Rigor, Relevance, and Results: The Quality of Teacher Assignments and Student Work in New and Conventional High Schools

Evaluation of the Bill & Melinda Gates Foundation's High School Grants

Prepared by:

The American Institutes for Research

SRI International

AMERICAN INSTITUTES FOR RESEARCH®

SRI International®

Rigor, Relevance, and Results: The Quality of Teacher Assignments and Student Work in New and Conventional High Schools

July 2005

Prepared for:

The Bill & Melinda Gates Foundation
P.O. Box 23350
Seattle, WA 98102

Prepared by:

The National Evaluation of High School Transformation
http://www.air.org/publications/pubs_ehd_school_reform.aspx

American Institutes for Research
1000 Thomas Jefferson St., NW
Washington, DC 20007
http://www.air.org

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
http://ctl.sri.com

The National Evaluation of High School Transformation is a collaborative effort between the American Institutes for Research and SRI International. This work, which began in 2001, is supported through funding by the Bill & Melinda Gates Foundation.

Rigor, Relevance, and Results: The Quality of Teacher Assignments and Student Work in New and Conventional High Schools is part of an ongoing series of reports based on the evaluation of the Bill & Melinda Gates Foundation's high school grants. The views, findings, conclusions, and recommendations expressed herein are those of the authors and do not necessarily express the viewpoint of the foundation.

Authors:

Karen Mitchell, Jamie Shkolnik, Mengli Song, Kazuaki Uekawa, Robert Murphy, Mike Garet, and Barbara Means

Principal Investigators:

Becky Smerdon
American Institutes for Research

Barbara Means
SRI International

Rigor and Relevance Study Team:

Linda Choi
Matt Gaertner
Mike Garet
Steven Leinwand
Connie Liu
Jamie Shkolnik
Mengli Song
Kazuaki Uekawa
American Institutes for Research

Larry Gallagher
Amy Hafter
Catherine Hall
Lauren Hersh
Barbara Means
Karen Mitchell
Robert Murphy
Janet Voight-Miro
SRI International

Outside Experts:

Michelle Bissonette, Los Altos High School
David Jolliffe, DePaul University
Carmen Manning, University of Wisconsin, Eau Claire

Contents

Exhibits

Tables

Figures

Executive Summary

While some secondary students are tackling not just college-preparatory but also college-level work, others take general or remedial courses that do not provide them with the skills required for higher education. Low-income and minority students are particularly likely to either drop out or take courses that leave them underprepared for higher education. Little more than half of the African-American and Latino youth who start ninth grade finish high school with a diploma. Fewer than 30% are ready to enter higher education without doing remedial work.

The Bill & Melinda Gates Foundation's High School Grants initiative seeks to catalyze the creation of a new kind of American high school—one where *every* student has a challenging academic program that provides sound preparation for college and for family-wage jobs and the demands of good citizenship. The foundation believes that to better serve all students, high schools need to become places that combine rigor in the academic program of every student (not just those in an honors or higher track) with relevance to their interests and potential career choices, supported by positive relationships that can inspire students both academically and personally.

When the foundation launched its national education initiative, it began giving grants to nonprofit organizations charged with creating high schools that would embody these ideals. These intermediary organizations received grants to establish high schools, either by starting new schools or by redesigning existing comprehensive high schools into smaller, more focused units (separate "learning communities" or "academies"). In either case, the foundation expected the resulting schools to be characterized by a coherent vision, high expectations for all their students, and powerful teaching and learning. The foundation expected that schools created under its initiative would be small in size (typically no more than a hundred students per grade), not because "small" was an end in itself but because the greater personal attention that comes with a small size is conducive to implementation of the practices and climate of effective schooling.

An evaluation team composed of researchers from the American Institutes for Research (AIR) and SRI International has been studying the schools being created or redesigned through this initiative since 2001. This report takes an in-depth look into the classrooms of foundation-affiliated schools, examining teaching and learning to see whether the students in these schools have challenging learning opportunities and, if they do, whether students rise to the challenge and produce high-quality student work.

