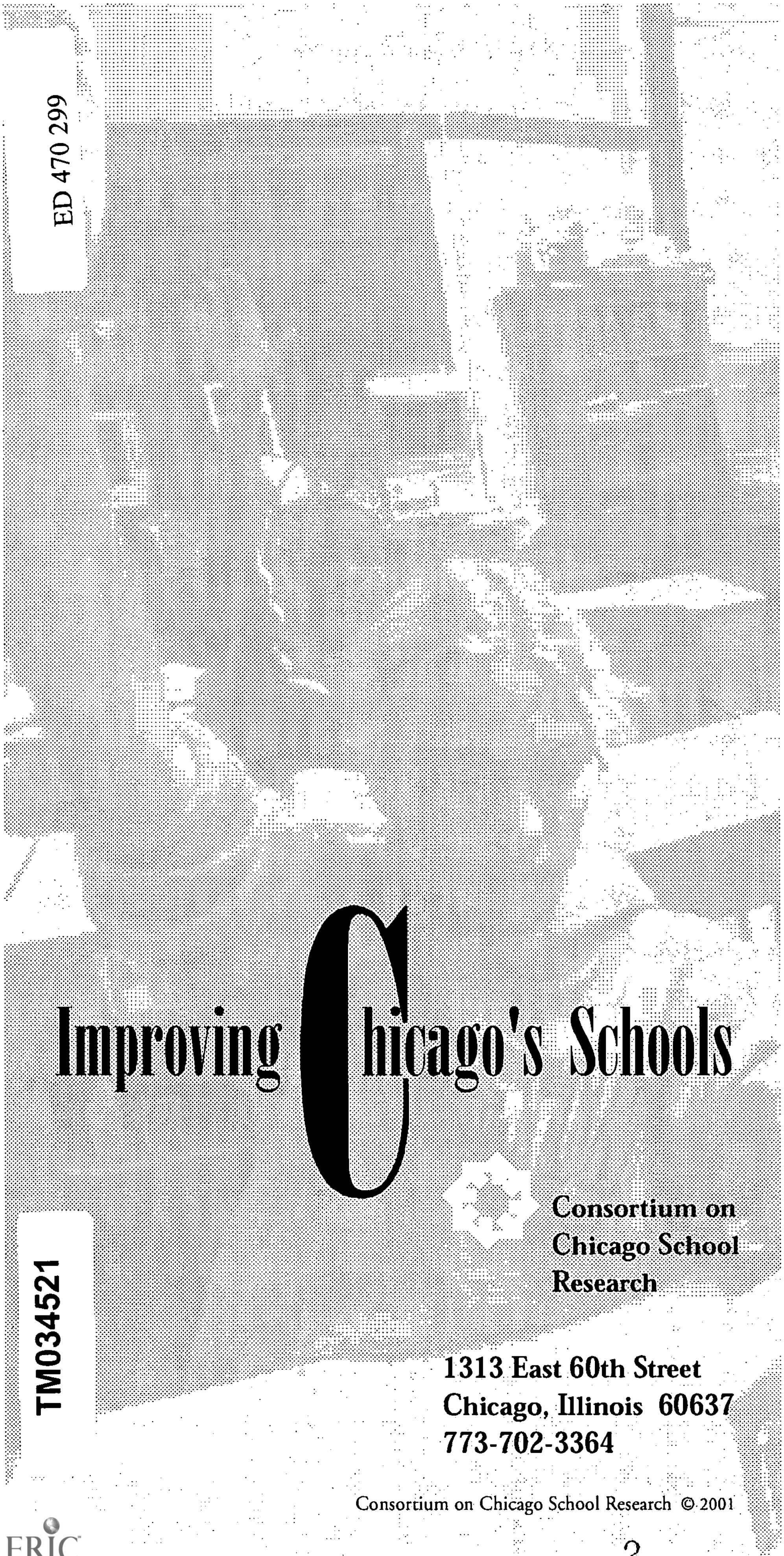
DOCUMENT RESUME

ED 470 299 TM 034 521

AUTHOR	Newmann, Fred M.; Bryk, Anthony S.; Nagaoka, Jenny K.
TITLE	Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence? Improving Chicago's Schools.
INSTITUTION	Consortium on Chicago School Research, IL.
PUB DATE	2001-01-00
NOTE	49p.; For a related report, see TM 034 517.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Assignments; *Difficulty Level; Disadvantaged Youth; Elementary Education; *Elementary School Students; Elementary School Teachers; *Mathematics Instruction; Urban Schools; *Writing Instruction
IDENTIFIERS	*Chicago Public Schools IL

ABSTRACT

This study of Chicago teachers' assignments in mathematics and writing in grades 3, 6, and 8 shows that students who received assignments requiring more challenging intellectual work also achieved greater than average gains on the Iowa Tests of Basic Skills in reading and mathematics, and demonstrated higher performance in reading, mathematics, and writing on the Illinois Goals Assessment Program. Data were from assignments collected in 1997, 1998, and 1999 with these sample sizes: (1) grade 3 writing, all 3 years, 1,785 assignments; (2) grade 6 writing, all 3 years, 1,686 assignments; (3) grade 8 writing, all 3 years, 1,425 assignments; (4) grade 3 mathematics, all 3 years, 1,794 assignments; (5) grade 6 mathematics, all 3 years, 1,522 assignments; and (6) grade 8 mathematics, all 3 years, 1,278 assignments. Assignments were scored by teams of teachers and scores were equated across years. Contrary to some expectations, the study found some high quality assignments in some very disadvantaged Chicago classrooms. It was evident that all students in these classes benefited from exposure to such instruction. Results suggest that if teachers, administrators, policymakers, and the public at large place more emphasis on authentic intellectual work in classrooms, yearly gains on standardized tests in Chicago could surpass national norms. Four appendixes contain details on study methodology. (Contains 5 figures, 24 endnotes, and 39 references.) (SLD)



Authentic Intellectual Work and Standardized Tests:

Conflict or Coexistence?

Improving Chicago's Schools

Fred M. Newmann
Anthony S. Bryk
Jenny K. Nagaoka

Consortium on Chicago School Research

1313 East 60th Street
Chicago, Illinois 60637
773-702-3364

January 2001

Executive Summary

Student success in contemporary society requires not only basic knowledge and skills but also the capacity to engage in more complex intellectual activity. Discussion of best instructional practices or forms of assessment, however, frequently poses a dichotomy between teaching approaches that enhance basic skills versus those that aim at more ambitious intellectual work, implying a trade-off between these two educational goals. The evidence presented here suggests that this debate rests on a false dichotomy.

Prior studies have documented that when teachers organize instruction around assignments that demand higher order thinking, in-depth understanding, elaborated communication and that make a connection to students' lives beyond school, students produce more intellectually complex work. This study of Chicago teachers' assignments in mathematics and writing in grades 3, 6, and 8, shows that students who received assignments requiring more challenging intellectual work also achieved greater than average gains on the Iowa Tests of Basic Skills in reading and mathematics, and demonstrated higher performance in reading, mathematics, and writing on Illinois Goals Assessment Program. Contrary to some expectations, we found high quality assignments in some very disadvantaged Chicago classrooms and that all students in these classes benefited from exposure to such instruction.

We conclude, therefore, assignments calling for more authentic intellectual work actually improve student scores on conventional tests. The results suggest that, if teachers, administrators, policymakers, and the public at-large place more emphasis on authentic intellectual work in classrooms, yearly gains on standardized tests in Chicago could surpass national norms.

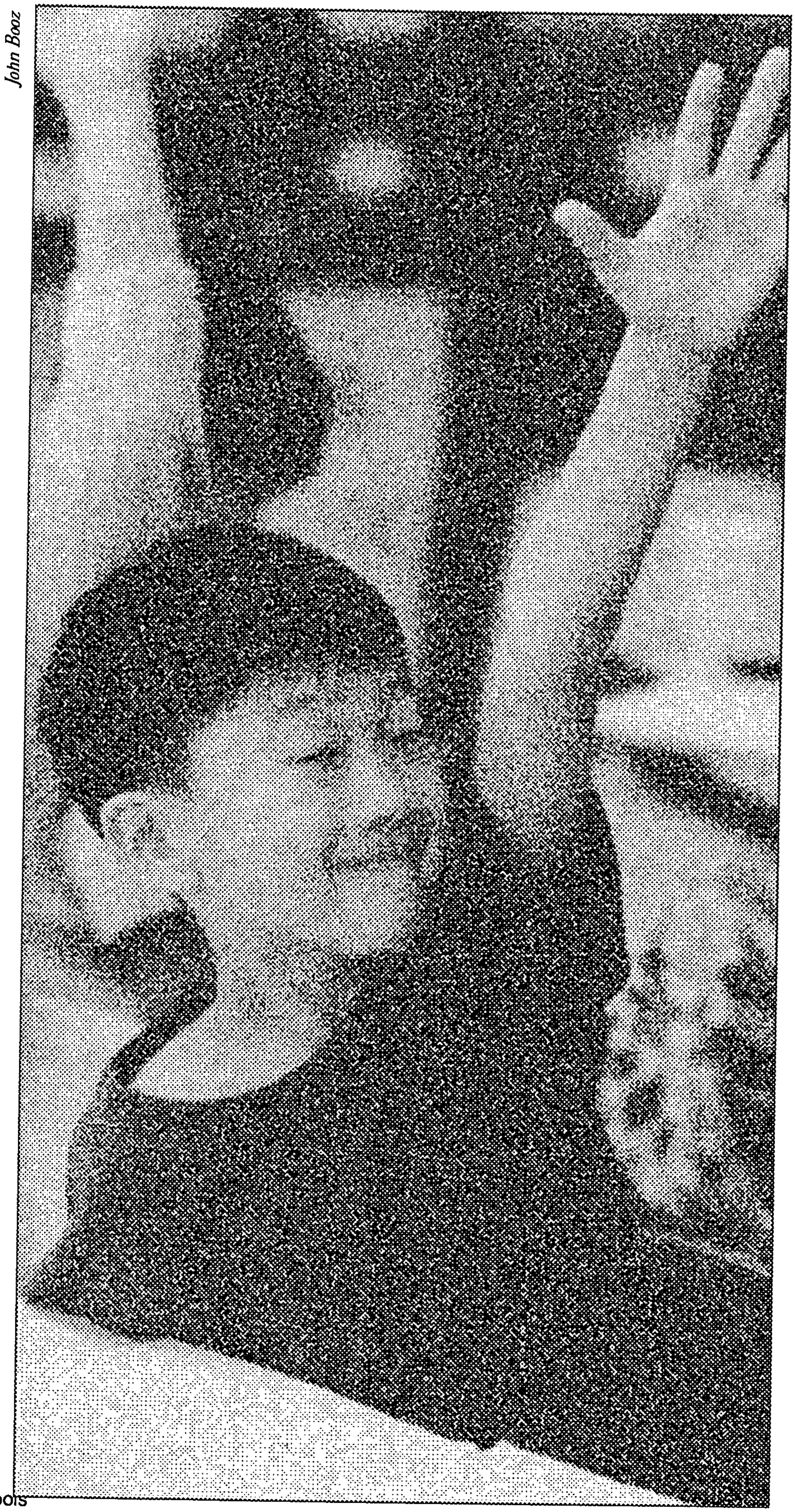
Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence?

Fred M. Newmann
University of Wisconsin

Anthony S. Bryk
University of Chicago

Jenny K. Nagaoka
Consortium on Chicago School Research

January 2001



John Booz

Foreword

In 1993, Ambassador Walter Annenberg announced a $500 million challenge grant to improve public education in the United States. Cities wishing to receive a portion of that grant were invited to submit proposals describing how the funds would be used to stimulate educational innovation and collaboration in their public school systems. A group of Chicago school reform activists and education stakeholders, including parents, teachers, principals, community leaders, and foundation officers, organized to write a proposal to include Chicago among the sites receiving a grant. They were successful. In January 1995, the Annenberg Foundation awarded a five-year grant of $49.2 million to establish the Chicago Annenberg Challenge. An additional $100 million in matching funds was pledged by local donors.

The Chicago Annenberg Challenge was organized to distribute and manage these monies among networks of schools and external partners throughout the city. Its mission is to improve student learning by supporting intensive efforts to reconnect schools to their communities, restructure education, and improve classroom teaching. The Chicago Challenge funds networks and external partners that seek to develop successful, community-based schools that address three critical education issues through whole-school change: school and teacher isolation, school size and personalism, and time for learning and improvement. More than half of Chicago's public schools will have participated in an Annenberg-supported improvement effort by the end of the grant period in 2001.

This report is part of a series of special topic reports developed by the Chicago Annenberg Research Project. This series focuses on key issues and problems of relevance to the Chicago Annenberg Challenge and to the improvement of Chicago public schools generally. It complements a series of technical reports that focus specifically on the work and accomplishments of the Chicago Annenberg Challenge. Among the topics examined to date in the special topics report series are the quality of intellectual work in Chicago elementary schools; social support, academic press, and their relationship to student achievement; and this report, *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence?*

The work of the Chicago Annenberg Research Project is intended to provide feedback and useful information to the Chicago Challenge and the schools and external partners who participate in its efforts to improve educational opportunities for Chicago's children and youth. This work is also intended to expand public discussion about the conditions of education in the Chicago Public Schools and the kinds of efforts needed to advance meaningful improvements. This effort to stimulate new avenues of discussion about urban school improvement is an important aspect of Ambassador Annenberg's challenge to engage the public more fully in school reform.

Acknowledgments

Many organizations and individuals contributed to this report. The authors are deeply grateful for their assistance. The Chicago Annenberg Challenge provided the primary funds for the study and for the development and distribution of the report.

