child is given a conventional test that is not traditional to his or her culture, the child, most likely, is going to demonstrate a fairly low level of performance. On the other hand, if the same child is given a chance to be reevaluated after the test-specific intervention, his or her performance might be drastically different.

One of the most important applications of dynamic testing has been in work with disadvantaged children who have performed exceptionally poorly on conventional static tests (e.g., Feuerstein et al., 1979, Feuerstein, Rand, Falik, & Feuerstein, 2003). The category *disadvantaged* (or, sometimes, *challenged*) *students,* in contrast to *advantaged* (*nonchallenged*) *students,* is used to refer to a large class of pupils viewed as having reduced learning opportunities. This reduction can be due to deficient previous education, lack of match in previous and current cultural and educational practices, or to apparent learning disability or mental deficiency. The claim that these students should be tested dynamically is motivated by the belief that dynamic testing in its proper application can help reduce educational inequalities by providing what are seen as more compassionate, fair, and equitable means for assessing students' learning capacities. For disadvantaged children, quantifying their learning in action, with the assistance of and under the supervision of an adult, might be the only way to evaluate their true level of functioning.

The idea of developing a methodological paradigm that goes beyond the measurement of developed abilities and that quantifies the potential that will be a main force in students' learning is an extraordinarily appealing idea to scientists and laypeople alike. A number of synonymous or nearly synonymous concepts, traditionally unified under the name "dynamic testing/assessment" (e.g., interactive testing/assessment, process testing/assessment, measuring the zone of proximal development, assisted testing/assessment, and tests of learning potential), have been suggested for this paradigm (see Grigorenko & Sternberg, 1998; Sternberg & Grigorenko, 2002, for more details).

The focus of the research project described here is on the development of content-specific forms of dynamic assessment (DA) designed to reduce costs while maintaining the philosophical foundations of DA. There are 2 primary studies as part of this project. Study 1 investigates group-administered dynamic assessment. Study 2 investigates individually-administered dynamic assessment.

Method

Design Overview

There are 2 main studies in the project. The first investigates group-administered dynamic testing; the second considers individually-administered dynamic assessment. To reduce redundancy in data collection, common control groups were used where appropriate across studies. Let us describe the overall design before describing the participant samples, materials, and detailed procedure.

The overall design of Study 1 is represented in Figure 1. It consists of a common pretest to all participants, plus administration 1 of 3 interventions: ([1] Dynamic Intervention, [2] triarchic-control intervention, or [3] a standard-control intervention), and followed by 1 of 2 possible posttests: (1) a group-administered dynamic posttest or (2) the same posttests interspersed with a filler activity).

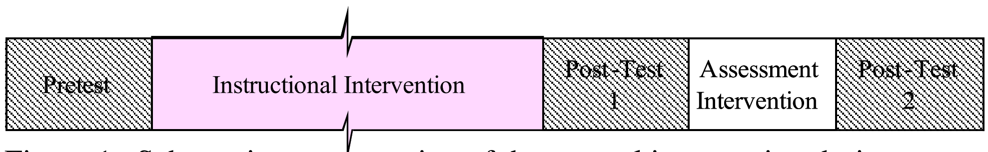

Figure 1. Schematic representation of the general intervention design.

In Study 2, students were provided with dynamic and triarchic instruction (as in conditions 1.1. and 1.2), but assessed individually rather than in a group format. The individual assessment format includes a matrix of prompts by modality (Grigorenko, Birney, Jeltova, & Sternberg, 2002) (see Appendix). The first 3 prompts relate to reading comprehension, attention, and basic problem-solving skills. The next 3 levels of prompts are related to conditional, procedural, and factual knowledge of mathematics. The student's math skills are assessed across memory, analytical, creative, and practical cognitive modalities. Adaptive testing is used as each modality includes a calibrating item and easier and harder items. It is a strength-based assessment in that it diagnoses both deficits and strengths while identifying ways for remediation.

Experimental Conditions

There were 7 different conditions across the 2 studies as described in Table 1. The experimental design for Study 1 is more clearly described in Figure 2.

As summarized in Table 1, participants received 1 of 3 types of instruction: [1] combined dynamic & triarchic instruction (conditions 1.1, 1.2, and 2.1), [2] triarchic instruction (conditions 1.3 and 1.4), or [3] standard instruction (conditions 1.5 and 1.6). The experimental (target) intervention being evaluated was a combination of dynamic and triarchic instruction with group-administered dynamic assessment. The first control entailed only instruction based on the triarchic curriculum and was thus a direct test of the efficacy of the features of the dynamic instructional component while holding content constant. The second control is a standard instructional intervention, that is, students are taught using the schools existing curriculum.

Also, as summarized in Table 1, students were assessed either in groups (conditions 1.1 - 1.6) or individually (condition 2.1). There are two types of group-administered assessments. The main difference was the content of the 30-minute period of time between the two posttests (post-B1 and post-B2). For the experimental group, the 30-minute instructional period includes a discussion of problem-solving strategies and a demonstration of 2 of the items from the first posttest using problem-solving strategies. For the control condition, students were given an unrelated filler activity between posttests B1 and B2.

Table 1

List of the 7 Independent Project Conditions

Condition Number	**Description**	**Code**
STUDY 1		
Condition 1.1	Dynamic + Triarchic Instruction: Group DA with 30 min process training intervention	DI-GDA
Condition 1.2	Dynamic + Triarchic Instruction: Group DA with 30 min filler intervention	DI-FDA
Condition 1.3	Triarchic Instruction: Group DA with 30 min process training intervention	TI-GDA
Condition 1.4	Triarchic Instruction: Group DA with 30 min filler intervention	TI-FDA
Condition 1.5	Standard Instruction: Group DA with 30 min process training intervention	SI-GDA
Condition 1.6	Standard Instruction: Group DA with 30 min filler intervention	SI-FDA
STUDY 2		
Condition 2.1	Dynamic + Triarchic Instruction: Individual DA	DI-IDA

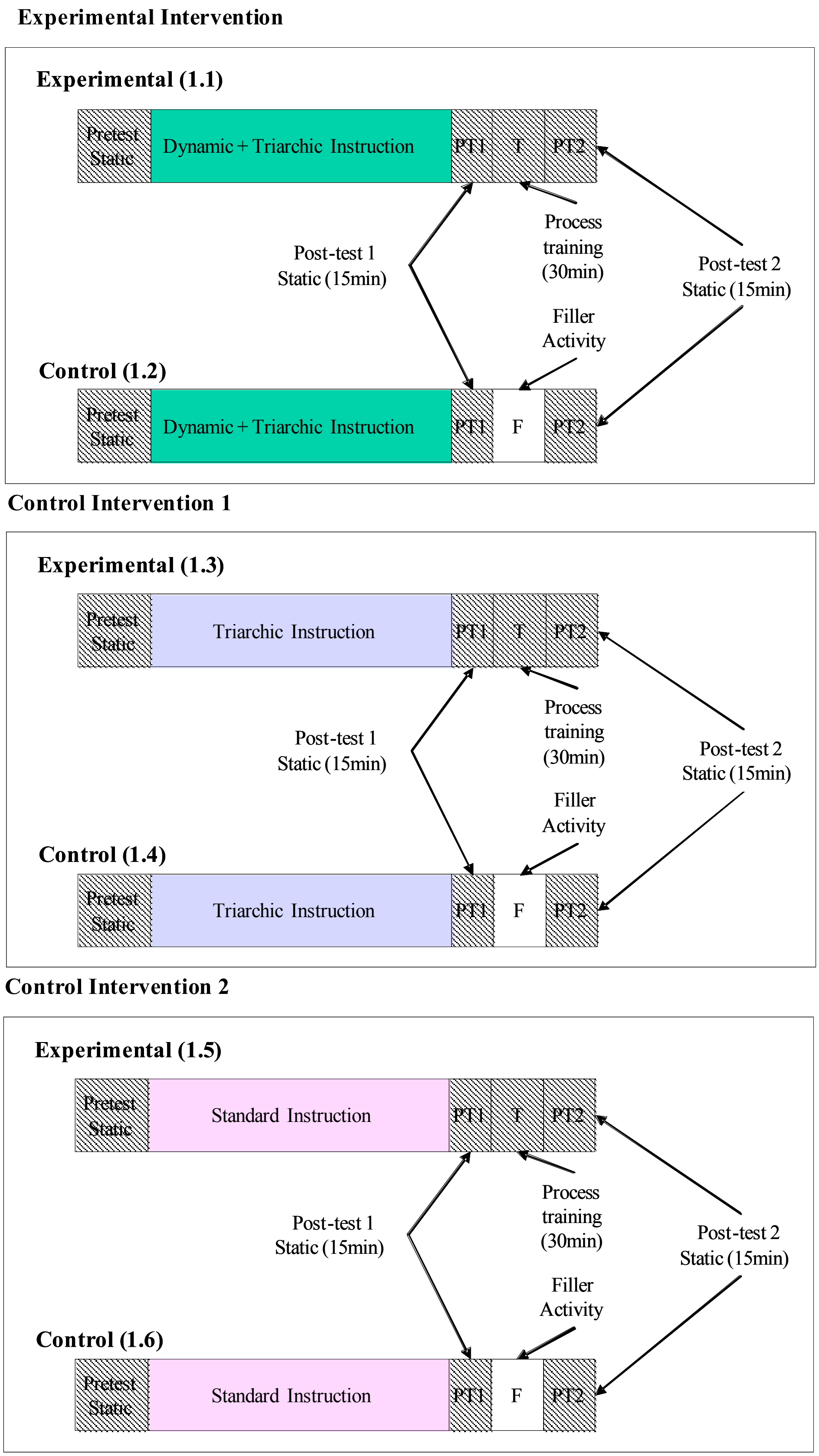

Figure 2. Schematic representation of Study 1.

Participants

In total, 1,500 students and 71 classroom teachers, in 24 schools across 6 school districts participated in the study. The distributions of the sample across experimental condition, district, gender, and ethnicity are reported in Tables 2 to 4. We sampled students from 4 ethnic groups: White, Asian American, African American, and Hispanic American[1].

As a function of the geographical location of the study, there were a majority of White students (*n*=594), and the breakdown was fairly equivalent among the other 3 ethnicities, with approximately 300 students in each group (321 Asian American, 246 African American, and 292 Hispanic American). All students were given instruction and/or dynamic assessments (either individually or group administered) nurturing (instruction) and measuring (assessment) their developing expertise in mathematics. Control participants were divided into no-treatment and irrelevant-treatment instructional groups. The 6 school districts were all located in the Northeast (Connecticut and New York States), but represented a mix of socio-economic groups, with median incomes ranging from $33,809 (New London, CT) to $60,556 (Stamford, CT). The majority of teachers were female (84.51%, versus 14.08% male and 1.41% no gender recorded) and White (73.24% versus 5.63% African American, 7.04% Asian Americans, 11.27% Hispanic Americans, and 2.82% Other), and they represented a range of level of experience, having taught between 1 and 38 years, with a median teaching experience of 16.28 years. A total of 10.81% of the teachers held a BA or BS as their highest degree, while 67.57% held an MA or an MS.

[1] Note: "White" includes students with origins in Europe, the Middle East, North Africa, and Arab countries. "Asian American" includes students of East Indian, Pakistani, Burmese, Hong Kong, and Thai origins. "African American" includes students from Jamaican.