Data Sources

This report uses data collected from schools in 2002–03 and 2003–04 to explore the nature of teaching practices and the quality of students' work. It examines samples of assignments and student work gathered from English/language arts and mathematics classrooms. Assignments were collected from 10th-grade English/language arts classes in eight large Washington state high schools in 2002–03, the year in which teachers at those schools were working on plans to redesign their schools into small learning communities. Additionally, assignments and work were collected in school year 2003–04 from 12 new high schools and 4 additional comprehensive high schools planning for redesign (located outside the state of Washington). We asked each participating teacher to provide four assignments that were typical of the assignments given to their students on a day-to-day basis, and four others that challenged their students and showed what their students know and can do at high levels. For three of the assignments, teachers provided the work that their students produced.

To gauge the rigor of teachers' assignments, we had teams of master teachers score them with rubrics capturing the extent to which the assignments required students to move beyond the reproduction of information to construct knowledge, communicate clearly and well, and use language and mathematics conventions accurately and effectively. Relevance was assessed through a separate scoring of assignments on the extent to which they called for work with authentic purposes and the extent to which they asked students to make choices about what and how they would learn. Thus, the rigor score captures the intellectual demands of assignments, while the relevance score deals with real-world applicability and students' involvement in shaping their own learning experiences. Teachers judged the quality of students' work by examining the extent to which student products demonstrated the construction of knowledge, deep conceptual understanding of important content, reasoning and problem-solving facility, effective communication, and accurate use of language and mathematics conventions.

The data were used to address five research questions:

1. How do the rigor and relevance of learning opportunities in new high schools compare with the rigor and relevance of learning opportunities in comprehensive high schools?

2. How does teacher feedback on students' work in new high schools compare with the feedback provided by teachers in comprehensive high schools?

3. How does the quality of students' work in new high schools compare with the quality of students' work in comprehensive high schools?
4. To what extent are rigorous and relevant assignments associated with high-quality student work?
5. To what extent do differences in the quality of student work relate to school differences on jurisdiction-sponsored achievement tests?

Interview and observation data were used to help explain some of the factors that prompt and stand in the way of rigorous and relevant learning opportunities and high-quality student work.

The Nature of Teacher Assignments in New High Schools and Comprehensive High Schools

The data examined by this report suggest that new high schools participating in the foundation's initiative are different from large comprehensive schools, not just in their size and structure but also in terms of the learning opportunities provided in their classrooms. We have found that:

- ***Assignments given in the new high schools are more rigorous than the assignments given in the comprehensive high schools.*** English/language arts assignments in new schools are more likely to entail the construction of knowledge and elaborated communication. Mathematics assignments in new schools also tend to be more rigorous than those in the comprehensive high schools, but the difference is very small.
- ***Assignments in the new high schools place a strong emphasis on embedding learning opportunities in real-world settings and giving students a voice in shaping these opportunities.*** English/language arts and mathematics assignments in these schools are more likely to have real-world connections and to incorporate elements of student choice, compared with assignments given by teachers in the comprehensive high schools.
- ***Rigor and relevance are not incompatible.*** Most rigorous assignments are also relevant.
- ***Teacher feedback on students' work in English/language arts is more informative in new high schools than in comprehensive high schools, though there is room for improvement in both.*** There is also much room for improvement in the quantity and quality of mathematics teachers' feedback in both types of schools.

The Quality of Student Work in New High Schools and Comprehensive High Schools

The data also reveal differences between the quality of student work produced in the two types of schools.