Many Chicago Annenberg Research Project researchers worked to collect assignments from teachers in the Chicago Annenberg schools participating in this research. The fine cooperation of teachers and other staff members in these schools has been critical to the completion of this report. Field researchers collecting assignments and work at 19 elementary schools from 1997 to 1999 included the following faculty, graduate students, and independent researchers: Janet Bresden, Rita Brusca-Vega, Robert Casselman, JoAn Chun, Karen DeMoss, Ruanda Garth-McCullough, Tania Gutierrez, Sara Hallman, Rodney Harris, Patty Horsch, Deborah Johnston, Joe Kahne, Diane King, Nicolas Leon, Pauline Lipman, Gudelia Lopez, Anna Lowe, Bridgette McCullough, Fikile Mazibuko, Audra Millen, Jane Montes, Pam Nesselrodt, James O'Brien, Tamara Perry, Sarah Phillips, Seri Porter, Therese Quinn, BetsAnn Smith, Terry Stirling, Amy Weaver, Kim Williams, Kristin Williams, and David Yasutake.

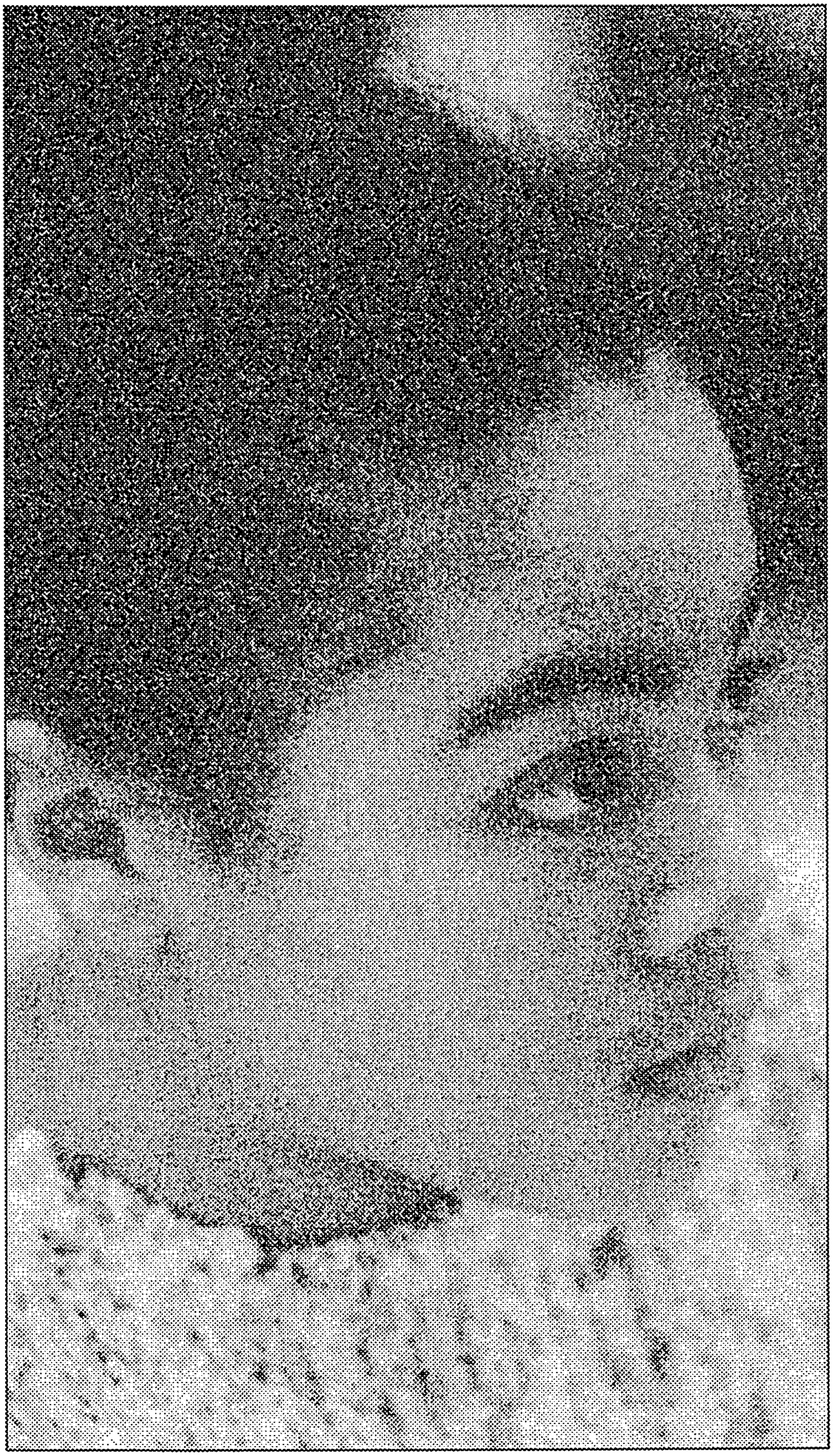
One hundred thirty-four teachers from the Chicago Public Schools worked hard during the summers of 1997, 1998, and 1999 to become trained in the scoring of teachers' assignments, and they contributed impressive professional expertise to complete the scoring activities. The teachers were trained to conduct the scoring by teams of authorities in mathematics education and the teaching of writing. The mathematics team, led by Eric Gutstein, consisted of Judy Merlau, Jean Biddulph, Joe Tillman, and Dan Miltner. The writing team, led by David Jolliffe, consisted of Carmen Manning, Kendra Sisserson, and Annie Knepler.

We would like to thank staff in the Office of Accountability at the Chicago Public Schools for their ongoing assistance and especially for providing the test score data essential for conducting this study. In particular, Joseph Hahn, Andrea Ross, Gudelia Lopez, and Sandra Storey helped us with this data.

Members of the Lead Team of the Chicago Annenberg Research Project gave important counsel during all phases of the report, from planning through publication. Core administrative staff supported the collection of the teacher assignments and managed logistics of handling the vast amounts of collected documents and the data generated when the documents were scored. Special thanks goes to Loretta Morris who has provided leadership, skill, and attention to detail without which this research would not be possible. Other professional staff members supporting the management of this process include Sabrina Billings, Karen DeMoss, Verity Elston, Nicolas Leon, Gudelia Lopez, Tamara Perry, and BetsAnn Smith.

Sandra Jennings, Sarah-Kay McDonald, and Carolyn Saper provided design, production, and editorial services. The photographs are by John Booz.

John Booz



I. Introduction

Since the early 1980s, a diverse array of national commissions, elected officials, and professional leaders have warned the American public time and time again that students' low performance in basic skills poses a serious threat to the economic future of our children and the country at large. Increasingly, parents, employers, higher education institutions, and policy makers are demanding that elementary and secondary schools do a better job of teaching basic skills to all students. For years, students' scores on standardized tests of basic skills carried few, if any, serious consequences for school administrators, teachers, or students. But now, students who fail to achieve minimum test cutoff scores can be required to attend summer school or receive other remedial instruction, and they may even be retained in their former grade until they pass these tests. Consistently low scores in the school can lead to the state or district taking over a school and possibly closing the school and then reopening it with a new program and staff (sometimes referred to as "reconstitution"). On the positive side, schools that show exceptional improvement in test scores may be rewarded with financial bonuses, and in some cases these bonuses go directly to teachers.[1]

The increasingly serious consequences for low student performance on standardized tests has been coupled with renewed attention to questions about how best to organize classroom instruction. Within the "back-to-basics" movement, it is widely believed that more sustained attention to didactic methods is essential.[2] From this perspective, the best way to teach is to present students with the desired information and ask them to memorize it, whether this be facts, definitions, algorithms, vocabulary lists, rules of communication, procedures (such as how to conduct an experiment or make a graph), and so on. Through various drills, exercises, and tests, students are expected to recall and repeat what they have memorized. This kind of mental work, while perhaps meticulous, is not especially complex, because the student is simply asked to reproduce material in the same form in which it was learned. The expected answers are brief, and usually only one answer is acceptable as correct.

Many educators have decried this "back-to-basics" movement. They argue that it endorses ineffective teaching methods and that its educational aims are too limited. Critics of didactic methods claim that instructional strategies dominant in the twentieth century will not serve our children well in the twenty-first century. Instead, they argue that instruction must be reorganized around assignments that make much more complex intellectual demands on students. Students should be expected to interpret and synthesize information, show relationships between various kinds of information, explain why some answers are better than others, and solve unfamiliar problems that might have more than one plausible solution. Proponents here believe that the best way to teach basic skills is to embed practice of these skills in the completion of more complex intellectual work.

These arguments raise special concerns in urban contexts where students' basic skills test scores have historically been very low.[3] Many urban school reformers contend that, while the vision of intellectually complex instruction may be attractive as an ultimate goal, it is not appropriate as a top priority for students in highly disadvantaged urban schools. These leaders worry that if teachers attempt more complex intellectual assignments in their classrooms, many disadvantaged students will never master the basics that they so desperately need to succeed on the standardized tests, which are increasingly important as gatekeepers of success in our society. To ensure that basic skills are mastered, they argue that students must concentrate mainly on straightforward memorization, drill, and repetitive practice—activities central to "didactic" instruction.

This report presents substantial new evidence on the issue raised by these opposing perspectives: *What happens to students' scores on standardized tests of basic skills when urban teachers in disadvantaged schools assign work that demands complex thinking and elaborated communication about issues important in students' lives?* We report results from a three-year study (1997-1999) of teaching and learning in more than 400 Chicago classrooms from 19 different elementary schools. We analyzed the intellectual demands of more than 2,000 classroom assignments in writing and mathematics from these schools and then linked these demands to learning gains on standardized tests in reading, mathematics, and writing for almost 5,000 Chicago students. This is the largest single body of evidence ever collected in disadvantaged urban schools on this question (see the sidebar "What We Know From Prior Research").

Examining Instruction: The Importance of Intellectual Demands

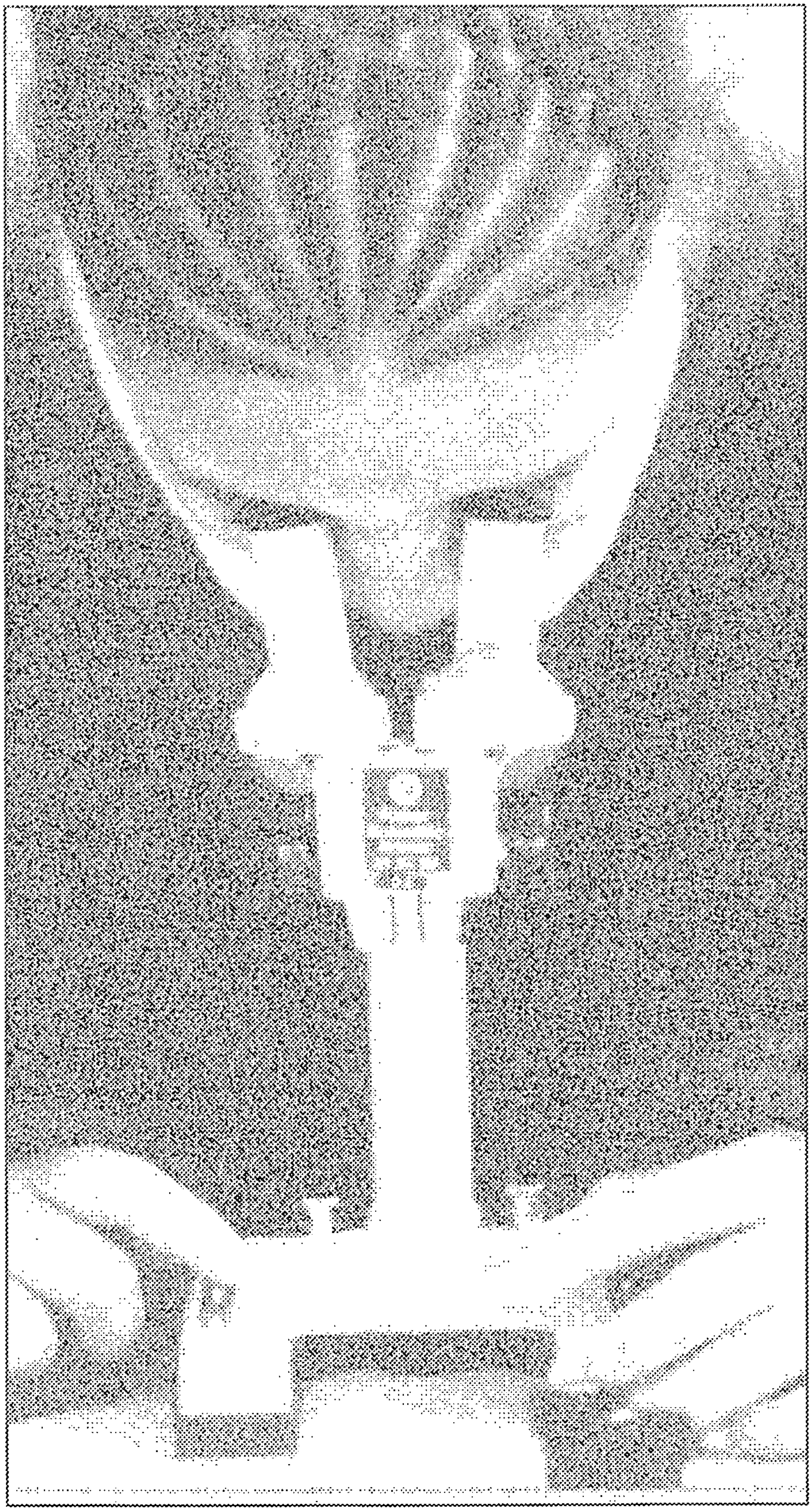
Both researchers and practitioners take different perspectives on the two approaches to instruction described above.[4] Comparisons between these approaches help emphasize the different assumptions made about the nature of knowledge, the roles of teacher and students in the classroom, and the behaviors of teachers and students. Our research focuses on the actual intellectual demands made on students. With didactic instruction, students are expected to reproduce information they have learned in the form of short answers to questions specified by teachers or on tests. Students are expected to learn facts, definitions, algorithms, and conventions of communication (e.g., punctuation) that have been presented in text or lecture, and to restate them in the same form they were learned. In didactic assessments, students may be asked to apply the knowledge they have memorized to new situations (e.g., using a word in a new sentence or solving a computation problem not previously encountered), but the assessments call for short, unelaborated answers, and the main criterion for mastery is whether the answer conforms to what the teacher has predetermined to be correct. In contrast, in interactive instruction, students are often asked to formulate problems, to organize their knowledge and experiences in new ways to solve them, to test their ideas with other students, and to express themselves using elaborated statements, both orally and in writing. For interactive assessments, in addition to correct answers, teachers also evaluate whether students used a sensible process to arrive at the answer. Such assignments challenge students to produce reasoning or justification for answers that may or may not have been anticipated by the teacher when the assignment was given.