Table 2

Number of Teachers and Students by District and Experimental Condition

District,		**Experimental**		**Triarchic Control**		**Standard Control**		**Individual**	
state		**DI-GDA**	**DI-FDA**	**TI-GDA**	**TI-FDA**	**SI-GDA**	**SI-FDA**	**DI-IDA**	**Total**
	Teachers	5	2	0	0	0	0	3	**10**
Vernon, CT	Students	113	36	0	0	0	0	61	**210**
	Teachers	0	0	17	6	0	0	0	**23**
Hamden, CT	Students	0	0	335	106	0	0	0	**441**
	Teachers	1	0	0	0	4	6	4	**15**
New York, NY	Students	26	0	0	0	96	167	85	**374**
	Teachers	0	4	2	2	0	0	0	**8**
New London, CT	Students	0	73	50	44	0	0	0	**167**
	Teachers	0	0	0	6	0	0	1	**7**
Stamford, CT	Students	0	0	0	121	0	0	22	**143**
	Teachers	0	0	0	0	8	0	0	**8**
Danbury, CT	Students	0	0	0	0	165	0	0	**165**
	Teachers	**6**	**6**	**19**	**14**	**12**	**6**	**8**	**71**
All Districts	**Students**	**139**	**109**	**385**	**271**	**261**	**167**	**168**	**1500**

Table 3

Experimental Condition by Gender

	Experimental		**Triarchic Control**		**Standard Control**		**Individual**	
	DI-GDA	**DI-FDA**	**TI-GDA**	**TI-FDA**	**SI-GDA**	**SI-FDA**	**DI-IDA**	**Total**
Female	72	50	183	113	123	83	92	716
Male	65	58	197	155	127	76	76	754
Unknown	2	1	5	3	11	8		30
Grand Total	**139**	**109**	**385**	**271**	**261**	**167**	**168**	**1500**

Table 4

Experimental Condition by Ethnicity

	Experimental		**Triarchic Control**		**Standard Control**		**Individual**	
	DI-GDA	**DI-FDA**	**TI-GDA**	**TI-FDA**	**SI-GDA**	**SI-FDA**	**DI-IDA**	**Total**
African American	5	20	106	72	29	1	13	246
Asian American	27	4	27	14	81	138	30	321
White	96	62	174	113	84	11	54	594
Hispanic American	8	21	70	65	56	7	65	292
Mixed	1		3	2		2	5	13
Other		1		2		1	1	5
Unknown	2	1	5	3	11	7		29
Grand Total	139	109	385	271	261	167	168	1500

Materials

In this section, we will first describe the instructional interventions (dynamic & triarchic, triarchic, or standard), and then describe the assessments (group or individually administered). Instructional and assessments materials were developed for 4 mathematical content areas appropriate to fourth grade: Number Sense and Place Value, Equivalent Fractions, Measurement, and Geometry, as described below.

Number Sense and Place Value—In this unit, students use number lines to identify and understand negative numbers and the ordering of numbers. They are led to an understanding of how to use the place-value structure of the Base 10 number system and how to identify factors and generate equivalent representations of numbers to use in problem solving. In addition, students explore even/odd numbers, square numbers, and prime numbers.

Equivalent Fractions—This unit is intended as a follow-up to an introductory fractions unit. In it, students develop an understanding of the concept of equivalence, model equivalent fractions with concrete manipulatives, identify and generate equivalent fractions (denominators less than 12), and apply the concept of equivalent fractions in practical and problem solving situations.

Measurement—Students learn to measure quantities (including time, length, perimeter, area, weight, and volume) in everyday and problem situations. They compare, contrast, and convert within systems of measurements (customary and metric) and estimate measurements in everyday and problem situations. In addition, students learn about the use of appropriate units and instruments for measurement.

Geometry—Students are engaged in the identification and modeling of simple 2-dimensional and 3-dimensional shapes and develop an understanding of their properties (review perimeter, area, and volume). Students are expected to understand and identify geometric concepts such as "congruent," "similar," and "symmetric." Finally, students combine, rotate, reflect, and translate shapes.

Interventions

Dynamic + Triarchic Instruction (Experimental Intervention)

The Triarchic units that are used are identical in the Experimental and the Control Intervention 1 conditions (Triarchic-Only condition). Both sets of teachers are instructed in Triarchic theory, and are given training exercises that focus on creative and practical intelligence as well as more traditional analytic and memory skills (see the Triarchic Intervention for more details). The general philosophy of the dynamic assessment that we adopted in the development of a group administered procedure is the integration of instruction with assessment (Grigorenko et al., 2002). The written content of the units remained identical, but the pedagogical framework differed.

The main difference between the dynamic + triarchic-instruction and the triarchic-only instruction, was that teachers were given instruction on the implementation of DA principles and processes in a 2-day workshop. They were then instructed to use the same triarchic instructional materials as those used in triarchic-only intervention, but to infuse the DA principles learnt in the 2-day workshop into their teaching.

Triarchic Instruction (Control Intervention 1)

The theoretical basis for the triarchic instructional units is Sternberg's (1985, 1988, 1997, 2005) triarchic theory of successful intelligence. According to this theory, intelligence results from information-processing components being applied to experience for the purposes of adaptation to, shaping of, and selection of environments. In particular, it is the ability to achieve success in life, given one's own personal standards, within one's sociocultural context; in order to adapt to, shape, and select environments; via recognition of and capitalization upon strengths and recognition of and correction of or compensation for weaknesses; through a balance of analytical, creative, and practical skills. Intelligence and the intellectual skills that constitute it are seen to form the basis of intellectual achievements and are forms of developing expertise—they can be developed, just like any other forms of expertise (Sternberg, 1998, 1999).

The curricula used in the current project were built on the principles of the triarchic theory of successful intelligence and thus focuses on 3 main types of thinking and instruction:

Analytical thinking occurs when the reasoning processes are applied to relatively familiar types of problems in their abstracted form. Analytical thinking is involved when people analyze, evaluate, judge, compare and contrast, and critique. For example, a student might be asked to evaluate the assumptions underlying a logical argument or to compare and contrast the themes underlying 2 short stories.

Creative thinking occurs when the components of information processing are applied to relatively novel types of problems. Creative thinking is involved when people create, invent, discover, explore, suppose, and imagine. For example, a student might be asked to create a poem or to invent a better mousetrap.

Practical thinking occurs when the components of information processing are applied to highly contextualized, everyday problems. Practical thinking is involved when people apply, use, utilize, implement, and contextualize. For example, a student might be asked how the lessons of the Vietnam War are and are not relevant to modern-day conflicts, or how to apply algebraic techniques to determining compound interest on an investment.

The units were developed and evaluated as part of another larger project and are appropriate for fourth grade students across the USA. The units used in the current project were Number Sense (a training unit), Geometry, Measurement, and Equivalent Fractions, roughly in this order.

Standard Instruction (Control Intervention 2)

This was the instruction focused on the general content of the triarchic units that was being implemented in the classrooms at the time.

Assessments

Group-Administered Dynamic Assessment (GDA)

The focus of assessment in Study 1 is group administration. The general rationale is that Dynamic Assessment methodologies are more likely to be taken up in practice if it fits well within the practical constraints of the classroom setting. The group-administered approach adopted in this study used Sternberg and Grigorenko (2002b) sandwich model. The implementation of this model entailed a pre-intervention-post design implemented over a 1 hour session.

There are 2 types of group-administered assessments, with the main difference having occurred in the content presented in the 30-minute period of time between the two posttests (post-B1 and post-B2). For the experimental group, the 30-minute instructional period after the first posttest includes a discussion of problem-solving strategies and a demonstration of 2 of the items from the first posttest using problem-solving strategies. Students are encouraged to participate in the discussion of how to break down the chosen test item, what steps are needed to solve it and what the thinking is behind the choice of steps. Table 5 describes the problem-solving strategy taught to students. Once the 2 items had been demonstrated, the students were encouraged to try their best on the second posttest and to employ the problem-solving strategies and thinking-skills that were just reviewed with them. At this point, the second 15-minute posttest was distributed.

Table 5

Problem-solving Strategy

Strategy for problem solving IDEA	
I	**Identify the question. What do you need to solve?**
D	**Define the ways to solve the problem.**
E	**Evaluate each way you came up with in D.**
A	**Apply your way and write down your answer. Your answer here should match your answer in I.**

The group-administered dynamic assessment consisted of a 15-20 minutes test (post-B1), followed by an in-depth 20-30 minutes teacher-led class discussion focusing on the problem-solving principles required. This was then followed by a second 15-20 minutes test (post-B2). It is important to note that the complete 3-part procedure was considered the dynamic assessment. That is, the dynamic assessment consisted of "posttest 1—process-training—posttest 2" design.

Group-administered Control (Filler) Assessment (FDA)

As a control and to allow further investigation of the nature of the group-administered dynamic assessment, a subset of students completed the same post-B1 and post-B2 assessments, but this time, instead of an intervening process-training session, an unrelated filler activity was conducted. This allowed us to investigate (a) the extent to which the process-intervention is a necessary component of the dynamic assessment, and (b) the magnitude of practice effects. Note: The expectation was that the effects of the group administered DA (posttest 1—intervention—posttest 2) are not simply due to practice from posttest 1 to posttest 2. This is because the rank-ordering from posttest 1 to posttest 2 is not expected to necessarily remain the same.

Individual-administered Dynamic Assessment (IDA)

Study 2 focused on individual assessment, although in practice this was interspersed with group assessment. The pre- and post-assessments serve to identify learners' strongest cognitive modality (e.g., practical vs. analytical), as well as their initial level of competence in a given content area (e.g., equivalent fractions). For example, the pretest may reveal that a student is most competent in solving math problems in a practical modality, then in analytical, memory, and finally a creative modality, as evidenced by the minimal number of prompts needed to assist the student with problem-solving and more difficult items. These results suggest the student needs to be helped with transferring his/her competences in the practical modality to the analytical and memory modalities, as these are the 2 commonly used modalities in traditional testing (e.g., standardized national testing). Further, the number and nature of prompts required for the student to solve problems across different modalities will suggest to the teacher how to best plan instruction. Emphasis may be placed on building conditional knowledge and general problem-solving skills (e.g., recognizing type of the problem).

For each unit, teachers were provided with 5 items of increasing difficulty for each cognitive modality (memory, analytical, practical, and creative), or a total of 20 items per instructional unit. The teacher and 2 assistants administered the test individually to every child.

Children were started on item #3 (calibrating item) for one of the modalities. If the answer was incorrect, the child was given the first of 6 graduated prompts. If the child still could not answer (or answered incorrectly) the second prompt was given, and so on until a correct answer was provided, or all of the prompts were used. If the first item presented (Item #3) was answered correctly with 3 or fewer prompts, the child was then asked to answer item #4. If the child got this item correct with 3 or fewer prompts, he or she went on to item #5, and then on to the next cognitive modality. If a child needed 4 or more prompts on item #3, the examiner then gave them item #2. If this item was passed with 3 or fewer prompts, the child was then given item #4.

For each modality, 3 of the 5 items were usually given, unless the examiner felt they needed more information about the child's knowledge within a specific modality.

In the Appendix we present the matrix used for individual assessment, along with the guidelines provided to teachers.

Procedure

The intervention materials are 4 mathematics units presented using one of the following: Enhancement of the triarchic curriculum using dynamic pedagogy (developed by Columbia Teachers College), Triarchic curriculum alone, or Standard curriculum.

Before the intervention is implemented, students' baseline performance is assessed using the following measures: the Mill Hill (Raven, Raven, & Court, 1995), the Cattell

Culture Fair test of g reasoning (Cattell & Cattell, 2002), and a math baseline test using items from the fourth grade Ohio Proficiency Tests.