- ***Students in the new high schools do higher-quality work in English/language arts, compared with students in the comprehensive high schools.*** In the new high schools, student work in English/language arts is more likely to demonstrate a deep conceptual understanding of content, clear communication, facility with language, and the construction of new knowledge.
- ***Students in the comprehensive schools do higher-quality work in mathematics, compared with students in the new high schools.*** In their class work, students in the comprehensive schools are more likely to show that they know and understand important math content, are facile with mathematics conventions, and possess skill and understanding in problem solving and reasoning.
- ***In English/language arts, both rigor and relevance in assignments lead to higher-quality student work,*** and relevance is the more important of the two. Assignments with a strong emphasis on embedding learning opportunities in real-world settings and giving students a voice in shaping their assignments lead to better-quality student work.
- ***In mathematics, rigor matters for student work quality, but relevance does not.*** Relevance in mathematics assignments is not correlated with the quality of student work.
- ***Students who do higher-quality work in school do better on standardized achievement tests,*** although the results are not strong. We see some support for this relationship, but the relationship hinges on what the tests measure and the content of the courses.

We do not yet have samples of assignments from the same schools over multiple years. Hence, we cannot draw conclusions about trends in the rigor and relevance of teachers' assignments over time as teachers spend more time in innovative new schools. We also cannot address differences in the quality of the resulting student work. We are currently collecting additional assignments and student work to address this question. Next year's report will look at trends over time in these foundation-affiliated schools and will compare assignments and work with artifacts of teaching and learning from nearby comprehensive high schools unaffiliated with the foundation.

Implications for the Initiative[1]

From these findings, we derive several implications.

- ***Professional development around teaching practices that incorporate both rigor and relevance is critically important for faculty in reforming schools.*** In almost every school, teachers asked for help in developing and honing their practice. They lamented the limited availability of useful professional development materials, offerings, and coaching. Especially important is professional development on the implementation of innovative practices within the context of current federal and state accountability requirements.

- ***Not every intermediary organization and district has a clear vision of effective instruction and appropriate curricular materials for high-need student populations.*** The organizations receiving foundation funds vary markedly in their histories and the nature of their expertise (AIR/SRI, 2003, 2004). Some of the organizations specialize in supporting a grassroots planning process rather than in providing instruction. Some have a commitment to a given instructional approach (typically project-based learning) but believe that teachers should develop the curriculum materials to use with this approach based on their particular students' interests and needs. Schools that do not receive curriculum resources from their grantee organization have the added burden of developing curriculum while trying to start or redesign a school; some resort to packaged software or traditional district textbooks.

- ***Teachers need compelling illustrations of the kinds of rigorous, relevant assignments that students with backgrounds similar to those of their own students can do.*** Although teachers have concerns about their students' academic preparation, there is proof from some of the classrooms in this study that low-income, historically underserved students can rise to the challenge of highly relevant, rigorous assignments. Good models of high-quality, relevant assignments and student work are needed to support teachers' work and stimulate students' efforts. The presence of tools per se, however, is rarely sufficient to bring about changes in instructional practice. Teachers need supported time for interacting with the resources and, ideally, with other teachers who have used and are currently using them.

- ***Innovative schools could learn from examples of mathematics assignments that are both rigorous and relevant.*** Given many teachers' concerns about the quality of their efforts in mathematics, the foundation might want to consider funding an online library of mathematics assignments that have real-world connections and are linked to high standards, supplemented with professional development. There are lessons to be learned from the new schools where mathematics assignments are both rigorous and relevant. One possibility would be to train master teachers at each school using the rubrics developed for the teacher assignment and student work scoring as a starting point for discussing real assignments given by teachers. Many of the schools started under this initiative stress a project- or problem-based approach that is theoretically compatible with teaching mathematics concepts and skills but difficult to reconcile with the specifics of district and state standards for algebra and geometry courses. Rather than expecting individual grantees or schools to solve this problem for themselves, the foundation could fund an organization with instructional development expertise in secondary mathematics to develop materials and facilitate the sharing of information among teachers.