Advocates of didactic instruction tend to support the goals of interactive instruction in the long run, but claim that before students are asked to apply, interpret, analyze, and integrate knowledge, they should first learn the basics as represented by questions on standardized tests. Advocates of interactive instruction generally agree that basic skills are important to learn, but they contend that didactic instruction is ineffective in teaching these skills. Moreover, they worry that an exclusive focus on didactic instruction of basic skills undermines efforts to promote more complex thinking and understanding. Supporters of interactive instruction contend that the basics can be learned well from interactive instruction and that interactive instruction also teaches to yield more complex understanding.[5]

What We Know From Prior Research

The limited available evidence from prior studies suggests that students exposed to teaching that demands complex intellectual work are likely to do as well as or better than students exposed to basic-skills-only instruction. These studies include research on the teaching of mathematics, reading, and writing to disadvantaged students (Knapp, Shields, and Turnbull 1992; D' Agostino 1996; Lee, Smith, and Newmann 2000), teaching mathematics in grades one, two, and eight (Carpenter, Fennema, Peterson, Chiang, and Loef 1989; Cobb, Wood, Yackel, Nicholls, Wheatley, Trigatti, and Perlwitz 1991; Silver and Lane 1995), teaching reading in grades one, two, and three (Tharp 1982), teaching mathematics and science in high school (Lee, Smith, and Croninger 1997), and teaching social studies in high school (Levin, Newmann, and Oliver 1969). Still, prior research has not adequately addressed the question we raise here. One limitation is that research relevant to the issue has tended to focus more on specific teaching practices and techniques such as class discussion versus teacher lecture, or cooperative learning activities versus individual seatwork, than on the intellectual demands embedded in classroom assignments. Second, even studies that have examined effects of intellectual demands in the classroom on standardized test scores have not included large numbers of students across different grade levels and subjects.

John Booz



II. Examining Classroom Assignments in Chicago: A Framework of Authentic Intellectual Work

The Consortium on Chicago School Research began a major longitudinal study in 1996, supported by the Chicago Annenberg Challenge to (1) examine the nature of instruction in Chicago classrooms; (2) document whether it is improving over time; and (3) analyze the key factors contributing to improvements where they have occurred. A central focus of this work has been the intellectual demands embedded in the classroom assignments that teachers ask students to complete. To guide our analysis of the intellectual demands embedded in classroom instruction, we adopted the framework of "authentic intellectual work" originally developed through federally funded national research by the Center on Organization and Restructuring of Schools.[6] The rationale and key ideas set forth in this framework were detailed in an earlier Consortium report (Newmann, Lopez, and Bryk, 1998). For those unfamiliar with this earlier work, we provide a brief review below.

Using Our Minds Well in Contemporary Society

Students must learn basic skills, that is, essential knowledge, procedures, and conventions in writing, speaking, and computing to participate successfully in contemporary society. But contemporary demands of productive work, responsible citizenship, and successful management of personal affairs also extend well beyond giving correct answers and following proper procedures for the work traditionally assigned in school. What are these additional intellectual demands?[7]

We considered the kinds of mastery demonstrated by successful adults who work with knowledge, such as scientists, musicians, childcare workers, construction contractors, health care providers, business entrepreneurs, repair technicians, teachers, lobbyists, and citizen activists. We do not expect children to achieve the same level of mastery as accomplished by adults in these roles, but identifying the kind of intellectual work in these professions can suggest some general standards for intellectual performance necessary for student success in contemporary society.

Consider, for example, an engineer designing a bridge. To complete the bridge design successfully, the engineer relies on extensive factual knowledge from engineering, architecture, science, and mathematics. But the particular context for the bridge, such as its length, height, peak points of stress and load, and the impact of local variation in weather conditions, require the engineer to organize, analyze, and interpret all this background information to make a unique product. Consider also a citizen trying to make an informed decision about whether an elected officeholder has done a good enough job to be reelected over the challengers, or trying to make a convincing public statement to increase local funding for school security. Finally, consider a single mother of pre-school children who calculates the costs and benefits of working, paying for childcare, and deciding how to choose among childcare providers. Adults in these diverse endeavors of work, citizenship, and personal affairs face a set of intellectual challenges that differ from those that students commonly experience in schools.[8] These intellectual challenges can serve as guidelines for education that extends beyond the basics to more complex intellectual work.

Adults in . . . diverse endeavors of work, citizenship, and personal affairs face a set of intellectual challenges that differ from those that students commonly experience in schools.

Criteria for Authentic Intellectual Work

Compared to the work of students in school, which often seems contrived and superficial, the intellectual accomplishments of adults in diverse fields seem more meaningful and significant. As a shorthand for describing the difference between the intellectual accomplishment of skilled adults and the typical work that students do in school, we refer to the more complex adult accomplishments as "authentic" intellectual work. "Authentic" is used here not to suggest that conventional work by students is unimportant to them and their teachers, or that basic skills and proficiencies should be devalued, but only to identify some kinds of intellectual work as more complex and socially or personally meaningful than others. More specifically, authentic intellectual work involves original application of knowledge and skills, rather than just routine use of facts and procedures. It also entails disciplined inquiry into the details of a particular problem and results in a product or presentation that has meaning or value beyond success in school. We summarize these distinctive characteristics of authentic intellectual work as **construction of knowledge**, through the use of **disciplined inquiry**, to produce discourse, products, or performances that have **value beyond school**.

Construction of Knowledge

Skilled adults working in various occupations and participating in civic life face the challenge of applying basic skills and knowledge to complex problems that are often novel or unique. To reach an adequate solution to new problems, the competent adult has to "construct" knowledge, because these problems cannot be solved by routine use of information or skills previously learned. Such construction of knowledge involves organizing, interpreting, evaluating, or synthesizing prior knowledge to solve new problems.

Disciplined Inquiry

Construction of knowledge alone is not enough. The mere fact that someone has constructed, rather than reproduced, a solution to a problem is no guarantee that the solution is adequate or valid. Authentic adult intellectual accomplishments require that construction of knowledge be guided by disciplined inquiry. By this we mean that they (1) use a prior knowledge base; (2) strive for in-depth understanding rather than superficial awareness; and (3) express their ideas and findings with elaborated communication.

- **Prior knowledge base.** Significant intellectual accomplishments build on prior knowledge that has been accumulated in a field. Students must acquire the knowledge base of facts, vocabularies, concepts, theories, algorithms, and other conventions necessary for their inquiry. This is usually the central focus of direct instruction in basic skills.
- **In-depth understanding.** A knowledge base of value to students involves more than being familiar with a broad survey of topics. Knowledge becomes most powerful when students can use information to gain deeper understanding of specific problems. Such understanding develops as one looks for, imagines, proposes, and tests relationships among key concepts in order to clarify a specific problem or issue.
- **Elaborated communication.** Accomplished adults working across a range of fields rely upon complex forms of communication both to conduct their work and to present their results. The tools they use—verbal, symbolic, and visual—provide qualifications, nuances, elaborations, details, and analogies woven into extended narratives, explanations, justifications, and dialogue.

Value Beyond School

This third criterion indicates that significant intellectual accomplishments have utilitarian, aesthetic, or personal value. When adults write letters, news articles, organizational memos, or technical reports; when they speak a foreign language; when they design a house, negotiate an agreement, or devise a budget; when they create a painting or a piece of music—they try to communicate ideas that have an impact on others. In contrast, most school assignments, such as spelling quizzes, laboratory exercises, or typical final exams, have little value beyond school, because they are designed only to document the competence of the learner.[9]

The three criteria—construction of knowledge, through disciplined inquiry, to produce discourse, products, and performances that have meaning beyond success in school—afford a foundation of standards for the more complex intellectual work necessary for success in contemporary society. All three criteria are important. For example, students might confront a complex calculus problem demanding much analytic thought (construction of knowledge and disciplined inquiry), but if its solution has no interest or value beyond proving competence to pass a course, students are less likely be able to use the knowledge in their lives beyond school. Or a student might be asked to write a letter to the editor about a proposed welfare policy. She might say she vigorously opposes the policy but offer no arguments indicating that she understands relevant economic and moral issues. This activity may meet the criteria of constructing knowledge to produce discourse with value beyond school, but it would fall short on the criterion of disciplined inquiry, and thereby possibly represent only superficial awareness, not deep understanding, of the issue.

> ***Knowledge becomes most powerful when students can use information to gain deeper understanding of specific problems.***

How We Did the Study

Sample of Schools, Classrooms, and Students

To assess the quality of intellectual work in Chicago public elementary schools, the Chicago Annenberg Research Project (CARP) has been collecting samples of classroom assignments. These assignments represent the primary artifacts of instruction—what Chicago teachers actually ask their students to do. We initiated this work in 1997 with 12 elementary schools, expanding the sample to 18 schools in 1998, and 19 schools in 1999. A two-stage plan was used to select schools. First, we identified a subsample of Annenberg networks to study. We clustered all funded Annneberg networks together in terms of their basic change strategy (e.g., community groups working on increased parental involvement, university-based groups emphasizing professional development, cultural institutions engaged in curriculum partnerships with schools). We then selected a subsample of networks from each cluster. Next at stage two, we chose a sample of two schools from within each network, one of which appeared especially promising in terms of its capacity for improvement and the other less so.[1] We also tried to assure that the final sample represented a diverse cross section of CPS schools and students. Table 1 provides a demographic description of the sample.

In general, CARP sample schools are slightly more disadvantaged that the overall school system. The percent of students at or above national norms in CARP schools at the onset of our study in 1997 was somewhat lower than the CPS average, and the percent of low-income students was a bit higher. Nonetheless, the CARP sample still represents a good cross section of schools in the CPS.

Demographics of Chicago Annenberg Challenge Research Project (CARP) Field Sites Compared with All Schools Participating in the Chicago Annenberg Challenge (CAC) and all Chicago public elementary schools (CPS)

	CARP Field Sample	All CAC Schools	CPS
Percent at or above national average - Reading (1997)	24.5	28.3	29.9
Percent at or above national average - Math (1997)	31.0	34.9	37.1
Percent African-American	53.0	59.8	58.3
Percent Latino	38.9	29.4	28.3
Percent low income	89.4	87.6	84.7

We collected assignments from classrooms in grades three, six, and eight. These grades were selected because at the outset of the study these were the target grades for the statewide Illinois Goals Assessment Program (IGAP). This allowed us to link teacher assignments both to student performance on state tests of reading, writing, and mathematics and to results from the national norm-referenced tests of reading and mathematics used by the CPS, The Iowa Test of Basic Skills (ITBS), in these same grades. The first three years of the study examined the performance of about 1,800 third graders, 1,700 sixth graders, and 1,400 eighth graders in classrooms of participating teachers.

Two teachers at grades three, six, and eight in each participating school were asked to submit both "typical" and "challenging" assignments in writing and mathematics.[2] Project researchers explained to participating teachers that the challenging assignments should be those that the teacher considered to provide the best indicators of how well students understood the subject at a high level. In contrast, typical assignments should reflect the daily work occurring in the course of a regular school week. Teachers were asked to provide four typical assignments spread throughout the year and two challenging assignments per year for a total of six assignments.[3] Incomplete data were received from some classrooms. For the analysis reported here, we examined a total of 349 assignments from 74 teachers in 1997, 953 assignments from 116 teachers in 1998, and 715 assignments from 87 teachers in 1999.

Assessing the Intellectual Demands in Classroom Assignments

Each summer following the school year that assignments were collected, we recruited a group of teachers from other Chicago public schools to assess the authenticity of intellectual work demanded by the assignments. We selected and trained six separate teacher groups, one for each subject matter-grade level combination. Teachers rated each assignment using the three standards of intellectual challenge. The standards for construction of knowledge and connection to students' lives were scored on a three-point scale and elaborated communication on a four-point scale. Assignments were randomly assigned to scorers one standard at a time. As a result, three different raters typically reviewed each assignment, one for each standard. In order to control for potential rater bias, a second rater independently scored a random sample of assignments. A special design was established for this double scoring in order to accurately assess and adjust each assignment score for the differential effects associated with individual raters.[4]

Using Many-Facets Rasch Analysis (MFRA), we constructed overall measures for the intellectual quality of all assignments based on the scores on the three standards, with separate scales for each grade and subject. The MFRA statistically adjusted the original scores for differences in the severity of raters and the difficulty associated with achieving each different rating on each standard.[5]

To ensure measure consistency across years, a sample of assignments collected in 1997 and 1998 were rescored in 1999.[6] This allowed us to equate measures across years, adjusting for differences over time in the relative severity of scorers and standards. All results reported here are in terms of the scales developed from the scoring of the assignments collected in 1998-99.[7]

Number of Students Per Grade in the Study

	Writing			Math		
	Grade 3	Grade 6	Grade 8	Grade 3	Grade 6	Grade 8
1997	482	578	402	515	435	365
1998	784	622	658	761	631	575
1999	519	486	365	518	456	338
Total	**1785**	**1686**	**1425**	**1794**	**1522**	**1278**

Standardized Tests

We used two different standardized tests to assess basic skills: the Iowa Tests of Basic Skills (ITBS), which is administered by the Chicago Public Schools to measure student achievement; and the Illinois Goal Assessment Program (IGAP) test, which is administered by the Illinois State Board of Education.

The ITBS is the primary data source used by the Chicago Public Schools on student achievement in mathematics and reading. It is a nationally norm-referenced test administered in May of each school year to students in first through eighth grades. Since the CPS changes the form of the ITBS used each year, the Consortium on Chicago School Research undertook a cross-form and cross-level equating to establish a common metric, in logits, for any trend analyses.[8] The results reported here are in this logit metric.