In addition, students' content knowledge is assessed before and after the implementation of each curriculum unit. The assessments are the same for participants in all 3 intervention conditions. Table 6 lists the different assessments administered throughout the study.

Table 6

Assessments Administered to Participants

Measures	Year 1	Year 2
Reading		
Mill Hill (vocabulary test)	X	
Year 2 Baseline Reading		X
Mathematics		
Year 1 Math Baseline (Yale School Assessment Form A)	X	
Year 2 Math Baseline (Yale School Assessment Form B)		X
Post Math Assessment		
Year 1 Post Study Math Test (Yale School Assessment Form B)	X	
Year 2 Post Study Math Test (Yale School Assessment Form A)		X
Reasoning		
Cattell (reasoning test)	X	
Number Sense Units (demonstration/practice unit)		
Number Sense DA Individ Pretest (matrix)	X	X
Number Sense DA Individ Posttest (matrix)	X	X
Number Sense Static Group Pretest A	X	X
Number Sense Static Group Posttest B	X	X
Number Sense DA Group Posttest B1	X	X
Number Sense DA Group Posttest B2	X	X
Number Sense Workbook	X	X
Geometry Units		
Geometry DA Individ Pretest (matrix)	X	X
Geometry DA Individ Posttest (matrix)	X	X
Geometry Static Group Pretest A	X	X
Geometry Static Group Posttest B	X	X
Geometry DA Group Posttest B1	X	X
Geometry DA Group Posttest B2	X	X
Geometry Student Workbook		
Measurement units		
Measurement DA Individ Pretest (matrix)	X	X
Measurement DA Individ Posttest (matrix)	X	X
Measurement Static Group Pretest A	X	X
Measurement Static Group Posttest B	X	X
Measurement DA Group Posttest B1	X	X
Measurement DA Group Posttest B2	X	X
Measurement Student Workbook	X	X

Table 6 (continued)

Assessments Administered to Participants

Measures	Year 1	Year 2
Equivalent Fractions Units	X	X
Equivalent Fractions DA Individ Pretest (matrix)	X	X
Equivalent Fractions DA Individ Posttest (matrix)	X	X
Equivalent Fractions Static Group Pretest A	X	X
Equivalent Fractions Static Group Posttest B	X	X
Equivalent Fractions DA Group Posttest B1	X	X
Equivalent Fractions DA Group Posttest B2	X	X
Equivalent Fractions Workbook	X	X
Teacher Interview		
Geometry Teacher Interview	X	X
Measurement Teacher Interview	X	X
Equivalent Fractions Teacher Interview	X	X
Number Sense Teacher Interview	X	X
Student Interview		
Geometry Student Interview	X	X
Measurement Student Interview	X	X
Equivalent Fractions Student Interview	X	X
Number Sense Student Interview	X	X
Other measures		
Creative Story Writing (filler task)	X	X
Creative Collage Task (filler task)	X	X

Baseline	
Language proficiency	General vocabulary test
Reasoning skills	General reasoning test
Classroom instruction	Behavioral observations
Math achievement	Pre-project math test
During project (Pre- and post- unit)	
Math achievement	Unit specific assessment
ZPD	Unit-specific dynamic assessment
Teacher and student SE	Teacher and students' SE beliefs
Post-project	
Same assessments as for the baseline	

Results and Discussion

The main hypothesis is that, whereas learning gains in the experimental condition will exceed those in the control conditions across the 4 ethnic groups (Asian American, White, African American, Hispanic American), the difference will be especially pronounced in the ethnic minority groups. Thus, it is hypothesized that dynamic tests will reduce or eliminate differences among groups at the same time they provide more equitable, fair, and comprehensive assessments of skills.

Data Preparation

There are 3 intervention tests being analyzed (a fourth unit, *Number Sense*, was administered at the beginning and predominantly used as a "practice" unit for the teachers to familiarize them with the triarchic and dynamic methods). These were *Geometry*, *Measurement*, and *Equivalent Fractions*. The intervention measures consist of a pretest and 3-step posttest (test-training-test). Data consisted of student responses to multiple-choice and free response items.

Multiple-choice student data were entered by project personnel trained and supervised by the main investigator. The following procedure was used to train raters to score short-response items: First, raters were grouped in pairs. After a general introduction to the instruments, each rater-pair scored 30-50 tests using the specially developed rubrics. The rater-pairs needed to reach a level of .70 inter-rater reliability (correlation) on every item before being allowed to move on to independent scoring. Under the supervision of senior project staff, raters discussed individual scores for items that did not reach the .70 criterion. If disagreement was substantial, the process was repeated for a different selection of tests. Over the whole project, short response answers were rated by 31 independent raters. To equate rater severity, 30% of all data were scored by two independent raters. All data were entered into Excel spreadsheets, which were then imported and managed in an Access database.

To prepare data for analysis, Rasch analyses were performed to equate the pretest and posttest to the same scale. We use the FACETS approach to achieve this. FACETS program calibrates person ability, item difficulty, and rater severity on the same scale and therefore is particularly well suited to equating the pre and posttest. Common items are used to anchor the scale of the two tests.

Validity of the Individual Dynamic Assessments

One of the commonly reported strengths of using prompting is that the obtained data have the potential to provide a very rich source of information about students' understanding and achievement. The primary objective of our analyses was to demonstrate that IDA, as a measure of achievement, is (a) internally valid (i.e., measures the same construct across all 4 cognitive modalities) and (b) externally valid (i.e., has practical utility in predicting external measures of mathematics achievement).

To address the issue of internal consistency, Cronbach's α was calculated for each 20-item assessment. It did not fall below 0.87 for any test, as specified in Table 7. This indicates that the testing procedure is yielding internally consistent results across all the instructional units throughout the study. This is important considering that the tests contain items that are balanced across levels of difficulty and learning modalities (e.g., analytical, practical, creative, and memory). While the items presented in analytical and memory-based modalities appear to be more transparent and directly linked to specific math competencies, the items presented in creative and practical modalities may be perceived as being very different and possibly measuring different competencies. Our data refute this concern and indicate that the individual dynamic assessments measured the same construct (competencies) across all 4 cognitive modalities (memory, analytical, practical, and creative).

Table 7

Individual Dynamic Assessments—Descriptive Statistics for Proportion of Prompts Required for Solution*

Test	*N*	Minimum	Maximum	Mean	Std. Deviation	Cronbach's Alpha
Geo_pre	38	.00	7.00	3.95	1.70	.905
Geo_post	38	.00	7.00	2.27	1.39	.870
Meas_pre	55	.00	7.00	4.42	1.36	.870
Meas_post	55	.00	5.00	3.54	1.40	.886
EF_pre	35	.00	7.00	3.09	1.35	.880
EF_post	20	.00	7.00	2.14	1.36	.886
Valid N (listwise)	20					

* Aggregated across all modalities and items

Simple one-tailed paired-sample *t*-tests were conducted to test student gain from pre to posttests. Significantly fewer prompts were required at posttest than at pretest in Geometry ($p<.000$), Equivalent Fractions ($p<.000$), and in Measurement ($p<.010$) (see Figure 3). This indicates that IDA was sensitive to gains made during instruction. Furthermore, the number of prompts that students required at each pretest in dynamic assessment was decreasing across the units ($F=8.09$, $p<.000$) with pairwise comparisons using Bonferroni procedure being significant across all units ($p<.000$). These statistically significant results translate into important practical implications. Further analyses need to examine whether the decrease in the number of prompts required at pretest was due to the students' and teachers' growing comfort with the IDA procedure, i.e., students internalizing the prompting procedure early in the process, which then lead to a decrease in their need for external prompting. Future research needs to investigate whether the first tier of prompting (prompts 1-3) that focuses on reading comprehension and general problem solving/test taking skills is what needs to be used once the students internalize the knowledge-based prompting in tier 2. Observational data suggest that students begin to internalize the prompting procedure after the first unit, and then start to anticipate and self-administer prompts (e.g., a student would prompt him/herself out loud before the examiner delivers the prompts).

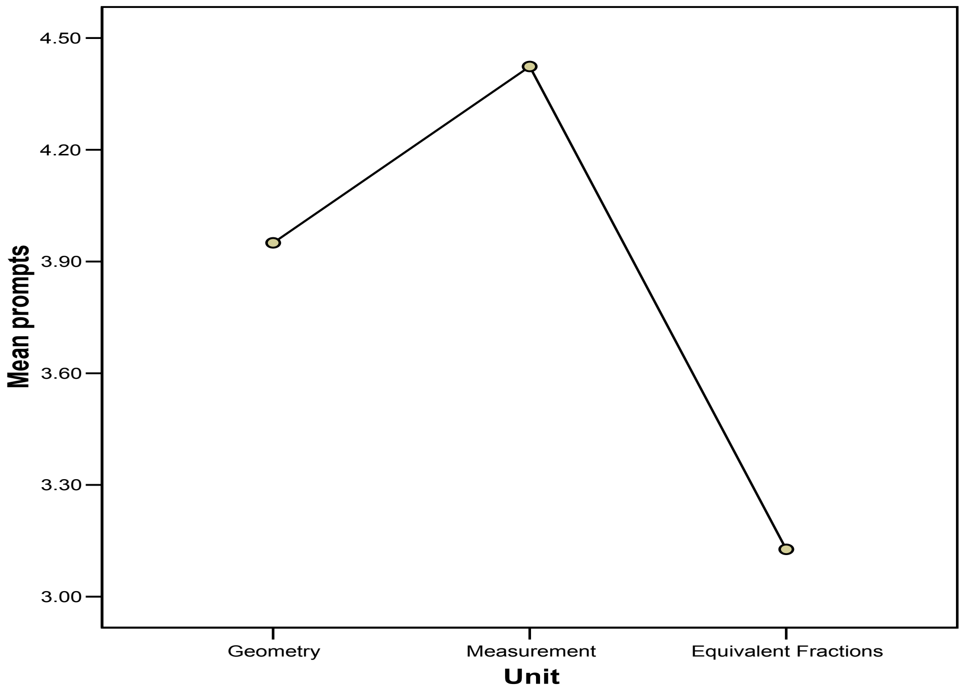

Figure 3. Mean number of prompts administered at pretest across the units.

To gauge the IDA's prompting procedure as a measure of achievement, the distribution of the level of prompts (0-6) required in each modality (analytical, memory, creative, practical) was investigated for each of the 3 units. Analyses revealed moderate degree of variability in item difficulty as a function of modality and the specific content of the unit. Overall, it was found that the distribution of prompts administered across items and modalities was bi-modal with about 15% of students requiring the first 3 prompts to solve the problem correctly, with only about 10% requiring 6 prompts, and with about 40% either answering the items correctly without prompting or failing it despite the prompting. The 25% of students who gain from prompts (i.e., have the potential to answer the questions correctly) represent a substantial group of students who may be otherwise underperforming on tests. For example, during pretest for the *Measurement* unit, upon receiving 3 prompts the number of students answering the items correctly increased by 17% for Analytical modality (by 36% upon receiving all 6 prompts), by 22% for Practical and Creative modality (by 36% upon receiving all 6 prompts), and by 9 % for Memory. At posttest for the *Measurement* unit, the number of students answering the items correctly increased on average by 12% upon receiving first

3 prompts. Also, as can be seen in Table 8, the items presented in creative modality at pretest consistently required more prompting than items in other modalities even though the items were created to be equivalent in content and level of difficulty. At posttest, however, creative items required either the same or a lower number of prompts suggesting that students were comfortable with this modality. Future research needs to investigate whether the schools' supposed underutilization of students' creative thinking skills is related for elevated level of prompting at pretest.