- ***Foundation-affiliated schools should offer structured mathematics classes, rather than relying primarily on mathematics learning through internships or unstructured projects.*** Incorporating mathematics into internships and student-designed projects is difficult because of the specific content that students are expected to learn in high school algebra and geometry. Students' college prospects will be better served by mastering the required high-level content than by relying entirely on internships or unstructured projects that incorporate mathematics at a more basic level (such as basic operations, percentages, and averages, as opposed to linear equations and rates of change). Some schools are successfully integrating high-level mathematics content with science in their project-based curriculum. These projects are structured, have been refined over several years in the classroom, and are supported with instructional materials and guides for implementation.

- ***Schools serving high-need students should provide supplementary academic supports,*** including academic coaching during the day and after-hours homework support. School staff cited lack of tutoring services as a barrier for many of the students in foundation-affiliated schools. In some sites, business and community partners provide mentors or tutors. Federal and state programs supporting after-school academic activities are a potential source of funding. Private nonprofit and community-based organizations are potential partners in providing such supports.

Introduction

The Need for Change in Secondary Education

America's high schools offer strikingly different experiences for different students. While some secondary students are tackling not just college-preparatory but also college-level work, others take general or remedial courses that do not provide them with the skills required for higher education. Low-income and minority students are particularly likely to either drop out or take courses that leave them underprepared for higher education. Little more than half of the African-American and Latino youth who start ninth grade finish high school with a diploma. Fewer than 30% are ready to enter higher education without doing remedial work (Green & Winters, 2005).

The Bill & Melinda Gates Foundation's High School Grants initiative seeks to catalyze the creation of a new kind of American high school—one where *every* student has a challenging academic program that provides sound preparation for college and for family-wage jobs and the demands of good citizenship. The foundation believes that to better serve all students, high schools need to become places that combine rigor in the academic program of each and every student (not just those in an honors or higher track) with relevance to their interests and potential career choices, supported by positive relationships that can inspire students both academically and personally.

The foundation's national high school reform work began in 2000–01 with the award of grants to 12 nonprofit organizations charged with creating high schools that would embody these ideals. These intermediary organizations received grants to establish high schools, either by starting new schools or by redesigning existing comprehensive high schools into smaller, more focused units (separate "learning communities" or "academies"). In either case, the foundation expected the resulting schools to be characterized by a coherent vision, high expectations for all their students, and powerful teaching and learning. These expectations were articulated by the foundation in the form of seven attributes of high-performing high schools and three attributes of powerful teaching and learning.[2] The foundation expected that schools created under its initiative would be small in size (typically no more than a hundred students per grade) not because "small" was an end in itself, but because the greater personal attention that comes with a small size is conducive to implementation of the practices and climate of effective schooling.[3]

The foundation has recently expanded its work to include additional reform organizations and jurisdictions. It encourages the states and districts in which it works to create diversified portfolios of schools with different types of instructional models and themes so that students and families can choose the school designs that best meet their needs. The foundation envisions a mix of traditional, theme-based, and student-centered schools in its jurisdictions, with traditional schools providing a rigorous, college-preparatory treatment of core subjects; theme-based schools organizing instruction around themes, such as social justice or science and technology; and student-centered programs building individualized learning plans where students work on interesting, challenging projects designed with help from adults.

Sidebar: Three Types of School Designs

	Description	*Examples*
Traditional	These schools teach traditional subjects, but focus on rigorously preparing every student for college or work.	LaGuardia Middle College – Early College High School, New York, NY; Lionel Wilson College Preparatory Academy, Oakland, CA; Cristo Rey Jesuit High School, Chicago, IL; Frederick Douglass Academy, Harlem, NY
Theme-based	These schools organize coursework around a theme—such as the sciences, technology, or the arts—to engage students in a college-preparatory curriculum.	High Tech High, San Diego, CA; Tacoma School of the Arts, Tacoma, WA; Zoo School, Minneapolis, MN; Boston Arts Academy, Boston, MA; Expeditionary Learning Schools; Francis W. Parker Charter Essential School, Devon, MA; Fenway High School, Boston, MA
Student-centered	These schools create individualized plans for each student, often with students' input, and may focus especially on dropouts or at-risk youth.	The Met, Providence, RI; Minnesota New Country School, Henderson, MN; Maya Angelou Public Charter School, Washington, DC; Portland Community College, Portland, OR

Source: Bill & Melinda Gates Foundation (2004).