The Illinois Goal Assessment Program (IGAP) test is also a norm-referenced assessment that was used in the state of Illinois through 1998 to test mathematics, reading, and writing in grades three, six, and eight. The IGAP test, administered in February/March of each year, reports reading and mathematics on a 0 to 500 scale (national norm originally set at 250 with a standard deviation of 75) and the writing assessment on a 6 to 32 scale (no normative comparison available here). The IGAP mathematics test consists of multiple-choice items similar to the ITBS. The IGAP reading tests consists of two long reading passages each followed by a series of questions with multiple answers. Each question, however, may have more than one correct answer and the students must indicate whether each possible answer is correct or not. The IGAP writing test consists of an on-demand writing prompt that is scored by trained observers using a standard writing rubric.

In 1999, Illinois introduced a new standards-based assessment program, Illinois Standards Achievement Tests (ISAT), and changed target grades to three, five, and eight. Since no equating of ISAT and IGAP test scores exist, only the 1997 and 1998 IGAP data are used in this report.

---

[1] The data collected here were part of a larger Annenberg Research Project. For further details about the sample selection plan for schools, see Smylie et al. (forthcoming).

[2] Many elementary teachers in Chicago at grades three and six teach both language arts and math. In such situations, the same teacher provided both writing and math assignments. In contrast, departmentalization is more common at grade eight. As a result, mathematics and writing assignments in grade eight were more likely to come from different teachers.

[3] Data collection in 1997 occurred only during the spring semester. We asked teachers that year for two challenging and two typical assignments for a total of four assignments.

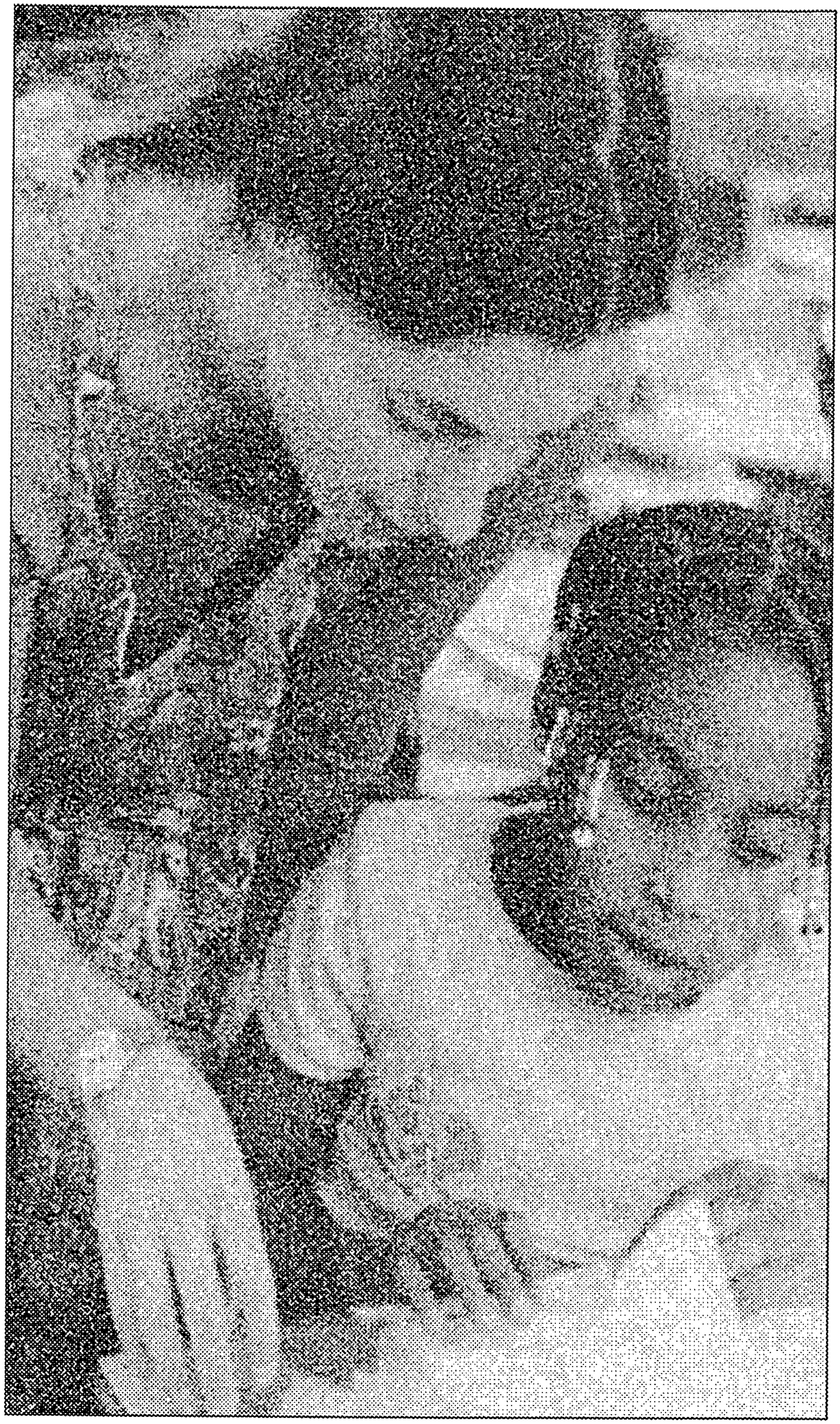
[4] Working in grade level teams for each subject (writing and mathematics) and scoring one standard at a time, each assignment was initially scored on that standard by one of the teachers in the team. Then each year at least 60 percent of the assignments were scored independently (i.e., without knowledge of the initial score) on one or more standards by a different teacher in the team. The assignments that were double scored were selected randomly from the pool of assignments, then assigned for double scoring through a systematic process in which each teacher in the team was paired with every other teacher to double score at least one of the standards.

[5] The Many-Facet Rasch model used here is: $\log (P_{nijk} / P_{nijk-1}) = B_n - D_i - C_j - F_k$ where P_{nijk} is the probability of assignment n being given a rating of k on standard i by judge j, P_{nijk-1} is the probability of assignment n being given a rating of k-1 on standard i by judge j, B_n is the intellectual challenge of assignment n, D_i is the difficulty of standard i, Cj is the severity of judge j and F_k is the difficulty of receiving a score in category k than a score of k-1. Thus, the final measure of the intellectual challenge in any assignment aggregates the score across all three standards, adjusted for the difficulty of each standard and the relative severity of the scorers. As in other item response theory applications, the final measure of assignment challenge exists in a logit metric. For reporting purposes, we converted them to a 0 to 10 point scale.

[6] The need to control for cross-year drift in assignment scoring became apparent in preliminary analyses of the 1998 data. We noticed some large differences in scores from the previous year, which raised concerns about the accuracy of the scoring across years. We conducted a test in which expert raters (i.e., the authorities in mathematics and writing who trained the teachers to do the scoring) scored a sample of assignments from 1997 and 1998 that had already been scored by the teachers. For this rescoring, we created a sample of paired assignments in which the 1997 and 1998 assignment had received identical scores from the teachers. As the expert raters scored these assignments they had no knowledge of the scores previously assigned. The activity identified a clear upward drift in scoring from 1997 to 1998. On the basis of these results, we developed the cross-year rescoring described in footnote 7 below.

[7] The cross-year rescoring design allowed us to estimate an adjustment for all the 1997 and 1998 assignments based on the difference between the original score and the 1999 rescores. This adjustment was then added to the scores for all of the 1997 and 1998 assignments to put them on the same scale as the 1999 assignments. Specifically, we calculated a Tukey's bi-weighted mean for the difference between the original scores and rescores for each year, subject, and grade. This bi-weighted mean was chosen as the adjustment statistic because it is a robust statistic that down weights the influence of extreme outliers in the data. This seemed most appropriate, given that our difference statistic involved a number of extreme values (i.e., the distribution had "fat tails.") For a further discussion of this statistic, see Mosteller and Tukey (1977). For a description of the actual trends in assignment scores from 1997 to 1999, see Bryk, Nagaoka, and Newmann (2000).

[8] For further details about the rationale and design of this test equating, see Bryk et al. (1998).


John Booz

III. Methods and Results

To assess the extent to which a particular assignment demanded authentic intellectual work, we translated the general criteria discussed above into more specific standards. Authorities in the fields of mathematics and writing created the standards in Figure 1 for assessing the extent to which any assignment called for authentic intellectual work.[10] Each of these standards was translated in turn into scoring rubrics. Figure 2 illustrates a rubric for one of the six standards.[11]

Each summer following the school year that assignments were collected, we recruited and trained a group of Chicago Public Schools teachers to apply these rubrics to the assignments that we had just collected. Assignments were randomly assigned to raters, one standard at a time. As a result, three different raters typically reviewed each assignment. In addition, a second set of raters independently scored a random subsample of assignments in order to control for potential rater bias. We employed a psychometric method, called Many-Facet Rasch Analysis (MFRA), to construct an overall measure of the intellectual quality of each assignment. This procedure allowed us to take into account, and adjust for, any observed differences among the raters in how they scored comparable assignments. (See the sidebar on pages 12-14 for a more detailed description of how we did the study, and the sidebar on page 20 to see examples of high and low scoring sixth grade assignments.)

Effects of Assignment Quality on Students' Basic Skills Learning

For each classroom in the study, we also drew individual student test score data from the Consortium's data archive. We included results

Figure 1

Standards for Assignments in Writing and Mathematics

A. Writing

Standard 1. Construction of Knowledge
The assignment asks students to interpret, analyze, synthesize, or evaluate information in writing about a topic, rather than merely to reproduce information.

Standard 2. Disciplined Inquiry: Elaborated Written Communication
The assignment asks students to draw conclusions or make generalizations or arguments and support them through extended writing.

Standard 3. Value Beyond School: Connection to Students' Lives
The assignment asks students to connect the topic to experiences, observations, feelings, or situations significant in their lives.

B. Mathematics

Standard 1. Construction of Knowledge
The assignment asks students to organize and interpret information in addressing a mathematical concept, problem, or issue.

Standard 2. Disciplined Inquiry: Written Mathematical Communication
The assignment asks students to elaborate on their understanding, explanations, or conclusions through extended writing; for example, by explaining a solution path through prose, tables, equations, or diagrams.

Standard 3. Value Beyond School: Connection to Students' Lives
The assignment asks students to address a concept, problem or issue that is similar to one that they have encountered or are likely to encounter in daily life outside of school.

from both the Iowa Test of Basic Skills (ITBS, a nationally norm referenced assessment administered annually by the CPS), and on the Illinois Goals Assessment Program (IGAP), a state test administered annually in elementary schools in grades three, six, and eight. In separate analyses, we estimated the effects of assignment quality on students learning of reading and mathematics, as measured by the ITBS, and on learning of reading, mathematics, and writing, as measured by the IGAP tests. For the reading and writing outcomes, we considered the intellectual demands embedded in classroom writing assignments. For the mathematics outcomes, we focused on the intellectual challenge in the mathematics assignments in students' classrooms. Our analyses took into account differences among classrooms in students' test scores from the prior year and controlled for differences in racial composition, gender, and socioeconomic status.[12] (Appendix A provides further technical details about how we developed an aggregate measure of the assignment quality in each classroom. Appendix B provides further details about the analysis model.)

Figure 3 displays the one-year learning gains on the ITBS for classrooms with challenging assignments as compared to national norms and as compared to Chicago classrooms where the intellectual quality of assignments were low.[13] We found a consistent positive relationship between student exposure to high-quality intellectual assignments and students' learning gains on the ITBS. Even after controlling for race, socio-economic class, gender, and prior achievement

Figure 2

Sample Scoring Rules for Writing Assignments: Elaborated Written Communication

4 = Explicit call for generalization AND support. The assignment asks students, using narrative or expository writing, to draw conclusions or to make generalizations or arguments, AND to substantiate them with examples, summaries, illustrations, details, or reasons.

3 = Call for generalization OR support. The assignment asks students, using narrative or expository writing, either to draw conclusions or make generalization or arguments, OR to offer examples, summaries, illustrations, details, or reasons, but not both.

2 = Short-answer exercises. The assignment or its parts can be answered with only one or two sentences, clauses, or phrasal fragments that complete a thought.

1 = Fill-in-the-blank or multiple-choice exercises.

Example: Grade 3 writing
High-scoring assignment
Elaborated Written Communication

In this assignment, third-grade students were asked to draw conclusions about how to show caring, and to substantiate them with reasons. They were asked to complete extended writing (i.e., at least six sentences) on this topic and to make sure that adequate support was included (i.e., it made sense to an adult). This assignment also scored high on the standards of Construction of Knowledge and Connection to Students' Lives.

"Write an essay on 'Showing Someone You Care.' Brainstorm words you might use in the essay. Write those words. Use these words in an essay. Write at least 6 sentences. Have a beginning, middle and a conclusion. Indent the first sentence of each paragraph. Spell words correctly, capitalize the first word in each sentence, have finger spaces between words, use correct ending marks, and write neatly. Re-read your essay after completing it, make corrections, re-read to an adult to make sure it makes sense.

Include in your essay whom you care about, give reasons why you care about them and what you can do to show them that you care."

differences among classrooms, the benefit of exposure to assignments that demand authentic intellectual work in writing and mathematics are quite substantial. In Chicago classrooms with high-quality assignments, students' record learning gains were 20 percent greater than the national average.[14] In contrast, Chicago classrooms where assignment quality reflects less demand, students gained 25 percent less than the national average in reading and 22 percent less in mathematics.[15]

The results for IGAP tests in reading, mathematics, and writing are presented in Figure 4. Since students take these tests only at grades three, six, and eight, we could not compute yearly gains in scores. Instead, we computed an adjusted mean outcome for each classroom on the IGAP reading, mathematics, and writing tests after controlling for differences among classrooms in student demographics and their prior year ITBS test scores in reading and mathematics. Thus, the results presented here can be interpreted as the "value-added" to student learning from a single year of classroom exposure to high-quality versus low-quality instructional assignments.[16]

To make these findings concrete, we consider two students who shared identical background characteristics, attended the same school, and had the same

Examples of Sixth Grade Assignments

High Scoring Writing Assignment

Write a paper persuading someone to do something. Pick any topic that you feel strongly about, convince the reader to agree with your belief, and convince the reader to take a specific action on this belief.