Table 8

Average Number of Prompts Needed in Each Unit at Pretest and Posttest Across Modalities

	Geometry		Measurement		Eq. Fractions	
	pre	post	pre	post	pre	post
Analytic	3.6	3.1	4.0	3.0	1.5	2.4
Practical	3.2	2.9	4.2	3.7	2.6	2.6
Creative	4.7	1.7	4.1	3.6	4.8	2.3
Memory	4.3	1.4	5.4	3.8	3.4	1.22
Total	**3.95**	**2.27**	**4.42**	**3.54**	**3.09**	**2.14**

In summary, these data analyses suggest that: (a) our IDA measures are internally consistent; (b) the findings are consistent with our expectations of an educational intervention, in that students required substantially fewer prompts at posttest than at pretest and, thus, a sensitivity of our measure; and (c) the prompting procedure helped to increase the number of problems solved.

Future research needs to address the role that reading comprehension and general problem solving prompts in IDA play in mathematical competence. Current research suggests unique roles of language in performance of students whose primary language is not English in content areas such as mathematics (e.g., Abedi & Lord, 2001). On average, students who answered 8 out of 40 questions correctly when assessed statically, answered 30 out of 40 questions correctly on a parallel version of the same test when assessed using IDA. In other words, when provided with reading comprehension prompts students were observed to solve over 50% of the initially failed items correctly. Reading comprehension may be critical to mathematics performance for students of diverse backgrounds.

In sum, obtained results indicate overall validity and reliability of the IDA. At the same time, important lines for future research were identified: The methodology for individual dynamic assessment combines several components and each component needs to be further validated and evaluated in terms of its unique contribution to the observed effects. These components are (a) evaluation and improvement of student–assessment fit by presenting tasks in different cognitive modalities, allowing for linguistic modifications, and differentiating between low math competence vs. poor problem solving skills vs. poor reading skills, (b) making testing adaptive, (c) using standardized, least-intrusive prompting procedures, and (d) building in redundancy and transparency in assessment that facilitates internalization and independent utilization of the prompting as a self-guiding procedure.

Baseline Performance

To control for individual differences in the experiences and knowledge that students bring to the study, a variety of baseline measures were included to be used as covariates in the analyses. Specifically, we administered the following assessments:

- Math Baseline: Yale School Assessment: Form A and Form B administered at baseline and post-study in year 1 and year 2, respectively. There are 9 items common across Form A and B allowing for equating.
- Math Post-study: As for Math Baseline. Students did the same test at post-study as they did at baseline. In year 1, this was Form A; in year 2 it was Form B.
- Cattell Culture Fair Test of "g": Test of general reasoning ability. Only administered in year 1. Two (sub)tests (out of 4) were administered, Test 2 (k=14 items) and Test 4 (k=8 items). Despite this being a commercially published assessment of general reasoning ability, the reliabilities of the scales were rather low, and analyses presented here are based on a subset of items that remained after psychometrically poor items were removed.
- Mill Hill Vocabulary Test: The Mill Hill is a commercially published 32-item general vocabulary test. It has also been used as an indicator of Crystallized Intelligence (Gc). The test was only administered in year 1. Rasch calibrations of persons and items were satisfactory.
- Reading baseline: The Yale School Program Reading test is a 20-item comprehension test in which students are required to read 3 different forms of information presentation and to respond to a combination of multiple-choice and short-answer questions associated with each. Reliability of the scale is good.

For these baseline measures, although we expected significant individual differences in the abilities and competencies tapped by these measures, we expected minimal differences between experimental conditions once individuals and classrooms were taken into consideration in the hierarchical models. Data show that while there is only minimal evidence overall that performance on the baseline measures differed as a function of the intervention conditions, each test is associated with significant individual differences at the individual level and therefore will be considered as covariates in the main analyses. As expected, there are differences in the post-study math score as a function of instructional condition. These differences are considered further on.

Impact of Experimental Condition

Table 9 reports the number of participants available for each core intervention measure by experimental condition. Table 10 reports the pair-wise data available for analyses.

Table 9.

Complete Data for Each Intervention Measure by Experimental Condition

Experimental Condition	Geo Pre	Geo B1	Geo B2	Meas Pre	Meas B1	Meas B2	EqF Pre	EqF B1	EqF B2	Listwise N
DI-GDA	131	131	131	51	26	26	62	113	108	24
DI-FDA	63	66	66	0	0	0	84	71	75	0
TI-GDA	343	346	342	310	303	291	201	179	201	145
TI-FDA	139	136	133	247	220	214	93	38	38	28
SI-GDA	166	142	98	249	215	194	84	71	67	57
SI-FDA	155	156	151	130	128	128	159	157	155	111
DI-IDA	74	NA	NA	105	NA	NA	61	NA	NA	43
Grand Total	**1071**	**977**	**921**	**1092**	**892**	**853**	**744**	**629**	**644**	**365**

Geo = Geometry; Meas = Measurement; EqF = Equivalent Fractions; pre = pretest; B1 = post-B1; B2 = post-B2; DI = Dynamic Instruction; TI = Triarchic Instruction; SI = Standard Instruction, GDA=Group-Administered Dynamic Assessment; FDA = Group-Administered Filler Assessment; IDA = Individually Administered Dynamic Assessment

Table 10.

Data Available for Analyses After Pair-wise Deletion to Allow Assessment of Post-B1 With Pretest as a Covariate

	DI-GDA	DI-FDA	TI-GDA	TI-FDA	SI-GDA	SI-FDA	Total
Geometry	128	47	328	125	96	144	868
Measurement	26	0	286	207	193	123	835
Equivalent Fractions	61	73	180	35	65	151	565

To test our main hypothesis, we considered the interactive effect of instructional condition and assessment-type on post-B2 performances. That is, we considered differences between the true "sandwich" dynamic assessment condition (Group Dynamic Assessment - GDA) where students received post-B1 and post-B2 with the process-training session between, and the reduced condition ("Filler" Dynamic Assessment -

FDA), which includes a filler activity between post-B1 and post-B2. This allows us to evaluate changes from post-B1 to post-B2 due to factors associated with repeated-testing alone (e.g., practice, fatigue). Assessment type is then of course crossed with the *instructional condition* (Dynamic, Triarchic, Standard). This is first to explore whether dynamic instruction results in improved performance over the triarchic-only condition and the standard instruction. The interactive effect is to explore whether instructional condition modifies the difference between the true (GDA) group-administered dynamic assessment which includes process-training as a core component, and the filler dynamic assessment (FDA). The primary outcome variable is performance on the post-B2 test. Pretest assessments at the start of the unit were entered as a covariate. Two HLM analyses for each measure were conducted: (a) a full interactive model and (b) a reduced, main-effects only model. Individual performances on the post-B2 assessment and the pretest assessment were modeled at level 1, and instructional/assessment condition was modeled at level 2 (as part of differences at the teacher-level data). Difference between the interactive and main-effects model provides a test of the explanatory contribution of the interaction terms.

We analyzed data separately for each 1 of the 4 instructional units: *Number Sense* (a training unit), *Geometry, Measurement,* and *Equivalent Fractions*. We will not here discuss the data for the training unit.

For the *Geometry* unit, the main effect of GDA indicates that performance on the GDA assessment (with process-training) is significantly higher than on the FDA assessment (filler activity), $\gamma03=0.46$, $t(59)=3.16$, $p=.003$. Investigation of the main effect for instructional condition indicated that while there is a trend toward the triarchic condition being superior to the dynamic condition, this difference was not significant, $\gamma01=0.26$, $t(59)=1.49$, $p=.142$. There was, however, a significant difference between dynamic instruction and the standard control condition, $\gamma01=-0.78$, $t(59)=-3.80$, $p=.001$. Individuals in the dynamic instructional condition performed significantly better on the geometry post-B2 measure than did those in the standard instruction control condition. The non-significant interaction is plotted in Figure 4.

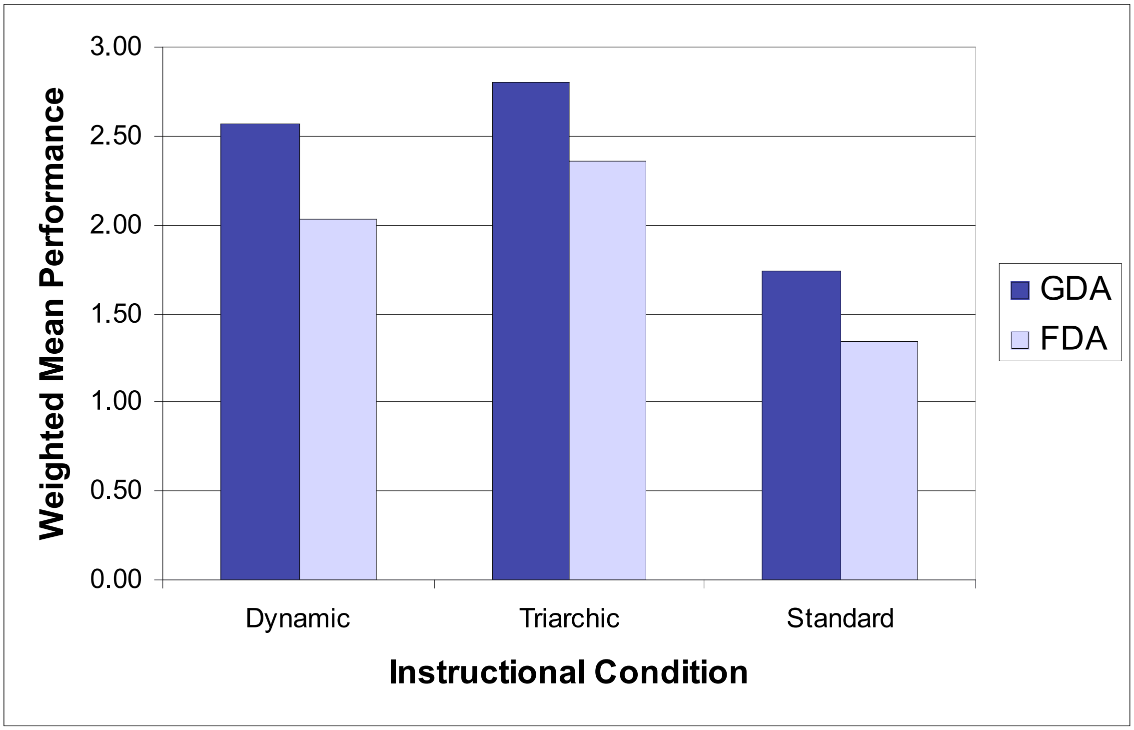

Figure 4. Weighted-mean performance on the post-B2 geometry test by instructional condition and assessment type.

For the *Measurement* unit, there were no participants in the Dynamic Instruction condition who were assessed with the FDA assessment (intervening filler activity); hence, the analyses are somewhat different from those reported for the *Geometry* and *Equivalent Fraction* units. The results indicated that although there is a clear trend toward the superiority of the Dynamic Instruction condition (with the GDA assessment) over the other conditions, this effect did not reach significance in the multi-level analysis[2]. Results indicate no significant differences between the instructional/assessment conditions (the test of the largest difference, between the DI-GDA and SI-GDA, was $\gamma03=-0.72$, $t(45)=-1.49$, $p=.143$). The estimated means (i.e., controlling for pretest score) are plotted in Figure 5.