Focusing on Relationships

Our previous evaluation report, *Creating Cultures for Learning: Supportive Relationships in New and Redesigned High Schools*, (AIR/SRI, 2005), examined the experiences of the first set of schools opened or preparing for redesign under the foundation's national education initiative. It evaluated the degree to which effective school environments were successfully implemented through the projects funded by the foundation. The report presented data suggesting that new, foundation-affiliated high schools were implementing the foundation's effective-school attributes and providing more positive school climates than traditional, comprehensive high schools. They reported that the new schools in the study had more positive school climates—in terms of both personalization of the learning environment and a common focus among teachers and students—than was found in the comprehensive high schools. The new schools also were marked by stronger relationships between students and teachers and by more staff collaboration and participation in decision-making than was typical of the comprehensive schools.

These differences between new and comprehensive high schools are encouraging, especially in light of the fact that the new schools generally enrolled higher proportions of students who were eligible for free or reduced-priced lunch and who were members of a racial/ethnic minority group. They also generally enrolled greater percentages of students with special education or English language acquisition needs. Additionally, the new high schools generally enrolled students who began high school academically behind students attending the comprehensive high schools.

In comparing the work of new schools in their first, second, and, in a small number of cases, third years of operation, the previous evaluation report also noted that after a strong initial year, several schools experienced a "second-year slump"; the family atmosphere that both students and adults described in their initial year gave way to growing pains as a new set of students and teachers came on board the second year. Both teachers and students described a change in social dynamics in their second year as schools doubled in size with the addition of a new class of ninth grade students and the teachers to work with them. Although still generally very positive about their schools and their teachers, students in second-year schools felt less special than students did the first year. In terms of teacher community, some schools experienced a schism between those teachers who had been the pioneers in the school's first year and the newcomers joining the staff later. Although the third-year school sample was small, it appeared that the schools recovered in terms of school climate in their third year of operation.

In the previous evaluation report, we also followed comprehensive high schools from their planning year into their first and second years of redesign and operation as small learning communities or academies. We found that the learning communities that resulted generally offered more positive climates than the comprehensive high schools from which they emerged. Comparing school climate and interpersonal relations in the smaller units with those of comprehensive schools in the year prior to redesign, evaluators found that implementation of the foundation's ideals had improved. The biggest positive change reported by students and staff was an improvement in the relationships between and among students and teachers. This change was particularly noteworthy because the people in the building—the student population and teaching staff—were basically the same groups that populated the earlier comprehensive high school.

The current report complements the earlier report's description of school-level changes with an examination of classroom teaching and learning opportunities. It focuses on instructional practice—an important area of investigation, given the fact that many carefully implemented education reforms have prompted only superficial changes in the classroom (e.g., Berends, Bodilly, & Kirby, 2002; Bodilly, 1988; Smith et al., 1998; Stringfield, Millsap, & Hermann, 1997). This report takes an in-depth look into the classrooms of foundation-affiliated schools, examining teaching and learning to see whether students in these schools are exposed to challenging learning opportunities and, if they are, whether students rise to the challenge and produce intellectually complex work.

Emphasizing Rigor and Relevance

The foundation's ideas about teaching and learning are not unprecedented. They rest on a knowledge base that includes two decades of learning research, data from proven school models, and lessons from exemplary established schools (Bransford, Brown, & Cocking, 1999; Newmann & Wehlage, 1995; Doran & Drury, 2002; Bottoms, et al, 2004; Bill & Melinda Gates Foundation, 2004, 2005). The foundation urges school leaders to make learning rigorous and relevant by:

- Building a culture of high expectations and academic challenge.
- Aligning curriculum, instruction and assessment with college admission standards.
- Creating opportunities for in-depth exploration of topics.
- Involving students in decisions about what and how they learn.
- Creating learning experiences that emphasize real-world connections and that relate to students' lives and aspirations.
- Setting clear learning goals.
- Providing intensive academic support so that all students perform to high levels.
- Regularly assessing and providing informative feedback on student products.
- Monitoring progress through multiple measures, including performance-based and standardized assessments.