Commentary

In this high scoring assignment, demands for construction of knowledge are evident because students have to select information and organize it into convincing arguments. By asking students to convince others to believe and act in a certain way, the task entails strong demands that students support their views with reasons or other evidence, which calls for elaborated written communication. Finally, the intellectual challenge is connected to students' lives because they are to write on something they consider to be personally important.

Low Scoring Writing Assignment

Identify the part of speech of each underlined word below. All eight parts of speech—nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions, and interjections—are included in this exercise.

1) My room is arranged for comfort and efficiency.
2) As you enter, you will find a wooden table on the left.
3) I write and type.
4) There is a book shelf near the table.
5) On this book shelf, I keep both my pencils and paper supplies.
6) I spend many hours in this room.
7) I often read or write there during the evening.

Commentary

This assignment requires no construction of knowledge or elaborated communication, and does not pose a question or problem clearly connected to students' lives. Instead it asks students to recall one-word responses, based on memorization or definitions of parts of speech.

High Scoring Mathematics Assignment

Pick a stock. You have $10,000 to invest. Calculate how many shares you can buy at the current price. Every week for the next 10 weeks you will check in the newspaper whether your stock has gone up or down. You will chart the progress of your stock on the bulletin board. The chart is organized to record prices in 1/4 points, but the newspaper reports the prices in 1/16 points, so you will need to convert.

At the end of the 10 weeks, determine whether you have made a profit or loss and predict what you think your stock will do based on the results of the chart. Decide whether you will buy more or sell your stock. At this point you will give an oral report on your stock, what happened to it, and what you decided to do.

Commentary

This assignment scores high on construction of knowledge because each week students must decide how to represent the current price of the stock on a chart different from that which appears in the newspaper. They also have the opportunity to draw conclusions about their current profits or losses, which involves deciding what numbers to add, subtract, and multiply in order to compute them. The assignment's demands for charting the stock requires some written mathematical communication, and one would assume that preparation for the oral report would entail some elaborated mathematical communication. By focusing on mathematics related to a stock that students choose to "own," the assignment draws connections to mathematics and students' lives beyond school.

Low Scoring Mathematics Assignment

Name ______________________

Adding and Subtracting Fractions and Mixed Numbers: Common Denominators

Add or subtract. Reduce if possible.

1. $\frac{2}{3} + \frac{1}{3}$	**2.** $\frac{3}{8} + \frac{1}{8}$	**3.** $\frac{3}{10} + \frac{1}{10}$	**4.** $\frac{1}{4} + \frac{3}{4}$
5. $\frac{7}{8} - \frac{5}{8}$	**6.** $\frac{4}{5} - \frac{2}{5}$	**7.** $\frac{9}{10} - \frac{3}{10}$	**8.** $\frac{5}{6} - \frac{3}{6}$

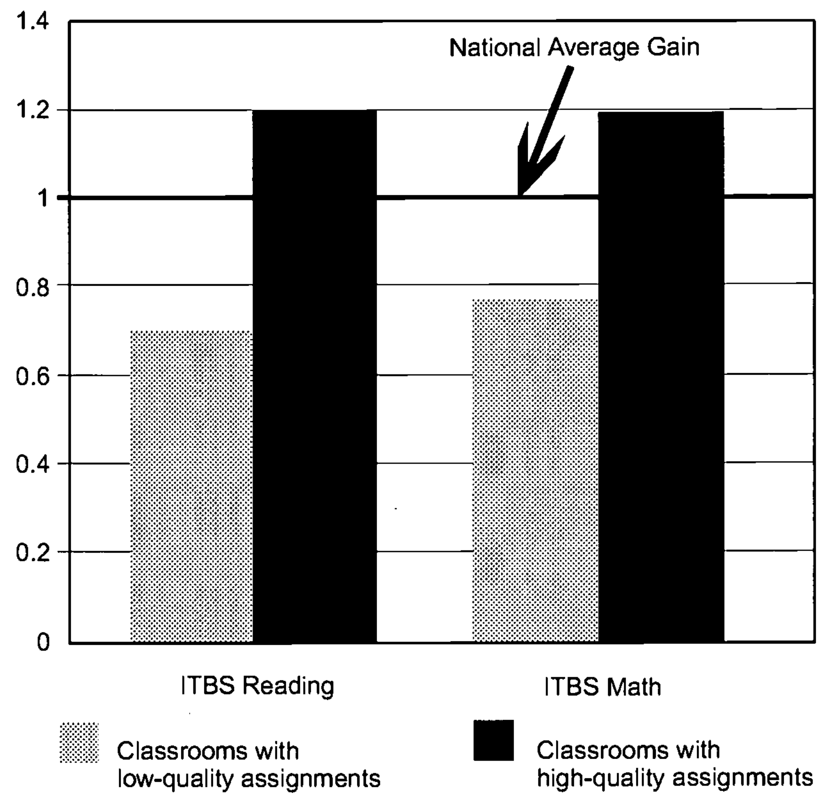
Commentary

This assignment requires no construction of knowledge to address a mathematical problem, no extended writing to explain mathematical conclusions, and it does not pose a mathematical problem connected to students' lives. Instead, it asks students only to fill in numerical answers to problems on addition, subtraction, and reduction of fractions based on memorized algorithms.

ITBS test scores from the prior year. Student A was assigned to a classroom that presented assignments demanding high levels of authentic intellectual work in both writing and mathematics. Student B, in contrast, attended a classroom where both writing and math assignments were weak. In the IGAP test assessments the following spring, student A would, according to our analyses, outperform his or her schoolmate, student B, by 32 points on the IGAP reading test and by 48 points on the IGAP math test. He or she would also be predicted to score 2.3 points higher on the IGAP writing rubric. These differences translate into standard effect sizes of 0.43, 0.64, and 0.52, respectively. In both substantive and statistical terms, these effects are quite large.[17 18]

Figure 3

Students Learn More in Classrooms with High Quality Assignments: Effects on One-Year ITBS Gains



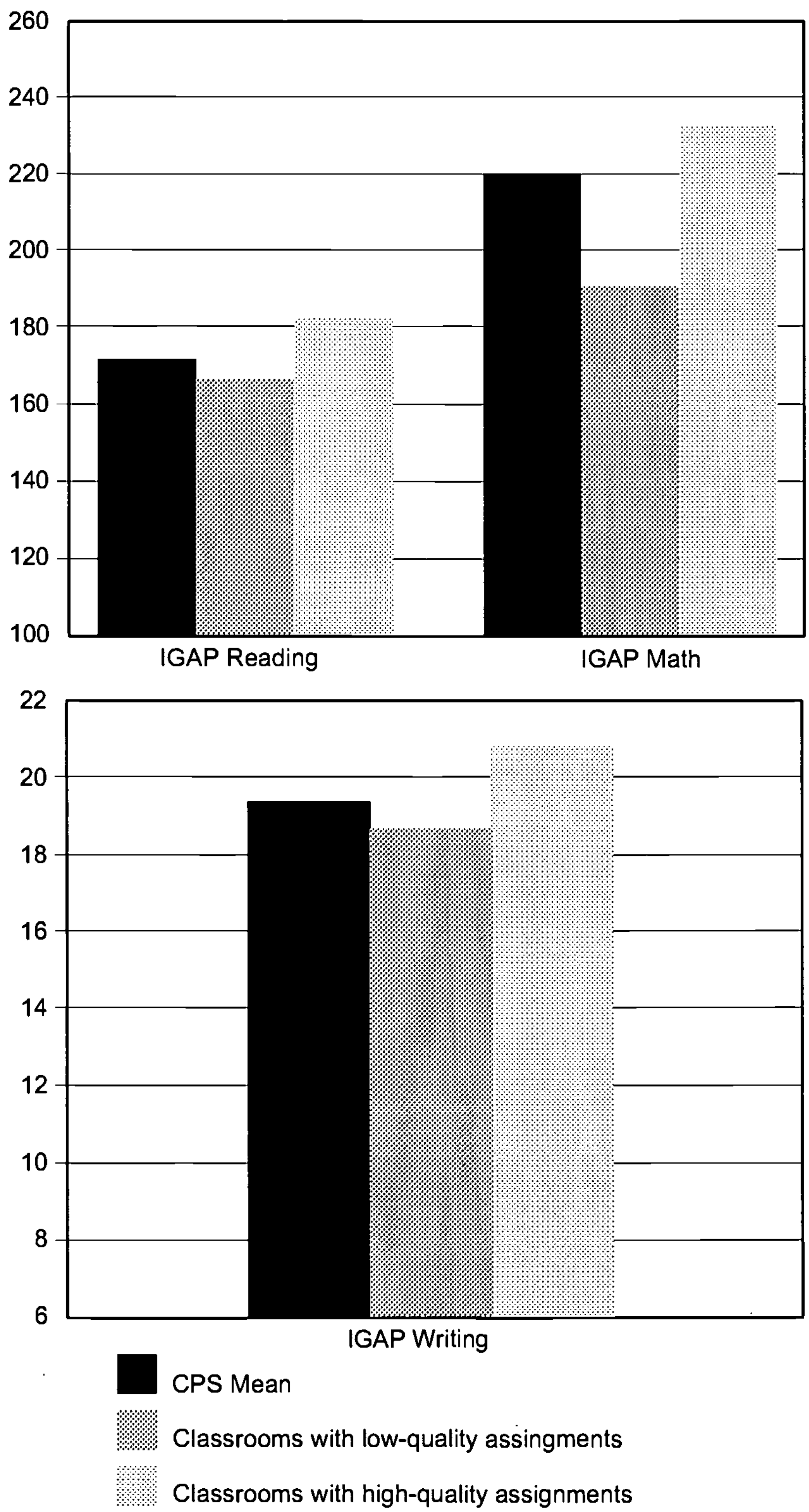
How is Exposure to Authentic Intellectual Work Distributed Among Different Kinds of Chicago Classrooms?

Having found substantial effects of assignment quality on students' basic skills learning, it became important to look more closely at which students actually are exposed to more challenging intellectual assignments. For example, if high-quality assignments occur primarily in classrooms with better-prepared students, this would suggest that the students most needing to improve their basic skills, have less opportunity to do so. With this concern in mind, we conducted a second round of analyses to examine relationships between classroom composition (i.e., prior achievement levels, race, gender, and socioeconomic status) and the intellectual challenge in the assignments assigned. (See Appendix C for further details on this analysis.)

To our surprise, we found virtually no relationship here (see Table 1). Across the fieldwork sample of 437 classes in 19 schools, we found only weak, statistically insignificant correlations between the quality of teachers' assignments and the racial or socioeconomic composition of their classrooms and the level of students' prior achievement. Variations in assignment quality were apparently more a function of teachers' dispositions and individual choices, than of any of the characteristics commonly used to describe students' capacity of students to engage challenging academic work. We find these results encouraging. Although high-quality assignments are not very commonplace in the Chicago Public Schools (see the companion

Figure 4

Students Learn More in Classrooms with High Quality Assignments: Effects of Assignment Quality on IGAP Test Scores



report by Bryk, Newmann, and Nagaoka, 2000), we have evidence that authentic intellectual assignments are occurring in at least some disadvantaged Chicago classrooms and, when they do, students on average learn more.

Do High and Low Achieving Students Benefit Equally from Exposure to Demands for Authentic Intellectual Work?

Another equity consideration, however, still remains. Perhaps not all students within a classroom benefit equally from high-quality assignments. For example, it would be troublesome if challenging intellectual assignments benefitted only the better-prepared students in a classroom, but not their more disadvantaged classmates. Generally, students with higher test scores at the beginning of a school year tend to have higher test scores at the end of the year. However, a key concern here is whether this relationship between entry and exit test scores is exacerbated (or attenuated) in classrooms that emphasize more challenging intellectual assignments. (See Appendix D for technical details on this analytic model.)

Figure 5 displays the results from these analyses. In general, students with high and low prior achievement levels appear to benefit about the same from exposure to authentic classroom assignments. To make our statistical results as concrete as possible we have computed ITBS learning gains for students with both high and low prior achievement levels, and then compared the relative size of these learning gains in classrooms with intellectually authentic versus non-authentic classroom assignments in writing and mathematics. Exposure to high-quality assignments in mathematics appears to benefit low achieving students somewhat more than their high achieving classmates. The gains for initially low achieving students were 29 percent larger in classrooms with high-quality assignments as compared to similar students in classrooms with low-quality assignments. For initially high achieving students, the incremental gain associated with exposure to high-quality assignments was 17 percent. In contrast in reading, higher achieving students appear to benefit a bit more than their lower achieving classmates. The comparable figures are 42 and 28 percent respectively.[19]

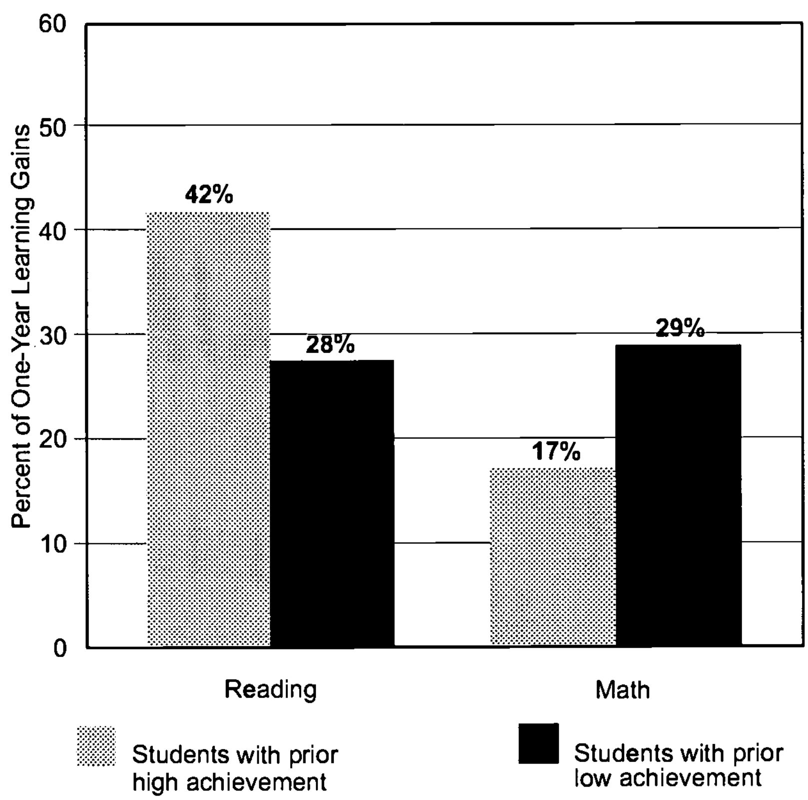
Taken together, these results indicate that a diverse array of students benefit from participation in a classroom with high-quality intellectual assignments. Both students with high and low prior achievement levels learn more over the course of an academic year than comparable students in classrooms with low-quality assignments. In short, authentic intellectual assignments enrich instruction not only for able children, but for all students.