[2] Ordinary-least-squares analyses indicate significant superiority of the Dynamic Instruction condition. See the full HLM analyses output for details.

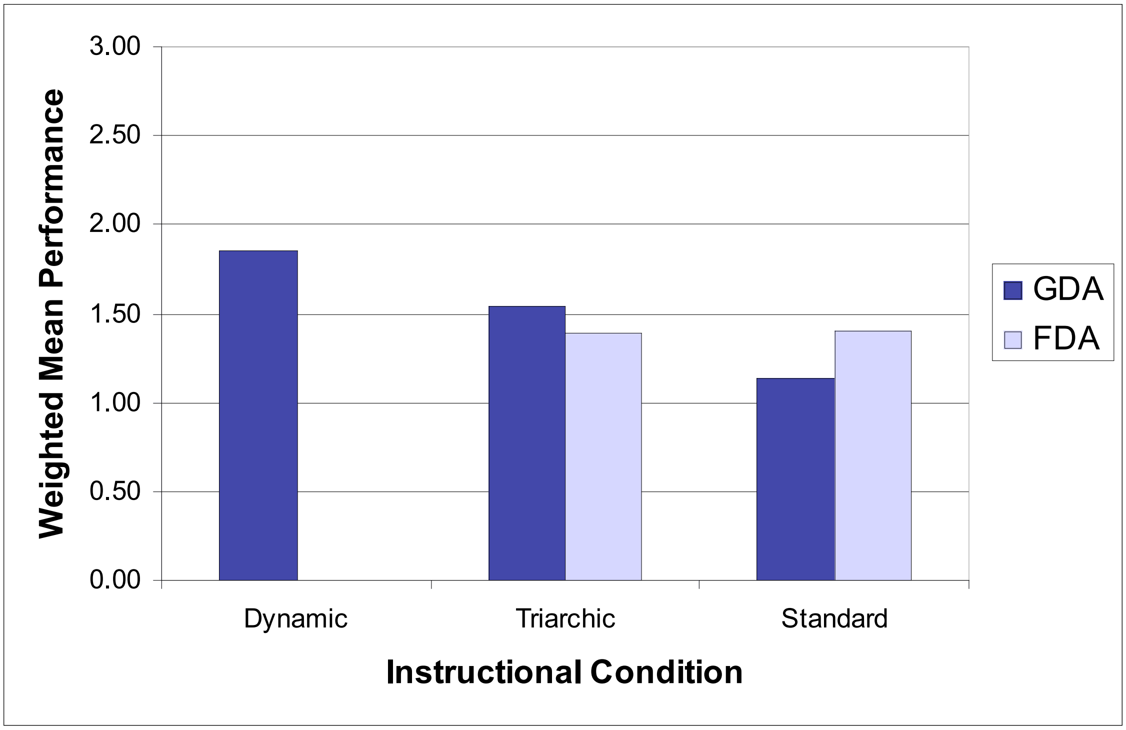

Figure 5. Weighted-mean performance on the post-B2 measurement test by instructional condition and assessment type (pretest score as a covariate).

Finally, for the *Equivalent Fractions* unit, the main effect of GDA was not significant, $\gamma 03=0.14$, $t(59)=1.07$, $p=.291$, indicating that while overall the trend is for the expected superiority of GDA assessment in the Dynamic and Standard conditions, there were no significant overall mean differences between the GDA and FDA assessments. Investigation of the main effect for instructional condition indicated that the dynamic instruction resulted in superior performance to both the triarchic, $\gamma 01=-0.31$, $t(59)=-2.18$, $p=.033$, and the standard control conditions, $\gamma 02=-0.33$, $t(59)=-2.71$, $p=.034$. That is, individuals in the dynamic instructional condition performed significantly better on the equivalent fractions post-B2 measure than those in the triarchic and the standard instruction control condition. The interaction is plotted in Figure 6.

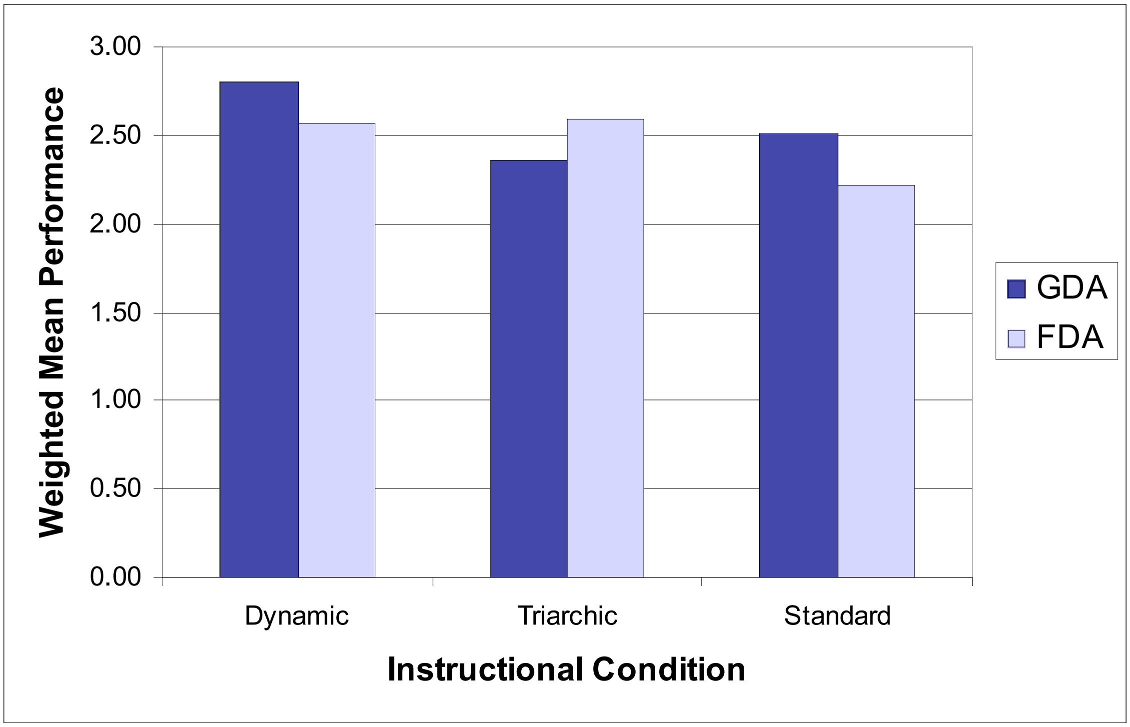

Figure 6. Weighted-mean performance on the post-B2 equivalent fractions test by instructional condition and assessment type (pretest score as a covariate).

Summary

Experimental Conditions

In study 1 (group-administered assessments, GDA), the comparisons across instructional units generated findings in support of our hypotheses. Specifically, after controlling for initial content abilities, there is an increasing advantage for the Dynamic + Triarchic instruction over the Triarchic-only instruction and standard instructional practices with time. In the first unit administered in this study (i.e., Geometry), the Triarchic-only instructional unit resulted in superior performance, followed closely by the Dynamic + Triarchic instruction. In the unit typically taught second (i.e., Measurement), there was a marginal advantage for the Dynamic + Triarchic instruction. In the unit typically taught last (Equivalent Fractions), the advantage for the Dynamic + Triarchic instruction is clear. First, it is further encouraging to us to find that infusion of Triarchic ideas into mathematics curriculum shows an advantage over standard instructional practices. Second, that the advantage of dynamic instruction takes some time to emerge is also consistent with evidence reported in the dynamic assessment literature. The dynamic assessment philosophy can at times require quite a fundamental change in the way instruction and assessment is conducted. For instance, from a purely pragmatic perspective, some lag due to resistance on both the student and teacher sides of the equation is likely to be expected. That we see the advantage of the Dynamic + Triarchic instruction within 3 units over a period of less than a semester, is encouraging to us for several reasons. First, much of the dynamic assessment philosophy has been focused on individuals which have been identified as disadvantaged in some way (e.g., low-SES, ethnic minority, or learning-disabled). While some of our participants do fall within such classifications, it is not the case for all the participants. Hence, we now have evidence to support what many have been arguing for some time, namely that dynamic instruction is advantageous to all students. Furthermore, we are in a position to argue that the day-to-day practical constraints of the classroom should not be used as an excuse to prevent the application of a combined dynamic instruction/assessment curricula. At least when dynamic assessment principles are paired with what could be argued to already be a rather dynamic mathematics curricula (i.e., the Triarchic units), positive outcomes across the board tend to be observed.

Group-administered Assessment

The second research question we were interested in has to do with the group-administered dynamic assessment approach more specifically. That is, we expected that the process-training between post-B1 and post-B2 (i.e., the GDA condition) would facilitate performance over and above the filler activity (i.e., FDA conditions), and that this advantage would be more pronounced for the Dynamic + Triarchic condition. We expected that the dynamic instruction would result in an advantage because the approach used during the process-training between post-B1 and post-B2 follows quite closely the dynamic instruction principles teachers used throughout the unit. That is, we argued that the reason for the GDA advantage was not simply due to teaching toward the test. However, the results suggested that while (a) performance in the dynamic instruction conditions was generally superior to the other conditions (certainly over time), and (b) the GDA (process-training) facilitated performance, this advantage was generally not more pronounced for the Dynamic + Triarchic condition. It would seem that at this stage, further research needs to investigate the group-administered assessment approach—more sophisticated analyses are being explored to tease apart some of the complexity surrounding this issue.

Differences Between Ethnic Groups

One of the strongest claims of dynamic assessment researchers is that minority achievement gaps can be reduced under dynamic assessment (Feuerstein et al., 1979, Sternberg & Grigorenko, 2002b). Recruitment in the current project targeted 4 major ethnic groups in the USA: White, African American, Hispanic American, and Asian American. The ethnicity hypothesis is that differences between ethnic groups will be more greatly reduced for DA (post-B2) relative to static assessment (post-B1), and that this effect will be greater when instruction is dynamic compared with triarchic-only, or standard control conditions.

In our data, the extent of the minority performance gap standard assessments is similar to what has been reported previously in the literature. The challenge of the dynamic instruction/assessment is to explore the extent that this disadvantage can be removed with training.

We next considered math post-study performance. The results indicate that even at post-study and after controlling for performance at baseline, when compared with White and Asian American groups, there is still a significant bias against the traditionally disadvantaged minority groups. Both African American and Hispanic American groups performed significantly poorer than the Caucasian group (Afr. Amer: $\gamma 20=-0.28$, $t(62)=-5.035$, $p<.001$; Asian American: $\gamma 30=0.10$, $t(62)=1.53$, $p=.132$; Hispanic American: $\gamma 40=-.17$, $t(62)=-2.52$, $p=.015$; Other: $\gamma 50=0.02$, $t(62)=0.37$, $p=.716$). However, this effect is mitigated by experimental condition. A second set of analyses explored the extent to which the minority gap is moderated by instructional condition, considering a dichotomous ethnicity variable (White and Asian American vs. African American and Hispanic American). This ensures a more sufficient sample size that would otherwise not be available if we continued to use the specific condition and ethnicity distinctions. The results, as plotted in Figure 7, indicate that, after controlling for baseline math achievement, the minority performance gap for the Dynamic Instruction condition (GDA and FDA) is smaller than for all other conditions. This trend was statistically significant for all but the Standard Instruction (SI-FDA) condition, DI-GDA vs. → DI-FDA: $\gamma 21=-0.02$, $t(57)=-0.26$, $p=.795$; TI-GDA: $\gamma 22=-0.24$, $t(57)=-2.80$, $p=.007$; TI-FDA: $\gamma 23=-0.30$, $t(57)=-3.07$, $p=.004$; SI-GDA: $\gamma 24=-0.40$, $t(57)=-2.07$, $p=.043$; SI-FDA: $\gamma 25=-0.25$, $t(57)=-1.38$, $p=.173$.