The relative importance of these prescriptions has shifted over time, with some gaining and others waning in importance in the foundation's grant making. However, the overriding press for rigorous and relevant learning opportunities and intellectually complex student work persists.

Foundation-affiliated schools have put a variety of structures in place to achieve these ends. As was mentioned earlier, some foundation-affiliated schools organize instruction around themes or individualized learning plans. Some support personalized instruction with structures such as advisories, multiyear teaching relationships, mentoring programs with business and community leaders, multidisciplinary courses, supplementary academic coaching, block scheduling, outside internships, and community service. School leaders have implemented these structures to help strengthen students' learning opportunities and provide a basis for rigorous and relevant instruction.

This report describes the learning structures that are in use in foundation-affiliated schools and examines the opportunities these structures provide. It examines the rigor and relevance of learning tasks and the quality of student work. The research is guided by an analytic framework developed by researchers examining school reform begun under the Chicago Annenberg Challenge (Newmann, Lopez, & Bryk, 1998; Bryk, Nagaoka, & Newmann, 2000; Newmann, Bryk, & Nagaoka, 2001). Their framework says that learning opportunities prompt intellectually complex work when they call for:

- The application of basic knowledge and skills to new, real-world problems.
- Problem solutions that require organization, interpretation, evaluation, and synthesis of information.
- Solutions grounded in solid information, concepts, and principles from the academic disciplines.
- Effective communication of conclusions.

The current work expands on the Chicago framework, which was developed for research in elementary and middle school settings, to incorporate secondary settings, where learning strategies are more varied, students bear greater responsibility for learning, and students have more frequent opportunities to refine their work based on teachers' feedback. We expand the model to include two additional important aspects of the foundation's tenets for teaching and learning:

- The choices that students make about what they will study and how they will demonstrate mastery.
- The opportunities and information that students receive to revise and improve their work.

With these additions, the model describes the types of intellectual demands the foundation seeks for students (AIR/SRI, 2004b). The expanded framework supports questions about the rigor of students' learning opportunities and the relevance of instruction. Table 1 summarizes the expanded framework and describes the characteristics of rigorous and relevant learning opportunities.

Table 1. Instructional Rigor and Relevance

Rigor	*Relevance*
Assignments ask students to ... ◆ Move beyond the reproduction of information, taking what they already know and can do and using their knowledge and skills to create or explore new ideas. ◆ Demonstrate conceptual understanding of important content. ◆ Organize, interpret, evaluate, and synthesize information. ◆ Communicate clearly and well. ◆ Revise work based on informative feedback.	Assignments ask students to ... ◆ Address questions or problems with real-world applications. ◆ Make choices about what they will study and how they will study it. ◆ Take on plausible writing roles and submit their work to real audiences.

Research conducted in the Chicago Annenberg Challenge sites and in other schools suggests that learning opportunities marked by these characteristics are likely to prompt intellectually complex student work. It suggests that when instruction makes high intellectual demands, students are more likely to rise to the occasion and produce high-quality work. This research suggests that rigorous, relevant learning opportunities open the door for work that reflects students' deep understanding, includes and explores students' new ideas, demonstrates reasoning and problem-solving skills, communicates well, and correctly applies procedures and conventions. Table 2 summarizes the characteristics of high-quality student work.

Table 2. High-Quality Student Work

Student products demonstrate ... ◆ The use of current knowledge and skills to create and explore new ideas. ◆ Deep conceptual understanding of important content. ◆ Reasoning and problem-solving facility. ◆ Extended, coherent, and well-organized communication. ◆ Effective use of language and mathematics conventions.