Table 1. Classroom Characteristics Are Not Associated With Students' Exposure to Assignments That Demand Authentic Intellectual Work (classroom-level correlation)

	Quality of Writing Assignments	Quality of Mathematics Assignments
Average Prior Achievement	0.06	0.06
Social Class	0.09	0.02
Percent Female	-0.03	0.00
Percent African-American	-0.04	-0.03
Percent Latino	0.05	0.02

Figure 5

Both Students with High and Low Prior Achievement Levels Benefit: "Value-Added" of Being Exposed to High Quality Assignments



John Booz

IV. Interpretive Summary

There is wide agreement that student success in contemporary society requires not only basic knowledge and skills but also the capacity to engage in more complex intellectual activity. Discussions of the most effective instructional practices or forms of assessment, however, frequently pose a dichotomy between teaching approaches that enhance basic skills versus those that aim at more ambitious intellectual work, implying a tradeoff between these two educational goals. The recent movement to increase student and school accountability through the use of high-stakes tests of basic skills has intensified the debate.

The evidence presented here suggests that this debate rests on a false dichotomy. We found that organizing instruction around challenging, authentic intellectual work can achieve both goals simultaneously. Prior studies have documented that when teachers organize instruction around authentic assignments, students produce more intellectually complex work.[20] This study demonstrates that such assignments also contribute to greater gains on the Iowa Tests of Basic Skills in reading and mathematics, and to better performance in reading, mathematics, and writing on the Illinois Goals Assessment Program tests. While authentic intellectual assignments may not be commonplace in Chicago, our findings also document that they do occur in some very disadvantaged Chicago classrooms and that all students benefit, when given the opportunity to be exposed to such instruction.

We conclude, therefore, that the fears that students will score lower on conventional tests due to teacher demands for more authentic intellectual work appear unwarranted. To the contrary, the evidence indicates that assignments calling for more authentic intellectual work actually improve student scores on conventional tests. The results suggest that if teachers, administrators, policy makers, and the public at large place more emphasis on authentic intellectual work in classrooms, yearly gains on standardized tests in Chicago could surpass national norms.

How Authentic Intellectual Work Improves Basic Skills

These findings may appear a bit startling and call for further explanation. How is it that authentic intellectual work actually enhances basic skills learning? How do demands for construction of knowledge and disciplined inquiry applied to issues that students face beyond school lead to improved student performance on tests that call primarily for recall or recognition of discrete facts, definitions, rules, and application of memorized algorithms? Since neither we nor others have conducted systematic research on this issue, we have no conclusive explanation. We suspect, however,

. . . participation in authentic intellectual activity helps to motivate and sustain students in the hard work that learning requires.

that two mechanisms are at work here. The first derives from the fact that conventional standardized tests make substantial demands for "mastery of vocabulary." That is, these tests, are in large part assessments of students' knowledge of the meaning of many words and concepts. This is most apparent in conventional tests of reading, but it applies to other subjects as well. Since writing tests assess students' proper use of words in sentences, paragraphs, etc., to convey the writer's intended meaning effectively, these also can be seen as tests of student understanding of words. Tests of mathematics and science also, in essence, ask students to show they understand the meaning of concepts and symbols such as add, divide, perimeter, percent, velocity, and temperature.

When teachers ask for authentic intellectual work, they may not use extensive drill and recitation to teach the meaning of words, but they do require students to think about and use words and concepts to solve real problems (rather than using them only to complete routine school exercises). When students "construct knowledge" through "disciplined inquiry," they must often consider alternative solutions, justify their conclusions with reasons and evidence, apply their knowledge to new contexts, develop deep understanding of topics (rather than only superficial awareness), and express themselves through elaborated communication (rather than in terse linguistic fragments). All of these activities emphasize, in one way or another, extensive use and application of words and ideas in varied contexts. As students study a topic in some depth, the concepts that they learn are less likely to remain as disconnected skills and facts, and more likely to be integrated within a larger cognitive schema that connects new bits of information to one another and to students' prior knowledge. This cognitively integrated knowledge is more likely to be owned or internalized by students.

Second, participation in authentic intellectual activity helps to motivate and sustain students in the hard work that learning requires. Since demands for authentic intellectual work pose questions of interest to students in their lives beyond school, students are more likely to care about both the questions they study and the answers they learn. Thus, such assignments enhance a student's willingness to put forth more serious effort in learning the material, as compared to exercises that have no personal meaning to the student beyond completing an assignment to please the teacher or attain a promotion standard.[21]

In sum, assignments that demand more authentic intellectual work elicit a combination of intensive thinking about and a deeper engagement in varied, interconnected applications of words, concepts, and ideas. This can help students to internalize these understandings as their own, and to use this knowledge on the intellectual work that conventional tests present.

Implications

Some readers may be tempted to conclude that our findings justify extensive promotion of project-based learning and the dismissal of lecturing and more traditional forms of teaching. We caution against any simple, unqualified interpretation of this sort. We have seen too many instances of superficial adoption of educational innovations that set back efforts to improve student learning. We wish to emphasize an important distinction often neglected in discussions about school improvement. Namely, no particular teaching practice or strategy assures that students will undertake work that makes high-quality intellectual demands on them. For example, we have observed numerous "hands-on" or "active-learning" classroom projects that, while entertaining for children, contain few demands for authentic intellectual work and afford little opportunity for academic growth. Conversely, we have also witnessed demanding, "teacher-centered" lecture and question-and-answer instruction that requires students to think deeply about issues important in their lives. Our key point is that it is the intellectual demands embedded in classroom tasks, not the mere occurrence of a particular teaching strategy or technique, that influence the degree of student engagement and learning.

Having said this, we do also need to recognize that some teaching practices are more likely to promote complex intellectual work than others. For example, students must have opportunities for extensive writing and conversation if they are to engage in elaborated communication necessary for disciplined inquiry. This point is supported by a companion study in this CARP series that reports higher standardized test score gains in classes where interactive pedagogy occurs.[22] So teaching technique does matter.

Nonetheless, it would be an over generalization of CARP findings to argue that the best or only way to increase conventional test scores is to ask for authentic intellectual work all the time and never to use more conventional, didactic teaching methods. Although more research is needed here, it remains reasonable to assume that authentic intellectual work in classrooms demands a variety of teaching behaviors. While such teaching may use more interactive methods, teachers in these same classrooms are also likely to use intensive didactic instruction to teach certain kinds of material and lessons. In short, we believe that efforts to enhance the quality of the intellectual work in Chicago classrooms requires attention to advancing teachers' expertise with a diverse mix of teaching strategies in order to best engage and support students as they encounter more challenging intellectual work.

Still others may read our results as suggesting the need for a systemwide curriculum consisting of more intellectually challenging tasks. To be sure, our findings indicate that many Chicago teachers need curriculum materials and classroom assessments that include more authentic intellectual challenge. But simply distributing such materials and mandating their use may not produce anything like the results documented in this study. As we noted in our 1998 baseline report on this topic, the mere appearance of more challenging assignments in classrooms is not enough. To use these materials well in their classrooms, many Chicago teachers will have to learn new teaching methods and acquire more subject matter knowledge as well. They will need support and assistance in integrating more challenging assignments with instruction targeted on basic skills, in evaluating the quality of students' answers that show more complex

. . . it is the intellectual demands embedded in classroom tasks, not the mere occurrence of a particular teaching strategy or technique, that influence the degree of student engagement and learning.

interpretation than conventional recitation questions, and, most of all, in teaching their students to succeed on the more demanding assignments.

In short, to significantly increase students' engagement in more authentic intellectual work in Chicago schools, teachers will need substantial new opportunities for professional development and more time to meet with colleagues and outside authorities to work on these concerns. Such effective professional development requires school leaders to place priority on a coherent program of instructional improvement. It also entails restructuring the organization of schools so that they become more productive workplaces for adults and foster the collegial professional activity necessary to sustain such teacher learning and instructional improvement.[23]

The school system can play a key role in moving this agenda forward. Many Chicago schools could benefit from better curriculum materials and classroom assessments with more emphasis on authentic intellectual work. While some promising developments have already emerged here, more are needed.[24] Additional financial resources and new incentives may also be needed to support and catalyze the kind of professional development just described. Local universities and other citywide and national groups can be drawn in to expand the capacity to support such teacher and school development. Existing barriers to improved teacher and student learning, such as the limited time for instruction and professional development, also need to be identified and reconsidered.

In closing, both the research reported here and in our 1998 study has shown that disadvantaged students in poor urban schools can engage in intellectually demanding academic work. When teaching emphasizes such intellectual activity in classrooms, Chicago youngsters have demonstrated both complex intellectual performance and simultaneously impressive gains on standardized tests. The challenge is now set out for adults to organize support along the lines we have suggested to make these opportunities available to students more regularly and more widely.

Appendices

Appendix A
Measuring the Intellectual Demand of Classroom Assignments

Different teachers and schools participated in this study over the course of three years. As a result, we have data from some teachers for three years (19 language arts teachers and 21 math teachers), some teachers for two years (35 language arts teachers and 33 math teachers), and many for only one year (102 language arts teachers and 91 math teachers). In addition, although we had a fixed data collection design, teachers actually provided us with a varying number of assignments, and the composition of these assignments in terms of challenging versus typical tasks also varied. Moreover, we also found considerable variability among task scores from the same classroom. Given the high degree of internal variability in task scores and the limited amount of data from some years in some classrooms for our key predictor variable, we needed to develop a special measurement model in order to extract the maximum information present in these data.

Specifically, in order to increase the reliability of our key predictor, we decided to combine data from each teacher across the one, two, or three years that he or she was in our study. In essence, we developed two teacher-level measures for the intellectual demands, one each for assignments in writing and math, aggregating across all assignments obtained from that teacher during the time that he or she participated in the study. In subsequent analyses where these data are used as a predictor variable, we impute these teacher-level measures back to all of the classes that that teacher taught. We refer to these as a "teacher-classroom" measure of the intellectual quality of assignments. We also developed separate measures for each classroom but found that there was too little reliable variance between classrooms within teachers to use them as predictors.

The actual analytic model used was as follows. Level 1 was a measurement model with two dummy variables to distinguish between mathematics and writing assignments. The outcome variable consisted of the assignment measures generated from the MFRA analysis. The predictors are two indicator variables to distinguish between the math and writing assignments. All of the elements in the level-1 model were weighted by the inverse of the standard error of the assignment measures. (These are produced as a byproduct of the MFRA analysis.) The major function of this level-1 measurement model is to take into account the unreliability of the assignment scores. Formally, the two coefficients produced here, π_{1jk} and π_{2jk} , can be thought of as latent "true scores" for the mathematics and writing tasks respectively. Each of these becomes an outcome variable in a level-2 model where we have multiple assignments per teacher-classroom.

Level 1

$$Y_{ijk} = \pi_{1jk}\,(\text{Math}) + \pi_{2jk}\,(\text{Writing}) + \varepsilon_{ijk}$$

where Y_{ijk} = assignment quality score, and
ε_{ijk} is now assumed N(0,1) given the re-weighting by the standard errors of measurement.

Level 2

$\pi_{1jk} = \beta_{10k} + \beta_{11k}$ (Challenging) + r_{1jk}
$\pi_{2jk} = \beta_{20k} + \beta_{21k}$ (Challenging) + r_{2jk}

Level 3

$\beta_{10k} = \gamma_{100} + \gamma_{101}$ (Grade 6) + γ_{102} (Grade 8) + u_{1k}
$\beta_{11k} = \gamma_{110}$
$\beta_{20k} = \gamma_{200} + \gamma_{201}$ (Grade 6) + γ_{202} (Grade 8) + u_{2k}
$\beta_{21k} = \gamma_{210}$

At Level 2, a dummy variable, distinguishing challenging from typical tasks, was grand-mean centered and its effect fixed. As a result, the intercept terms, β_{10k} and β_{20k} are the classroom mean score (in mathematics and writing, respectively), adjusted for differences among teachers in the number and types of tasks they submitted. Finally, at Level 3 (i.e., the teacher-classroom level), indicator variables for grades six and eight were included in order to adjust for grade-specific effects in the assignment rating system. The empirical Bayes estimates from this model for β_{10k} and β_{20k} were used as in subsequent analyses as the teacher-classroom measure of assignment demand in mathematics and writing, respectively.

Appendix B
HLM Model for Estimating the Effects of the Quality of Assignments on Basic Skills Learning

To analyze the effects of the quality of assignments on students' basic skills learning we employ a 3-level latent variable analysis model using HLM Version 5.0 (Raudenbush, Bryk, Fai, and Congdon, 2000, p. 207). This model was run separately for ITBS reading and mathematics outcome scores and IGAP reading, writing, and mathematics outcomes scores. The input data at Level 1 consisted of a teacher-classroom assignment measure (see Appendix A) and the student end-of-year test data stacked together from each teacher-classroom. Each level-1 outcome was weighted inversely proportional to it respective standard error. For the ITBS test data, the standard error is produced in the equating process and for the IGAP, the standard error is based on an HLM analysis of the test reliability. The standard error of the assignment measure is the square root of the posterior variance of the estimate generated by the analysis described in Appendix A. Note that, for those teachers who participated in the study for two or more years, the student data from multiple years was stacked together as one teacher-classroom. Two level-1 indicator variables distinguished between the teacher-classroom assignment measure and the student test scores. The coefficient π_{1jk} represents the "true" or latent value for each teacher-classroom's assignment score. Similarly, π_{2jk} represents the latent outcome score for each student. Each of these coefficients in turn becomes an outcome variable in the student model at Level 2.