Finally, we analyzed unit posttest performance by ethnicity. Analyses to date suggest no significant minority bias that occurs differentially as a function of instructional condition. There is some indication that if a minority bias is going to occur, it does so in the second (B2) post- assessments (which is regarded as the main outcome of the dynamic assessment), rather than in first (B1) post-assessment (which is considered to be the static component of the assessment). That is, there is more variability in the dynamic assessment post-B2 component.

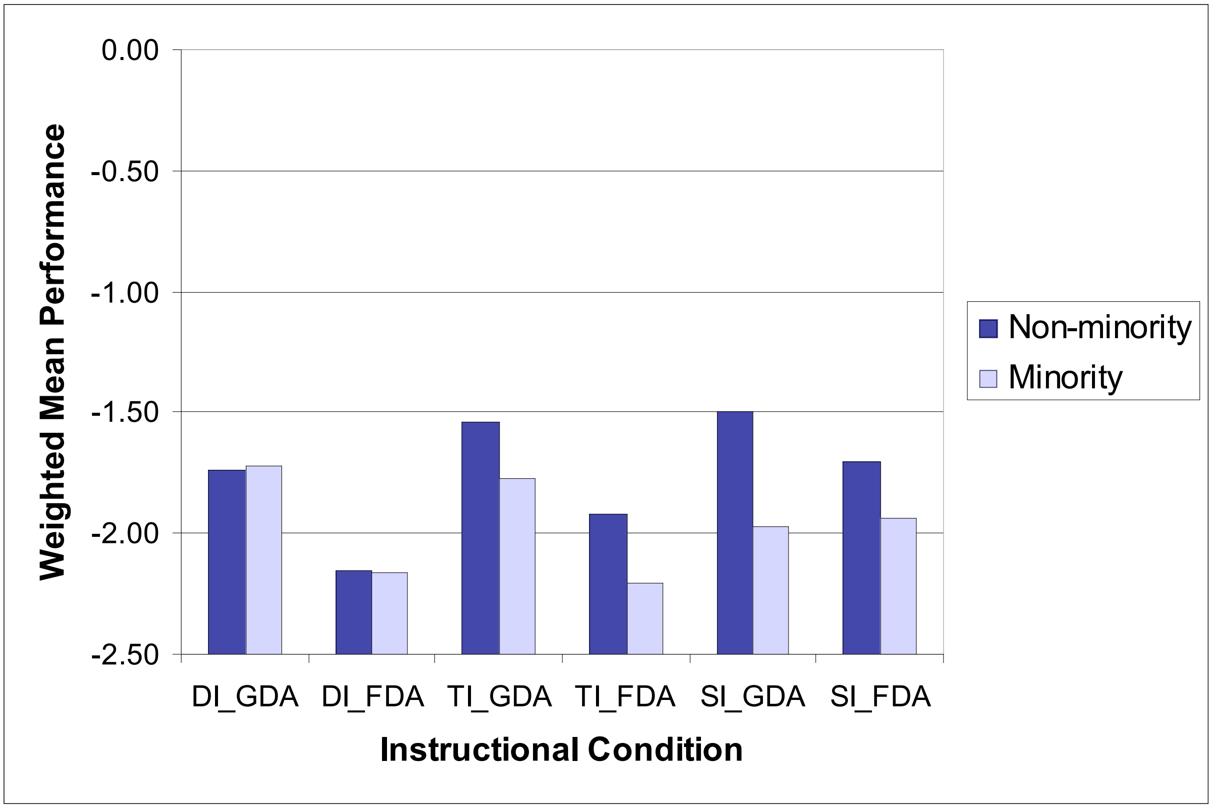

Figure 7. Weighted-mean performance scores across instructional condition for minority and non-minority students.

Conclusion

The study reported here represents an initial attempt to develop dynamic instruction and dynamic assessments to better gauge student knowledge and to further student learning. The following criteria have been adapted by many researchers in the field of educational and dynamic assessment to evaluate their methodologies: the underlying theory, variety of processes addressed in the learner, clear principles for examiner interaction, and clear links to instructional criteria, utility of obtained information (improved learner functioning in the classroom), inter-rater reliability, ease of infusing the methodology into everyday practice, and time- and cost-efficiency (Lidz, 1991). The present study yielded results that met all of the above criteria. The study developed very structured instructional and assessment materials that can be easily generalized to different schools. Heavy emphasis on structured (nearly manualized) approaches to instruction and assessment produced a very reliable and user-friendly methodology. While one of the common criticism of dynamic assessment is that it is very time consuming and subjective in nature, this study integrated instruction and assessment into one procedure in order to (a) make it more time-efficient, (b) foster the connection between instruction and assessment (particularly for individual assessments), and (c) link it to specific academic outcomes that serve as objective indicators of progress.

The data collected from participating students and teachers show that (a) it is possible to develop dynamic assessments that can be used to asses groups of and individual students in a regular classroom setting, (b) such dynamic assessments with a process oriented (rather than a filler) activity between post tests tends to lead to higher student achievement, and (c) dynamic instruction tends to reduce the achievement gap between minority and non-minority students. Previous attempts to reduce performance differences between majority and underrepresented minority students have not been altogether successful, so these results are promising as an avenue for further exploration. Triarchic dynamic instruction utilized in this study represents one type of differentiated adaptive instruction that teachers are strongly encouraged to practice by current legal mandates (e.g., No Child Behind, US Congress, 2002).

According to the National Standards for School Mathematics from the National Council of Teachers of Mathematics (NCTM, 2000) teachers need to provide students with real-life activities "that are based on significant and correct mathematics" and stress learning to reason mathematically, connect various math concepts, and communicate about mathematics. The NCTM urges teachers to provide learners with instruction and assessment that promotes equality and high expectations for all students. Research suggests that teachers are supportive of these standards and report a need for appropriate instructional tools to satisfy these standards and provide their students with appropriate educational experiences (e.g., Adams & Hsu, 1998; Shinn & Hubbard, 1993). However, the teachers also report little progress toward implementing process-oriented methods in classroom practices (Day & Cordon, 1993; Hirsch, 2005). There are gaps in knowledge of what and how to train teachers in using these practices and how to implement these assessment and instructional practices into classrooms. These gaps are even greater when examined in relation to math achievement in diverse learners (e.g., poor minority students). This study provides foundation for future solidification of dynamic methods to provide the teachers with tools and skills that can address the needs of diverse students. While group administered dynamic assessment may serve as an alternative to currently used static assessments to better gauge students' learning status and to link assessment with instruction more directly, individually administered assessment may provide a very useful blueprint for response-to-intervention and academic failure prevention practices in inclusive classrooms in contemporary schools.

The dynamic methods described and tested in the present study have the potential to be generalized to other content areas and student populations (e.g., science for English language learners in High School). Further research will help identify critical components of these methods and thus design a finely grained net for promoting academic competencies in students and fostering increased teaching competence in teachers. Correspondingly, we conclude that the combination of dynamic assessment and instruction is a promising educational practice, and these initial results warrant further exploration in other subject areas and at other grade levels.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234.

Adams, T. L., & Hsu, J. Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics, 4,* 174-181.

Alves-Martins, M., Peixoto, F., Gouveia-Pereira, M., Amaral, V., & Pedro, I. (2002). Self-esteem and academic achievement among adolescents. *Educational Psychology, 22,* 51-62.

Amabile, T. M. (1996). *Creativity in context*. Boulder, CO: Westview.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 181-215.

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

Beckman, J. F., & Guthke, J. (1995). Complex problem solving, intelligence status, and learning ability. In J. Funke & P. Frensch (Eds.), *Complex problem solving: The European perspective* (pp. 177–200). Hillsdale, NJ: Erlbaum.

Beckman, J. F. & Guthke, J. (1999). *Assessing reasoning ability*. Gottingen, Germany: Hogrefe.

Binet, A., & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique, 11,* 191-336.

Borman, G. D., & Overman, L. T. (2004). Academic resilience in mathematics among poor and minority students. *Elementary School Journal, 104,* 177-195.

Campione, J. C., Brown, A. L., & Bryant, N. (1985). Individual differences in learning and memory. In R. J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 103-126). New York: Freeman.

Carlson, J. S., & Wiedl, K. H. (1976). The factorial analysis of perceptual and abstract reasoning abilities in tests of concrete operational thought. *Educational and Psychological Measurement, 36*, 1015–1019.

Carlson, J. S., & Wiedl, K. H. (1978). Use of testing-the-limits procedures in the testing of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency, 11*, 559–564.

Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven Matrices. *Intelligence, 3*, 323-344.

Carlson, J. S., & Wiedl, K. H. (1992). The dynamic testing of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive Testing* (pp. 167-186). New York: Springer–Verlag.

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

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

Cattell, R. B. & Cattell, A. K. S. (2002). *Culture Fair Intelligence Tests*. Champaign, IL: Institute for Personality and Ability Testing.

Ceci, S. J., & Liker, J. (1986). Academic and nonacademic intelligence: An experimental separation. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 119-142). New York: Cambridge University Press.

Ceci, S. J., & Roazzi, A. (1994). The effects of context on cognition: Postcards from Brazil. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 74-101). New York: Cambridge University Press.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Cianciolo, A. T., & Sternberg, R. J. (2004). *A brief history of intelligence*. Malden, MA: Blackwell.

Cogan, J. C., Sternberg, R. J., & Subotnik, R. F. (2006). Integrating the other three Rs into the curriculum. In R. J. Sternberg & R. F. Subotnik (Eds.). O*ptimizing student success in schools with the other three R's: Reasoning, resilience, and responsibility* (pp. 227-238). Greenwich, CT: Information Age Publishing.

Davydov, V. (1988). Problems of developmental teaching: The experience of theoretical an experimental psychological research. *Soviet Education, 30*(8), 44-97.

Day, J. D., Engelhardt, J. L., Maxwell, S. E., & Bolig, E. E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. *Journal of Educational Psychology, 89*(2), 358-368.

Day, J. D. & Cordon, L. A. (1993). Static and dynamic measures of ability: An experimental comparison. *Journal of Educational Psychology, 85*, 75-82.

Delpit. L. (1995). *Other people's children: Cultural conflict in the classroom.* New York: New Press.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality and development*. Philadelphia: Psychology Press.

Elliott, J. (1993). Assisted testing: If it is "dynamic" why is it so rarely employed? *Educational and Child Psychology, 10,* 48-58.

Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.

Fagundes, D. D., Haynes, W. O., Haak, N. J., & Moran, M. J. (1998). Task variability effects on the language test performance of southern lower socioeconomic class African American and Caucasian five-year-olds. *Language, Speech, and Hearing Services in Schools, 29,* 148-157.

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device theory, instruments, and techniques*. Baltimore: University Park Press.

Feuerstein, R., Rand, Y., Falik, L., & Feuerstein, R. S. (2003). *Dynamic Assessment of Cognitive Modifiability*. Jerusalem, Israel: ICELP Publications.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171-191.

Flynn, J. R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 25-66). Washington, DC: American Psychological Association.

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

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.