Research Questions

The foundation's intentions for teaching and learning in its schools and this expanded framework provide important background for the current research. The data presented in this report were gathered to answer five evaluation questions:

1. How do the rigor and relevance of learning opportunities in new high schools compare with the rigor and relevance of learning opportunities in comprehensive high schools?
2. How does teacher feedback on students' work in new high schools compare with the feedback provided by teachers in comprehensive high schools?
3. How does the quality of students' work in new high schools compare with the quality of students' work in comprehensive high schools?
4. To what extent are rigorous and relevant assignments associated with high-quality student work?
5. To what extent do differences in the quality of student work relate to school differences on jurisdiction-sponsored achievement tests?

In answering these questions, qualitative data are used to help explain some of the factors that prompt and stand in the way of rigorous and relevant learning opportunities and high-quality student work.

Organization of the Report

The remainder of this report is divided into seven sections. The next section describes the design of the study and the plan for data collection. The section that follows compares the rigor and relevance of assignments in new schools with that in comprehensive high schools. The third section provides detail on teacher feedback in new and comprehensive schools. The fourth section compares the quality of students' work in new schools with the quality of students' work in comprehensive high schools. The fifth section studies the relationships between assignments and work, testing the assumption that challenging learning opportunities prompt high-quality student work. The sixth section relates the quality of students' in-class work to students' achievement test results. The final section of the report reviews the findings and discusses the implications of these results for the foundation's work with new and redesigned schools.

Study Design

Study Sample and Data Collection

A team of researchers from the American Institutes for Research (AIR), SRI International (SRI), and Fouts & Associates have been evaluating the foundation's high school reform initiative since 2000. The bulk of the evaluation activity examines the initiative's progress at the national and state levels rather than at the levels of schools and intermediary organizations designated to help establish and support the high schools. The evaluation seeks to explore and test the idea that schools with the characteristics described by the foundation yield better and more equitable outcomes for students. It evaluates the degree to which the school environments exhibit the attributes desired by the foundation and describes factors that are keys to this type of school success.

The data that are described in this report were collected from teachers and students in 24 foundation-affiliated schools. Twelve of the 24 sites are new schools, and 12 are traditional comprehensive high schools. The 12 new schools are located in 8 districts across the country. Eight of the 12 comprehensive high schools operate in Washington state, where the foundation initiated its high school reform work, and 4 are located elsewhere across the country. All 12 of the comprehensive sites are scheduled to redesign their schools in the very near future, assigning students and teachers to small learning communities or academies. The evaluation team will follow these schools and the new high schools, returning to all 24 sites for a second wave of data collection to examine advances in instructional practice as new schools move into their second and third years and as comprehensive high schools redesign into several smaller organizational units. It must be noted that the 12 comprehensive schools described in this report do not provide adequate matched-comparison data for the 12 new schools.[4]

In general, the new schools in the sample are populated by students at higher educational risk than students in the comprehensive high schools in the study; that is, there are higher proportions of students receiving free and reduced-price lunch, who have special learning and language needs, who had lower achievement test scores as freshmen, and who come from underserved minority backgrounds. The faculties in the new schools have fewer years of teaching experience than faculty in the comprehensive high schools. On average, the comprehensive high schools from Washington state have more advantaged student populations and more experienced faculty than the new schools and the comprehensive high schools in other states. The technical appendix of this report includes additional information about the characteristics of students and adults in the new and comprehensive high schools in the sample.

The characteristics of the instructional programs in new schools in the sample also differ from those of the comprehensive schools. Table 3 shows how many new schools are implementing traditional, theme-based, or student-centered instructional models. The data in Table 3 also show that academic structures that support personalized instruction are relatively common in the new schools; for example, these structures include advisories, individualized learning plans, project-based learning, supplementary academic coaching, and multiyear relationships with teachers. Additionally, the table shows that multidisciplinary courses, student internships, and community service are common components of academic programs in new schools.