We then entered at level 2 the student's ITBS scores in mathematics and reading from the previous year as key control variables. We also introduced dummy variables for whether the student was African-American, Latino, or female and a measure for SES and dummy variables for the year the student was in the classroom. The ITBS pretest scores were centered around the mean for the year and the grade of testing, so that the value of the pretest was the deviation from the year-grade group mean. The latent assignment measure, π_{1jk} , was specified fixed at Level 2 because all students within the same classroom shared the same assignments. Grade-level indicator variables were introduced in the level-3 model and effects coded so that the intercept, γ_{200}, represents the overall adjusted mean achievement for all three grades.

Finally, using a latent variable model formulation in HLM version 5, we moved the latent measure for teacher classroom assignments, β_{10k}, to the "predictor side" of the equation for β_{20k}. Given the formulation of the level-2 model, β_{20k} represents the adjusted mean test score outcome in classroom k after taking into account differences among classrooms in terms of measured student backgrounds. Thus, β_{20k} indicates the learning gain specific to classroom k. The latent variable model relates these classroom specific learning gains to the quality of the assignments teachers made.

We note that the latent variable model formulation was essential in these analyses given the relatively low reliability between classrooms in the assignment measures. Had we not employed such a model, we would have had to enter an errorful predictor at Level 3, which would have biased the final results.

Level 1 (measurement model)

$$Y_{ijk} = \pi_{1jk} \text{ (assignment dummy)} + \pi_{2jk} \text{ (ITBS dummy)} + \varepsilon_{ijk}$$

where Y_{ijk} = ITBS (or IGAP) score or assignment quality score, and
ε_{ijk} is now assumed N(0,1) given the reweighting by the standard errors of measurement.

Level 2 (students)

$$\pi_{1jk} = \beta_{10k}$$
$$\pi_{2jk} = \beta_{20k} + \beta_{21k} \text{ (ITBS reading pretest score)} + \beta_{22k} \text{ (ITBS math pretest score)} + \beta_{23k} \text{ (SES)} + \beta_{24k} \text{ (African-American)} + \beta_{25k} \text{ (Latino)} + \beta_{26k} \text{ (female)} + \beta_{27k} \text{ (Year 1998)} + \beta_{28k} \text{ (Year 1999)} + r_{2jk}$$

Level 3 (teacher-classroom level)

$$\beta_{10k} = \gamma_{100} + u_{1k}$$
$$\beta_{20k} = \gamma_{200} + \gamma_{201} \text{ (Grade 6)} + \gamma_{202} \text{ (Grade 8)} + u_{2k}$$
$$\beta_{21k} = \gamma_{210}$$
$$\beta_{22k} = \gamma_{220}$$
$$\beta_{23k} = \gamma_{230}$$
$$\beta_{24k} = \gamma_{240}$$
$$\beta_{25k} = \gamma_{250}$$
$$\beta_{26k} = \gamma_{260}$$
$$\beta_{27k} = \gamma_{270}$$
$$\beta_{28k} = \gamma_{280}$$

Final Latent Variable Model Formulation

$$\beta_{20k} = \gamma_{200} + \gamma_{201} \text{ (Grade 6)} + \gamma_{202} \text{ (Grade 8)} + \gamma_{203} (\beta_{10k}) + u_{2k}$$

Where β_{20k} indicates the size of the learning gain specific to classroom k,
β_{10k} is the latent assignment quality measure in classroom k (in either writing or mathematics), and
γ_{203} estimates the effect of the quality of classroom assignments on learning gains.

Note: For the reading and writing test outcomes, we used the classroom writing assignments score in the level-1 model above. For analyzing the mathematics outcomes, we used the quality score from the classroom mathematics assignments.

HLM Results for Estimating the Effect of Assignment Quality on ITBS Gains

	Reading	Mathematics
Class Assignment		
Intercept	5.89 *	5.91 *
Test Score		
Intercept	-0.14 *	0.11 *
Grade 6	0.03	0.29 *
Grade 8	1.00 *	1.20 *
Reading Pretest	0.48 *	0.13 *
Math Pretest	0.24 *	0.70 *
SES	0.002	-0.01
Female	0.01	-0.01
African-American	-0.16 *	-0.10 *
Latino	-0.10 *	0.02
Year 1998	0.12 *	0.15 *
Year 1999	0.01	0.04
Latent Variable Regression		
Outcome: Test Score		
Intercept	-0.83 *	-1.11
Grade 6	0.03	0.29 *
Grade 8	1.00 *	1.20 *
Class Assignments	0.12 *	0.21 +

HLM Results for Estimating the Effect of Assignment Quality on IGAP Test Scores

	Writing	Reading	Mathematics
Class Assignment			
Intercept	5.90 *	5.89 *	5.89 *
Test Score			
Intercept	19.59 *	174.21 *	216.21 *
Grade 6	0.14	-11.31	-5.45
Grade 8	1.71 *	-5.56	-2.69
Reading Pretest	0.76 *	46.45 *	15.43
Math Pretest	1.06 *	28.93 *	54.04
SES	-0.11	-1.05	-4.29
Female	1.01 *	2.90	-2.28
African-American	0.29	-21.27	-14.64 *
Latino	0.47 +	-11.05	-2.24
Year 1998	0.41	22.07 *	10.74 *
Latent Variable Regression			
Outcome: Test Score			
Intercept	14.65 *	139.79 *	28.57
Grade 6	0.14	-11.31	-5.45
Grade 8	1.71 *	-5.56	-2.69
Class Assignments	0.84 *	5.83 *	31.84 *

* Significant at .05 or higher   + Significant at .1 or higher

Appendix C
HLM Model for Estimating the Effects of Classroom Composition on Exposure to Challenging Intellectual Assignments

The purpose of this analysis was to determine whether the level of intellectual challenge in assignments from different classrooms depended on student demographics, namely race, prior achievement, socioeconomic class, and gender. Separate analysis was run for writing and math tasks. The quality of assignments, measured now at the classroom by year rather than a teacher-classroom composite, was used as the outcome variable in the level-1 model. Level 1 again was a measurement model that adjusts for the varying reliability of the classroom assignment scores.

The next two levels of the model consisted of years nested within classrooms. This allowed us to take into account that while some teachers appeared in the study two or more years, their classroom composition might vary from year to year. The latent classroom quality score for each year was predicted by the classroom means for prior achievement in reading and mathematics, SES, race, and gender. The year of data collection was also included as a predictor. Prior year test scores were centered around the mean for the year and grade of testing. The remaining demographic variables were grand-mean centered. At Level 3, effects-coded indicator variables for grade six and grade eight were used as controls on the adjusted classroom assignment quality score.

Level 1 (measurement model)

$$Y_{ijk} = \pi_{1jk} \text{ (assignment dummy)} + \varepsilon_{ijk}$$

where Y_{ijk} = assignment quality score, and
ε_{ijk} is now assumed N(0,1) given the reweighting at Level 1 by the standard errors of measurement.

Level 2 (years)

$$\pi_{1jk} = \beta_{10k} + \beta_{11k} \text{ (Reading pretest score)} + \beta_{12k} \text{ (Math pretest score)} + \beta_{13k} \text{ (SES)} + \beta_{14k} \text{ (African-American)} + \beta_{15k} \text{ (Latino)} + \beta_{16k} \text{ (female)} + \beta_{17k} \text{ (Year 1998)} + \beta_{18k} \text{ (Year 1999)} + r_{1jk}$$

Level 3 (classroom)

$$\beta_{10k} = \gamma_{100} + \gamma_{101} \text{ (Grade 6)} + \gamma_{102} \text{ (Grade 8)} + u_{1k}$$
$$\beta_{11k} = \gamma_{110}$$
$$\beta_{12k} = \gamma_{120}$$
$$\beta_{13k} = \gamma_{130}$$
$$\beta_{14k} = \gamma_{140}$$
$$\beta_{15k} = \gamma_{150}$$
$$\beta_{16k} = \gamma_{160}$$
$$\beta_{17k} = \gamma_{170}$$
$$\beta_{18k} = \gamma_{180}$$

HLM Results for the Relationship of Classroom Composition to Exposure to Challenging Assignments

	Writing	Mathematics
Class Assignment		
Intercept	5.81 *	5.75 *
Grade 6	0.82 *	0.22 *
Grade 8	0.44 *	-0.47 *
Reading Pre-test	0.06	0.06
Math Pre-test	0.08	0.06
SES	0.05	0.04
Female	-0.46	-0.25
African-American	0.18	-0.34
Latino	0.39	-0.20
Year 1998	0.46 *	0.22 *
Year 1999	0.20 *	0.15 *

* Significant at .05 or higher

Note: For purposes of simplicity of exposition, the fixed effect coefficients above were converted to correlations to present in the text of this report. Note that all of the classroom composition measures had a nonstatistically significant relationship for mathematics and writing.

Appendix D
An HLM Model to Estimate Differential Within Classroom Effects of Exposure to Challenging Intellectual Assignments

The purpose of this analysis was to examine whether the intellectual challenge of assignments caused increased differentiation in basic skills acquisition among students within classrooms. Specifically, do academically more able students benefit from challenging instruction but at a price to their less advantaged classmates? Data from all three years were included, but separate analyses were conducted for writing and math tasks.

Level 1 was a measurement model that adjusted for the reliability of the standardized test scores (i.e., ITBS reading and math; IGAP reading, math, and writing). The outcome here was the reported student test score weighted inversely by the standard error of that score. The key student level predictor in this analysis was prior year achievement. This was entered at Level 2. The measure of the intellectual demand of classroom assignment level was entered at Level 3, along with information about the particular grade level.

Level 1 (measurement model)

$$Y_{ijk} = \pi_{1jk} \text{ (dummy variable)} + \varepsilon_{ijk}$$

where Y_{ijk} = standardized test score i for child j in classroom k, and
ε_{ijk} is now assumed N(0,1) given the reweighting at Level 1 by the standard errors of measurement.

Level 2 (students)

$$\pi_{1jk} = \beta_{10k} + \beta_{11k} \text{ (pretest score)} + \beta_{12k} \text{ (SES)} + \beta_{13k} \text{ (African-American)} + \beta_{14k} \text{ (Latino)} + \beta_{15k} \text{ (female)} + r_{1jk}$$

Level 3 (teacher-classroom)

$$\beta_{10k} = \gamma_{100} + \gamma_{101} \text{ (Grade 6)} + \gamma_{102} \text{ (Grade 8)} + \gamma_{103} \text{ (assignment score)} + u_{10k}$$
$$\beta_{11k} = \gamma_{110} + \gamma_{111} \text{ (Grade 6)} + \gamma_{112} \text{ (Grade 8)} + \gamma_{113} \text{ (assignment score)} + u_{11k}$$
$$\beta_{12k} = \gamma_{120}$$
$$\beta_{13k} = \gamma_{130}$$
$$\beta_{14k} = \gamma_{140}$$
$$\beta_{15k} = \gamma_{150}$$

The analysis problem here resembles that of Appendix B where we used a latent variable HLM model to assess the affects of assignment quality on student learning gains. Now we wish to enter assignment score as a predictor at Level 3 for both the adjusted average classroom achievement, β_{10k}, and for β_{11k}, which measures the pre-post relationship separately for each classroom. We are specifically concerned now about the effect of assignment quality on this latter relationship. This is estimated by γ_{113}. A positive value for this coefficient means that higher-achieving students within a classroom gain more from challenging instruction than do their lesser-prepared classmates. We would characterize this as a disequalizing effect. In contrast, a negative coefficient implies an equalizing effect.

Unfortunately, version 5.0 of HLM does not permit a model specification where multiple level 3 equations include a latent variable predictor. Instead, we employ a conventional (measured-variables) HLM where we used the Empirical Bayes estimate for the teacher classroom assignment score, derived in Appendix A, as the level-3 predictor. Formally, this EB estimate is an instrumental variable representation for the errorful observed data.

As a final check on the adequacy of this analysis, we compared the estimates for γ_{103} from this analysis with those estimated from the latent variable model described in Appendix A. The same basic pattern of effects occurred in both analyses. For ITBS reading, the structural effect estimates were slightly higher here (0.11 versus 0.15) than in the latent variable model. The reverse was true for ITBS math. In general, the two models produced the same statistical inferences and similar estimates of the magnitude of effects. This provides some additional assurance about the accuracy of the inferences for the differentiating effects, γ_{113}, reported here.

HLM Results on the Differential Effects by Prior Achievement

	Reading	Mathematics
ITBS Test Score		
Intercept	-0.91*	-0.49*
Grade 6	-0.07	0.29*
Grade 8	0.99*	1.20*
Class Assignment Score	0.14*	0.12*
Pre-test		
Intercept	0.43*	0.91*
Class Assignment Score	0.03	-0.03
SES	-0.02	-0.02
Female	-0.10	0.00
African-American	-0.19*	-0.12*
Latino	-0.12*	-0.01

* Significant at .05 or higher

Endnotes

[1] Early leaders in this area included the state of Kentucky and the Charlotte-Mecklenburg, N.C. school district.

[2] Lee, Smith, and Newmann (2001) explain that didactic instruction is sometimes referred to as teacher-directed, conventional, or highly structured instruction. They describe didactic instruction in more detail and contrast it with "interactive" instruction, which tends to include more emphasis on complex project-based intellectual work.

[3] Delpit (1995) argued that certain kinds of "progressive" pedagogy—for example, in teaching the writing process—neglect important needs of many low-income urban African-American students to learn basic skills necessary to succeed in education and society.