Gordon, E. W., & Armour-Thomas, E. (1991). Culture and cognitive development. In R. J. Sternberg & L. Okagaki (Eds.), *Directors of development: Influences on the development of children's thinking*. Hillsdale, NJ: Erlbaum.

Gordon, E. W., & Meroe, A. S. (1991). Common destinies—continuing dilemmas. *Psychological Science, 2*(1), 23-24.

Gordon, E. W., & Song, L. D. (1994). Variations in the experience of resilience. In M. C. Wang & E. W. Gordon (Eds.), *Educational resilience in inner city America* (pp. 3-25). Mahwah, NJ: Erlbaum.

Gordon, E. W., & Wilkerson, D. (1996). *Compensatory education for the disadvantaged: Program and practices, preschool through college*. New York: College Entrance Examination Board.

Gordon, E. W. & Yowell, C. (1994). Educational reforms for students at risk: Cultural dissonance as a risk factor in the development of students. In R. J. Rossi (Ed.), *Educational reforms and students at risk* (pp. 51-69). New York: Teachers College Press.

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

Grigorenko, E. L., Birney, D. P., Jeltova, I., & Sternberg, R. J. (2002). *Dynamic assessment: When assessment and instruction are two sides of the same coin.* Paper presented at the Annual conference of the National Research Center on the Gifted and Talented, Washington, DC.

Grigorenko, E. L., & Sternberg, R. J. (1997). Styles of thinking, abilities and academic performance. *Exceptional Children, 63*, 295-312.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*, 75-111.

Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests? *European Journal of Psychological Assessment, 12*, 1-13.

Hattie, J., & Marsh, H. W. (1996). Future directions in self-concept research. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 421-462). New York: John Wiley.

Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snook, S., & Sternberg, R. J. (2003). Identifying and assessing tacit knowledge: Understanding the practical intelligence of military leaders. *Leadership Quarterly, 14,* 117-140.

Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: A supplement to the graduate management admissions test. *Learning and Individual Differences, 16,* 101-127.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.

Hirsh, S. (2005). Professional development and closing the achievement gap. *Theory Into Practice, 44*(1), 38-44.

Hoffman, R. R. (Ed.). (1992). *The psychology of expertise: Cognitive research and empirical AI.* New York: Springer-Verlag.

Horn, J. L. (1994). Fluid and crystallized intelligence, theory of. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 443-451). New York: Macmillan.

Jensen, M. R. (1998). *Mindladder projects: Student, school, family and community learning in the information age*. Unpublished manuscript, International Center for Mediated Learning, American InterContinental University.

Jitendra, A. K., & Kameenui, E. J. (1993). Dynamic testing as a compensatory testing approach: A description and analysis. *RASE: Remedial and Special Education, 14,* 6-18.

Kamhi, A., Pollock, K., & Harris, J. (1996). *Communication development and disorders in African American children*. Baltimore: Paul H. Brooks.

Laboratory of Comparative Human Cognition. (1982). Culture and intelligence. In R. J. Sternberg (Ed.*), Handbook of human intelligence* (pp. 642-719). New York: Cambridge University Press.

Lane, J., Lane, A. M., & Kyprianou, A. (2004). Self-efficacy, self-esteem and their impact on academic performance. *Social Behavior and Personality, 32,* 247-256.

Laughon, P. (1990). The dynamic testing of intelligence: A review of three approaches. *School Psychology Review, 19,* 459-470.

Lave, J. (1989). *Cognition in practice*. New York: Cambridge University Press.

Lidz, C. S. (Ed.). (1987). *Dynamic assessment: An interactional approach to evaluating learning potential*. New York: Guilford Press.

Lidz, C. S. (1991). *Practitioner's guide to dynamic testing*. New York: Guilford Press.

Lidz, C. S. (1997). Dynamic assessment approaches. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary approaches to assessment of intelligence* (pp. 281-296). New York: Guilford.

Lidz, C. S., & Elliott, J. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. Greenwich, CT: Elsevier-JAI.

McClelland, D. C. (1985). *Human motivation*. New York: Scott Foresman.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1976). *The achievement motive*. New York: Irvington.

Missiuna, C., & Samuels, M. T. (1988). Dynamic testing: Review and critique. *Special Services in the Schools, 5,* 1-22.

National Council of Teachers of Mathematics. (2000). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

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

Nuñes, T. (1994). Street intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 2, pp. 1045-1049). New York: Macmillan.

Okagaki, L., & Sternberg, R. J. (1993). Parental beliefs and children's school performance. *Child Development, 64,* 36-56.

Piaget, J. (1978). *Success and understanding*. Cambridge, MA: Harvard University Press.

Raven, J. C., Raven, J., & Court, H. C. (1995). *Mill Hill Vocabulary Scales*. San Antonio, TX: Hartcourt Assessment.

Renzulli, J. S., Reis, S. M., Hébert, T. P., & Diaz, E. I. (1995). The plight of high-ability students in urban schools. In M. C. Wang & M. C. Reynolds (Eds.), *Making a difference for students at risk* (pp. 41-98). Thousand Oaks, CA: Corwin Press.

Resing, W. C. M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential testing* (pp. 219-242). Amsterdam: Swets & Zeitlinger.

Rittle-Johnson, B., Siegler, R., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An Iterative process. *Journal of Educational Psychology, 93*(2), 346-362.

Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction, 14,* 373-408.

Rubinstein, S. L. (1946). *Osnovy obshchei psikhologii* [Foundation of general psychology]. Moscow: Uchpedgiz.

Schön, D. A. (1983). *The reflective practitioner*. New York: Basic Books.

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

Shinn, M. R., & Hubbard, D. D. (1993). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. In E. L. Meyen, G. A. Vergason, & R. J. Whelan (Eds.), *Challenges facing special education* (pp. 193-226). Denver, CO: Love.

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

Sternberg, R. J. (1986). *Intelligence applied.* Orlando, FL: Harcourt Brace College Publishers.

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

Sternberg, R. J. (1990). *Metaphors of mind.* New York: Cambridge University Press.

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

Sternberg, R. J. (1994b). Cognitive conceptions of expertise. *International Journal of Expert Systems: Research and Application, 7,* 1-12.

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

Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27,* 11-20.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology, 3,* 292-316.

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

Sternberg, R. J. (2003). Teaching for successful intelligence: Principles, practices, and outcomes. *Educational and Child Psychology, 20*(2), 6-18.

Sternberg, R. J. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology*, *39*(2), 189-202.

Sternberg, R. J. (2006). Reasoning, resilience, and responsibility from the standpoint of the WICS theory of higher mental processes. In R. J. Sternberg & R. F. Subotnik (Eds.), *Optimizing student success in schools with the other three R's: Reasoning, resilience, and responsibility* (pp. 17-37). Greenwich, CT: Information Age Publishing.

Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg triarchic abilities test in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment, 17*(1) 1-16.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P. R., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly, 40,* 129-137.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Snook, S. Williams, W. M. Wagner, R. K., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

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

Sternberg, R. J., & Grigorenko, E. L. (1999). *Our labeled children: What every parent and teacher needs to know about learning disabilities*. Reading, MA: Perseus Publishing Group.

Sternberg, R. J., & Grigorenko, E. L. (2000). *Teaching for successful intelligence*. Arlington Heights, IL: Skylight Training and Publishing.

Sternberg, R. J., & Grigorenko E. L. (Eds.). (2002a). *The general factor of intelligence: How general is it?* Mahwah, NJ: Erlbaum.

Sternberg, R. J., & Grigorenko, E. L. (2002b). *Dynamic testing*. New York: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (2003). Teaching for successful intelligence: Principles, procedures, and practices. *Journal for the Education of the Gifted, 27*(2), 207–228.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment, 15,* 1-11.

Sternberg, R. J., Grigorenko, E. L., Ngrosho, D., Tantufuye, E., Mbise, A., Nokes, C., Jukes, M., & Bundy, D. A. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30*, 141–162.

Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance 15*(1/2), 143-160.

Sternberg, R. J., & Horvath, J. A. (Eds.). (1999). *Tacit knowledge in professional practice*. Mahwah, NJ: Erlbaum.

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. New York: Free Press.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist, 51*(7), 677-688.

Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence, 29,* 1-18.

Sternberg, R. J., Okagaki, L., & Jackson, A. (1990). Practical intelligence for success in school. *Educational Leadership, 48,* 35-39.

Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence*, *34*(4), 321-350.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology, 27,* 195-232.

Sternberg, R. J., & Spear-Swerling, L. (1996). *Teaching for thinking*. Washington, DC: APA Books.

Sternberg, R. J., & Subotnik, R. F. (Eds.). (2006). *Optimizing student success in schools with the other three R's: Reasoning, resilience, and responsibility*. Greenwich, CT: Information Age Publishing.

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998a). Teaching for successful intelligence raises school achievement. *Phi Delta Kappan, 79,* 667-669.

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998b). Teaching triarchically improves school achievement. *Journal of Educational Psychology, 90,* 1-11.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacit 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,* 912-927.

Taylor, A. R. (1991). Social Competence and the early school transition: Risk and protective factors for African-American children. *Educational and Urban Society, 24,* 15-26.

Taylor, O., & Payne, K. (1983). Culturally valid testing: A proactive approach. *Topics in Language Disorders, 3,* 8-20.

Tharp, R. G. (1989). Psychocultural variables and constants: Effects on teaching and learning in schools. *American Psychologist, 44*(2), 349-59

Vaughn-Cooke, F. (1986). The challenge of assessing the language of nonmainstream speakers. In O. Taylor (Ed.), *Treatment of communication disorders in culturally and linguistically diverse populations* (pp. 23-48). San Diego, CA: College-Hill Press.

Wiedl, K. H., Guthke, J., & Wingenfeld, S. (1995). Dynamic assessment in Europe: Historical perspectives. In J. S. Carlson (Ed.), *Advances in cognition and educational practice* (Vol. 3, pp. 33-82). Greenwich, CT: JAI Press.

Williams, W. M., Blythe, T., White, N., Li, J., Gardner, H., & Sternberg, R. J. (2002). Practical intelligence for school: Developing metacognitive sources of achievement in adolescence. *Developmental Review, 22*(2), 162–210.

Appendix

The Learner's Profile Scoring Matrix

Appendix

The Learner's Profile Scoring Matrix

General Description:

The first column contains labels for levels of probing. Prompts 1 to 3 deal with factors unrelated to math content knowledge such as reading comprehension, attention, and familiarity with open-ended questions etc. Prompts 4 to 6 are designed to target *Conditional*, *Procedural*, and *Factual* levels of knowledge.

The second through fifth columns represent the four cognitive modalities. There are 5 items within each modality. Items increase in difficulty from left to right. Item 1 is the easiest and Item 5 is the hardest within each modality.

Item 3 is the starting (entry) item within each modality. It is also a diagnostic item for each modality. The items were developed according to the nationwide standards for fourth grade mathematics. Thus, this instrument is a criterion-referenced assessment tool.

Administration Instructions:

General: The administration procedure follows an **up/down** procedure. Start with the average item (item #3). If the student produces the correct answer **before or on probe 4** (i.e., at the level of Conditional Knowledge) then move up to the next item, otherwise move down to an easier item that has not been attempted.

1. Select one of the modalities (Analytical, Practical, Creative, Memory). The modalities can be administered in any order.

    a. Start administration on item #3 (the shaded item).
    b. If the child produces the correct answer on item #3 **before or on probe 4** (i.e., at the level of Conditional Knowledge), **move up** and administer item #4.
    c. If item #4 is **passed with 0-4 probes**, go to item #5. After item #5, STOP and go to another modality.
    d. If item #4 **requires more than 4** prompts, then **move down** to item #2 to establish the basal level of the child's performance. If item # 2 requires more than 4 probes, go to item #1.
    e. If on item #3 the child produces a correct answer **after 4 probes**, **move down** to item #2.
    f. If item #2 requires **0-3 probes**, go to item #4.
    g. If item #4 requires **0-3 probes**, go to item #5.
    h. If item #2 requires **more than 3 probes**, go to item **#**1.

2. Choose another modality and repeat the above steps **1a** to **1h**.

3. <u>Prompts</u>. If the child has the conditional knowledge to be able to solve the problem after the conditional prompt (4), circle prompt 4. If the child has the conditional knowledge but cannot solve the problem, cross (x) prompt 4 and administer prompt 5 (procedural prompt). If the child has the procedural knowledge but cannot produce the correct answer, then cross prompt 5 and administer factual prompt 6.

4. While administering the items, take notes on the child's preferred way of solving the problems (e.g., drawing vs. mental representation), the vocabulary he or she uses (e.g., abstract terms vs. common terms), the degree of reliance on context (e.g., spontaneous introduction of context into the problem). The aim is to be more able to identify the child's individual preferences.

5. Upon administering each modality, ask the child about his/her metacognitive skills taking notes on the reverse side of the scoring matrix (see *suggested* questions).

6. The results are available immediately upon testing. The child can be provided with general feedback about his/her unique profile of strengths and weaknesses across different modalities at the end of testing.

Prompt	A	P	C	M	Total
0					
1					
2					
3					
4					
5					
6					

Student Id Label___________________

Unit: ______________ Teacher: ________________ Date: _______________

*** In addition to any other comments you may have, include some notes on the following:**

modality/ item / prompt	ANALYTICAL Easy-----Ave-----Hard					PRACTICAL Easy-----Ave-----Hard					CREATIVE Easy-----Ave-----Hard					MEMORY Easy-----Ave-----Hard				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
0. No prompts	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1. Reread	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2. Id the Q	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3. Paraphrase	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4. Conditional knowledge	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5. Procedural knowledge	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6. Factual knowledge	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
NOTES* (see below)																				

Were there any specific problems the student seemed to really enjoy or really dislike?
Which questions seemed to be of most help to the student?
Was there a particular modality that seemed to be more difficult to the student?

Metacognitive skills. Questions to ask after each modality (complete on reverse side)
Were there any problems you enjoyed working on?
Did you find the questions I was asking helpful?
Can you tell me what was going on in your head as you worked on (this) problem?

Summary Table

Prompt	A	P	C	M	Total
0					
1					
2					
3					
4					
5					
6					

Enter the number of items the student solved at each one of the prompting levels for each modality.
Easier items not attempted should be counted as having been answered correctly with no (0) prompts

Metacognitive Skills (ask these or similar questions after administering each modality)

Questions	**ANALYTICAL**	**PRACTICAL**	**CREATIVE**	**MEMORY**
Were there any problems you enjoyed working on?				
Did you find the questions I was asking helpful?				
Can you tell me what was going on in your head as you worked on (this) problem?				

Research Monograph

The National Research Center on the Gifted and Talented
University of Connecticut
2131 Hillside Road Unit 3007
Storrs, CT 06269-3007
www.gifted.uconn.edu

Editor
E. Jean Gubbins

Production Assistant
Siamak Vahidi

Also of Interest

State Policies Regarding Education of the Gifted as Reflected in Legislation and Regulation
A. Harry Passow and Rose A. Rudnitski

Residential Schools of Mathematics and Science for Academically Talented Youth: An Analysis of Admission Programs
Fathi A. Jarwan and John F. Feldhusen

The Status of Programs for High Ability Students
Jeanne H. Purcell

Recognizing Talent: Cross-Case Study of Two High Potential Students With Cerebral Palsy
Colleen Willard-Holt

The Prism Metaphor: A New Paradigm for Reversing Underachievement
Susan M. Baum, Joseph S. Renzulli, and Thomas P. Hébert

Attention Deficit Disorders and Gifted Students: What Do We Really Know?
Felice Kaufmann, M. Layne Kalbfleisch, and F. Xavier Castellanos

Gifted African American Male College Students: A Phenomenological Study
Fred A. Bonner, II

Also of interest from the

Research Monograph Series

Counseling Gifted and Talented Students
Nicholas Colangelo

E. Paul Torrance: His Life, Accomplishments, and Legacy
Thomas P. Hébert, Bonnie Cramond, Kristie L. Speirs Neumeister, Garnet Millar, and Alice F. Silvian

The Effects of Grouping and Curricular Practices on Intermediate Students' Math Achievement
Carol L. Tieso

Developing the Talents and Abilities of Linguistically Gifted Bilingual Students: Guidelines for Developing Curriculum at the High School Level
Claudia Angelelli, Kerry Enright, and Guadalupe Valdés

Development of Differentiated Performance Assessment Tasks for Middle School Classrooms
Tonya R. Moon, Carolyn M. Callahan, Catherine M. Brighton, and Carol A. Tomlinson

Society's Role in Educating Gifted Students: The Role of Public Policy
James J. Gallagher

Middle School Classrooms: Teachers' Reported Practices and Student Perceptions
Tonya R. Moon, Carolyn M. Callahan, Carol A. Tomlinson, and Erin M. Miller

Assessing and Advocating for Gifted Students: Perspectives for School and Clinical Psychologists
Nancy M. Robinson

Giftedness and High School Dropouts: Personal, Family, and School Related Factors
Joseph S. Renzulli and Sunghee Park

Assessing Creativity: A Guide for Educators
Donald J. Treffinger, Grover C. Young, Edwin C. Selby, and Cindy Shepardson

Implementing a Professional Development Model Using Gifted Education Strategies With All Students
E. Jean Gubbins, Karen L. Westberg, Sally M. Reis, Susan T. Dinnocenti, Carol L. Tieso, Lisa M. Muller, Sunghee Park, Linda J. Emerick, Lori R. Maxfield, and Deborah E. Burns

Also of interest from the

Research Monograph Series

Teaching Thinking to Culturally Diverse, High Ability, High School Students: A Triarchic Approach
Deborah L. Coates, Tiffany Perkins, Peter Vietze, Mariolga Reyes Cruz, and Sin-Jae Park

Advanced Placement and International Baccalaureate Programs for Talented Students in American High Schools: A Focus on Science and Mathematics
Carolyn M. Callahan

The Law on Gifted Education
Perry A. Zirkel

School Characteristics Inventory: Investigation of a Quantitative Instrument for Measuring the Modifiability of School Contexts for Implementation of Educational Innovations
Tonya R. Moon, Catherine M. Brighton, Holly L. Hertberg, Carolyn M. Callahan, Carol A. Tomlinson, Andrea M. Esperat, and Erin M. Miller

Content-based Curriculum for Low Income and Minority Gifted Learners
Joyce VanTassel-Baska

Reading Instruction for Talented Readers: Case Studies Documenting Few Opportunities for Continuous Progress
Sally M. Reis, E. Jean Gubbins, Christine Briggs, Fredric J. Schreiber, Susannah Richards, Joan Jacobs, Rebecca D. Eckert, Joseph S. Renzulli, and Margaret Alexander

Issues and Practices in the Identification and Education of Gifted Students From Under-represented Groups
James H. Borland

The Social and Emotional Development of Gifted Students
Carolyn M. Callahan, Claudia J. Sowa, Kathleen M. May, Ellen Menaker Tomchin, Jonathan A. Plucker, Caroline M. Cunningham, and Wesley Taylor

Promoting Sustained Growth in the Representation of African Americans, Latinos, and Native Americans Among Top Students in the United States at All Levels of the Education System
L. Scott Miller

Also of interest from the

Research Monograph Series

Evaluation, Placement, and Progression: Three Sites of Concern for Student Achievement
Samuel R. Lucas

Latino Achievement: Identifying Models That Foster Success
Patricia Gándara

Modern Theories of Intelligence Applied to Assessment of Abilities, Instructional Design, and Knowledge-based Assessment
Robert J. Sternberg, Elena L. Grigorenko, Bruce Torff, and Linda Jarvin

Giftedness and Expertise
Robert J. Sternberg, Elena L. Grigorenko, and Michel Ferrari

Academic and Practical Intelligence
Robert J. Sternberg, Elena L. Grigorenko, Jerry Lipka, Elisa Meier, Gerald Mohatt, Evelyn Yanez, Tina Newman, and Sandra Wildfeuer

Developing Creativity in Gifted Children: The Central Importance of Motivation and Classroom Climate
Beth A. Hennessey

Intelligence Testing and Cultural Diversity: Concerns, Cautions, and Considerations
Donna Y. Ford

Equity, Excellence, and Economy in a System for Identifying Students in Gifted Education: A Guidebook
Joseph S. Renzulli

The Feasibility of High-end Learning in a Diverse Middle School
Catherine M. Brighton, Holly L. Hertberg, Tonya R. Moon, Carol A. Tomlinson, and Carolyn M. Callahan

The Law on Gifted Education (Revised Edition)
Perry A. Zirkel

Nurturing Talent in Underrepresented Students: A Study of the Meyerhoff Scholars Program at the University of Maryland, Baltimore County
Beatrice L. Bridglall and Edmund W. Gordon

Also of interest from the

Research Monograph Series

The Schoolwide Enrichment Model Reading Study
Sally M. Reis, Rebecca D. Eckert, Fredric J. Schreiber, Joan Jacobs, Christine Briggs, E. Jean Gubbins, Michael Coyne, and Lisa Muller

Identifying Academically Talented Minority Students
David F. Lohman

Teachers' Guide for the Explicit Teaching of Thinking Skills
Deborah E. Burns, Jann Leppien, Stuart Omdal, E. Jean Gubbins, Lisa Muller, and Siamak Vahidi

Multiple Case Studies of Teachers and Classrooms Successful in Supporting Academic Success of High Potential Low Economic Students of Color
Carol Ann Tomlinson, Holly Gould, Stephen Schroth, and Jane Jarvis

Advanced Placement and International Baccalaureate Programs: A 'Fit' for Gifted Learners?
Holly Hertberg-Davis, Carolyn M. Callahan, and Robin M. Kyburg

The Effects of Dynamic Pedagogy on the Mathematics Achievement of Ethnic Minority Students
Edmund W. Gordon and Eleanor Armour-Thomas

NRC G/T

The National Research Center on the Gifted and Talented Research Teams

University of Connecticut

Dr. Joseph S. Renzulli, Director
Dr. E. Jean Gubbins, Associate Director
Dr. Sally M. Reis, Associate Director
University of Connecticut
2131 Hillside Road Unit 3007
Storrs, CT 06269-3007
860-486-4676

Dr. Del Siegle

University of Virginia

Dr. Carolyn M. Callahan, Associate Director
Curry School of Education
University of Virginia
P.O. Box 400277
Charlottesville, VA 22904-4277
804-982-2849

Dr. Tonya Moon
Dr. Carol A. Tomlinson
Dr. Catherine M. Brighton
Dr. Holly L. Hertberg-Davis

Yale University

Dr. Elena L. Grigorenko, Associate Director
Yale University
Center for the Psychology of Abilities, Competencies, and Expertise
340 Edwards Street, P.O. Box 208358
New Haven, CT 06520-8358

Dr. Linda Jarvin