[4] For example, Good and Brophy (2000) summarized differences in the two approaches as the transmission view and social construction view of teaching.

[5] Examples of proposals or programs that reflect more of an emphasis on didactic instruction include Bereiter and Engelmann (1966); Hirsch (1987); Stein, Silbert, and Carnine (1997); and Strickland (1998). Examples emphasizing interactive instruction include Brooks and Brooks (1993); Driver (1995); Goodman and Goodman (1979); and Thomas (1991). The Web site www.publicagenda.org provides periodic surveys of public attitudes and summaries of arguments related to these approaches, (e.g., under Framing the Debate, "Perspective #2: Creating Student-Centered Schools."

[6] The main public report of this research was Newmann and Wehlage (1995) which was distributed by the Wisconsin Center for Education Research, the American Federation of Teachers, the Association for Supervision and Curriculum Development, the National Association of Elementary School Principals, and the National Association of Secondary School Principals. A more detailed account is available in Newmann and Associates (1996). This research has been referenced in numerous professional publications, and it has been the subject of professional development for numerous schools and districts in the United States and abroad. The standards for intellectual quality have been integrated into Michigan's state curriculum standards and assessments and are the focus of research and school development projects in Australia. To be sure, while this is not the only conceptual framework with associated scoring rubrics that could be applied to assess the intellectual challenge of classroom assignments, the framework is considered useful in many places for helping teachers move beyond traditional teaching of basic skills to more ambitious intellectual work.

[7] Much of the material in this section is taken from Newmann, Lopez, and Bryk (1998). The perspective on authentic intellectual work was originally proposed by Archbald and Newmann (1988), then revised and further developed by Newmann, Secada, and Wehlage (1995). The first empirical research using this conception of achievement was presented in Newmann and Associates (1996).

[8] For evidence of the complex intellectual demands in the contemporary workplace see Cappelli, Bassi, Katz, Knoke, Osterman, and Useem (1997); Decker, King Rice, Moore, and Rollefson (1997); Murnane and Levy (1996); and National Center on Education and the Economy (1990). Explanations of the complex intellectual demands of democratic citizenship are available in Aristotle (trans. 1946); Barber (1984), Dewey ([1916]1966); and Jefferson (1939 version).

[9] The call for "relevant" or "student-centered" curriculum is, in many cases, a less precise expression of the view that student accomplishments should have value beyond simply indicating school success. Nevertheless, while some people may regard the term "authentic" as synonymous with curriculum that is "relevant," "student-centered," or "hands-on," we do not. Value beyond the school is only one component of authentic intellectual work.

[10] For both writing and mathematics, elaborated communication is the only criterion specified for disciplined inquiry. Yet, our definition of criteria for authentic intellectual work also identified prior knowledge base and in-depth understanding as aspects of disciplined inquiry. Writing and mathematics authorities concluded it was unnecessary to specify a demand for prior knowledge, because they assumed that in both writing and mathematics, all three standards imply that students are asked to use prior knowledge. The writing standards did not include specific reference to in-depth understanding, because authorities in the teaching of writing could not identify a specific set of concepts or problems that must be understood for student writers to demonstrate disciplined inquiry in writing. The mathematics authorities concluded that the standards for construction of knowledge and written mathematical communication, when considered together, constitute a requirement that students show in-depth understanding of a mathematical concept, problem, or issue.

[11] The manual for scoring assignments in writing and mathematics is available from the Consortium on Chicago School Research.

[12] The social class variable used in this analysis is a composite indicator developed by the Consortium to measure the concentration of poverty. It consists of the students' census block group variables for the percentage of males over age 18 employed one or more weeks during the year and the percentage of families above the poverty line.

[13] We designated as higher quality those assignments that were rated, on the standards for authentic intellectual work, in the top quartile of all assignments in the subject and grade level. We designated as lower quality those assignments that were

rated in the bottom quartile of all assignments in the subject and grade level.

[14] The national gain comparison used in these displays was estimated from the norming tables provided by the ITBS. For the three grades under study here, (third, sixth, and eighth), we examined the amount of learning in the course of one year reflected in these tables for students who were at national norms, e.g., a student moving from a grade equivalent of 2.8 to 3.8 over the course of third grade. Since this amount varies depending upon the particular combination of test forms used to estimate the one-year learning gain, we averaged these results across the different forms used by the CPS during this period.

[15] The advantage was computed by subtracting the percent gain of the average yearly Chicago gain associated with exposure to lower levels from the percent gain associated with exposure to higher levels of authentic intellectual work.

[16] Since the IGAP assessments were typically administered in March of the academic year, technically these results are associated with seven months of instruction.

[17] Based on CPS data, we estimated a simple student level standard deviation of about 4.4 for the IGAP writing assessment. We used this estimate to compute an effect size for this measure.

[18] For purposes of computing these effect sizes, we used the standard deviation of 75 from the norming sample for the IGAP reading and math tests.

[19] We note that the differences reported here and in Figure 5 (e.g., 17 versus 29 percent in reading) did not achieve statistical significance at the .05 level. This is consistent with an inference that students' prior achievement level does not affect the benefit derived from exposure to high-quality classroom assignments.

[20] Newmann and Associates (1996): Newmann, Lopez, and Bryk (1998).

[21] For evidence of the connection between authentic intellectual work and student engagement, see Newmann and Associates (1996); Kane, Khattri, Reeve, Adamson, and Pelavin Research Institute (1995); Marks (2000); Avery (1999).

[22] See Lee, Smith, and Newmann (2001).

[23] For a further discussion of this theme, see the companion report on coherent instructional improvement (Newmann, Smith, Allensworth, and Bryk 2000).

[24] One exemplar in this regard is the CPS-Northwestern partnership around developing new science and technology curriculum for grades six through eight.

References

Archbald, Douglas A., and Fred M. Newmann (1988). *Beyond Standardized Testing: Assessing Authentic Academic Achievement in Secondary Schools.* Reston, VA: National Association of Secondary School Principals.

Aristotle (1946). *The Politics of Aristotle.* (Trans. Sir Ernest Barker). Oxford: Clarendon Press.

Avery, Patricia G. (1999). Authentic instruction and assessment. *Social Education* 65 no. 6: 368-73.

Barber, Benjamin R. (1984). *Strong Democracy: Participatory Politics for a New Age.* Berkeley, CA: University of California Press.

Bereiter, Carl and Siegfried Engelmann (1966). *Teaching Disadvantaged Children in the Preschool.* Englewood Cliffs, NJ: Prentice Hall.

Brooks, J. G., and Martin G. Brooks (1993). *In Search of Understanding: The Case for Constructivist Classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Bryk, Anthony S., Jenny K. Nagaoka, and Fred M. Newmann (2000). *Chicago Classroom Demands for Authentic Intellectual Work: Trends from 1997-1999.* Chicago: Consortium on Chicago School Research.

Cappelli, Peter, Laurie Bassi, D. Katz, David Knoke, P. Osterman, and Michael Useem (1997). *Change At Work.* New York: Oxford University Press.

Carpenter, Thomas P., Elizabeth Fennema, Penelope L. Peterson, Chia-Ling Chiang, and Megan Loef (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *American Educational Research Journal* 26, no. 4: 499-531.

Cobb, Paul, Terry Wood, Erna Yackel, Eastmond J. Nicholls, Grayson Wheatley, B. Trigatti, and M. Perlwitz (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education* 22, no. 1: 3-29.

D'Agostino, Jerome V. (1996). Authentic instruction and academic achievement in compensatory education classrooms. *Studies in Educational Evaluation*, 22 no. 2, 139-55.

Decker, Paul T., Jennifer King Rice, Mary T. Moore, and Mary R. Rollefson (1997). *Education and the Economy: An Indicators Report.* Washington, DC: National Center for Education Statistics, U. S. Department of Education, Office of Educational Research and Improvement.

Delpit, Lisa (1995). *Other People's Children: Cultural Conflict in the Classroom.* New York: New Press.

Dewey, John ([1916] 1966). *Democracy and Education.* New York: Free Press.

Driver, Rosalind (1995). Constructivist approaches to science teaching. In *Constructivism in Education*, edited by L. P. Steffe and J. Gale. Hillsdale, NJ: Erlbaum.

Good, Thomas L. and Jere E. Brophy (2000). *Looking in Classrooms* (8th ed.). Reading, MA: Longman.

Goodman, Kenneth S. and Yetta M. Goodman (1979). Learning to read is natural. In *Theory and Practice of Early Reading*, eds. L. B. Resnick and P. A. Weaver, Vol. 1. Hillsdale, NJ: Erlbaum.

Hirsch, Eric D., Jr. (1987). *Cultural Literacy: What Every American Needs to Know.* Boston: Houghton Mifflin.

Jefferson, Thomas (1939). *Democracy* (Ed. Saul K. Padover). New York: D. Appleton-Century Company.

Johnson, J. (1995). *Assignment Incomplete: The Unfinished Business of Education Reform: A Report From Public Agenda.* New York: Public Agenda.

Johnson, Jean and John Immerwahr (1994). *First Things First: What Americans Expect From the Public Schools: A Report from Public Agenda.* New York: Public Agenda.

Kane, Michael B., Nidhi Khattri, Alison L. Reeve, Rebecca J. Adamson, and Pelavin Research Institute (1995). *Assessment of Student Performance: Studies of Education Reform.* 3 vols. Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education.

Knapp, Michael S., Patrick M. Shields, and Brenda J. Turnbull (1992). *Academic Challenge for the Children of Poverty: Summary Report.* Washington, DC: Office of Policy and Planning, U. S. Department of Education.

Lee, Valerie E., Julia B. Smith, and Robert G. Croninger (1997). How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education*, 70 (April): 128-50.

Lee, Valerie E., Julia B. Smith, and Fred M. Newmann (2001). *Instruction and Achievement in Chicago Elementary Schools.* Chicago: Consortium on Chicago School Research.

Levin, Martin, Fred M. Newmann, and D. W. Oliver (1969). *A Law and Social Science Curriculum Based on the Analysis of Public Issues: A Final Report.* Washington, DC: Office of Education, Bureau of Research, U. S. Department of Health, Education and Welfare.

Marks, Helen M. (2000). Student engagement in instructional activity: patterns in elementary, middle, and high schools. *American Educational Research Journal* 37 no.1.

Murnane, Richard J., and Frank Levy (1996). *Teaching the New Basic Skills: Principles for Children to Thrive in a Changing Economy.* New York: Free Press.

National Center on Education and the Economy (1990). *America's Choice: High Skills or Low Wages? The Report of the Commission on the Skills of the American Workforce.* Rochester, NY.

Newmann, Fred M. and Associates (1996). *Authentic Achievement: Restructuring Schools for Intellectual Quality.* San Francisco, CA: Jossey-Bass.

Newmann, Fred M., Gudelia Lopez, and Anthony S. Bryk (1998). *The Quality of Intellectual Work in Chicago Schools: A Baseline Report.* Chicago: Consortium on Chicago School Research.

Newmann, Fred M., Walter G. Secada, and Gary G. Wehlage, (1995). *Successful School Restructuring: A Report to the Public and Educators.* Madison, WI: Wisconsin Center on Education Research, University of Wisconsin.

Newmann, Fred M., BetsAnn Smith, Elaine Allensworth, and Anthony S. Bryk (2001). *School Instructional Program Coherence: Benefits and Challenges.* Chicago: Consortium on Chicago School Research.

Silver, Edward A., and Suzanne Lane (1995). Can instructional reform in urban middle schools help students narrow the mathematics performance gap? *Research in Middle Level Education* 18, no. 2, 49-70.

Smylie, Mark A., Stacy A. Wenzel, Penny Bender Sebring, Elaine Allensworth, Tania Gutierrez, Sara Hallman, and Shazia R. Miller (forthcoming). *Improvement in Chicago Annenberg Schools, 1996-1999: Interim Technical Report of the Chicago Annenberg Research Project.* Chicago: Consortium on Chicago School Research.

Stein, M., J. Silbert, and Douglas Carnine (1997). *Designing Effective Mathematics Instruction: A Direct Instruction Approach.* Upper Saddle River, NJ: Merrill.

Strickland, Dorothy S. (1998). *Teaching Phonics Today: A Primer for Educators.* Newark, DE: International Reading Association.

Tharp, Roland G. (1982). The effective instruction of comprehension: results and description of the kamehameha early education program. *Reading Research Quarterly* 17, no. 4, 503-27.

Thomas, D. A. (1991). *Children, Teachers, and Mathematics.* Boston, MA: Allyn and Bacon.

About the Authors

Fred M. Newmann is Professor Emeritus of Curriculum and Instruction at the University of Wisconsin-Madison and a member of the Consortium's lead team for the Chicago Annenberg Research Project. He has worked as a high school social studies teacher, curriculum developer, teacher-educator and researcher of school reform. Having directed the National Center on Effective Secondary Schools and the Center on Organization and Restructuring of Schools, his recent publications propose criteria for authentic achievement and assessment and describe how restructuring in schools nationwide affects authentic student achievement, equity, empowerment, professional community, and accountability. He recently completed a national study of professional development to build capacity in low-income schools. Newmann co-authored the Chicago Annenberg Research Project report, *The Quality of Intellectual Work in Chicago Schools: A Baseline Report* (October 1998).

Anthony S. Bryk is the Director of the Center for School Improvement, Marshall Field, IV Professor of Urban Education and Sociology at the University of Chicago, and Senior Director of the Consortium. Professor Bryk has developed new statistical methods in education that have contributed to studies of school effects. Among his current research interests are the social organization of schools and the effects of school restructuring to improve student learning. As Director of the Center for School Improvement, he also leads the design team that is developing the North Kenwood/Oakland Professional Development Charter School for the Chicago Public School system.

Jenny Nagaoka is a Research Associate at the Consortium on Chicago School Research. She received her B.A. from Macalester College in St. Paul, Minnesota, and her M.A. in public policy from the Harris School of Public Policy at the University of Chicago. Prior to graduate school, she lived in Sendai, Japan, where she attended Tohoku University under a Fulbright Fellowship and was a junior high and high school English teacher.