Table 3. Characteristics of Instructional Programs in New High Schools[5,6]

		New Schools (n=12)
School designs	Traditional	33%
	Theme-based	33%
	Student-centered	33%
Structures that support personalized instruction	Advisories	78%
	Multiyear relationships with teachers	78%
	Mentoring by business, community leaders	56%
	Individualized learning plans	56%
	Project-based learning	78%
	Academic coaching	67%
Instructional schedules and structure	Traditional class schedules	22%
	Block scheduling	56%
	No class schedules	22%
Course offerings	Basic skills instruction	67%
	Advanced Placement, International Baccalaureate, honors classes	44%
	Multidisciplinary instruction	56%
	Local college and university classes	89%
	Online courses	56%
Outside requirements	Mandatory internships	67%
	Mandatory community service	67%

The evaluation team collected naturalistic data in the form of assignments given in class and work produced by students from the 24 new and comprehensive high schools to examine the rigor and relevance of assignments and the quality of student work in the two sets of schools. The courses from which English/language arts data were collected for this study included sophomore English, honors English, remedial English, humanities, and multidisciplinary or project-based learning sequences. The mathematics courses included algebra 1, geometry, algebra 2, honors algebra 2, trigonometry, integrated math, and multidisciplinary or project-based learning sequences.

At eight points during the school year, up to three faculty members who taught 10th-grade English/language arts and up to three who taught 10th-grade mathematics in each of the 24 schools submitted copies of the assignments they gave to their students. Teachers submitted four assignments that were typical of the work their students did on a day-to-day basis and four assignments that challenged their students to show what they knew and could do at high levels. For three of the assignments, teachers submitted the work produced by a random sample of eight students. The data collection was revised for schools where students design their own projects (project-based learning). In these schools, students work with their teachers or mentors to come up with a project proposal, which is essentially the assignment they need to complete. In these schools, teachers submitted three sets of proposals (assignments) and student work (projects) for the random sample of eight students.

The evaluation team also conducted site visits in a subsample of 9 of the 12 new schools, observing classrooms, interviewing teachers and other school leaders, and conducting focus groups with students. The 9 schools were selected to be representative of the characteristics of the 12 new high schools. See the technical appendix for additional detail about sampling and data collection.

Measures and Analyses

In the summers of 2003 and 2004, project staff worked with 24 master 10th-grade English/language arts and mathematics teachers to score the assignments and work against a set of scoring criteria originally conceived by Newmann and Bryk (Newmann et al., 1998; Bryk et al., 2000; Newmann et al., 2001). To align them more closely with the foundation's intentions for teaching and learning, the scoring criteria were refined to reflect the earlier-described revisions to the Chicago conceptual framework. The individual scoring criteria are described in later sections of this report.

The scores that were generated at the summer sessions were then combined by using Many-Facet Rasch Measurement (MFRM), where the individual scores that measured assignment rigor were used to estimate a total rigor measure, and similarly for assignment relevance and student work quality.[7,8] The estimated rigor measure characterizes the intellectual demands of the assignments teachers gave in English/language arts and mathematics; the relevance measure assesses the extent to which assignments have real-world applications and involve students in decision-making about what and how they learn. The student work measures describe the quality of students' efforts.

For assignments and work, the Rasch scores were divided into four reporting categories. Categories described assignments with *substantial* rigor, *moderate* rigor, *limited* rigor, or *little or no* rigor. Analogous scales were created for assignment relevance and student work quality. These reporting categories are used to present findings for the 24 schools; they support comparisons between the rigor and relevance of learning opportunities and the quality of student work produced in new and comprehensive high schools. The technical appendix describes the scoring criteria; the rigor, relevance, and student work measures; and the statistical properties of the original and summary scales.

An analytic technique called hierarchical linear modeling was used to test the statistical significance of the differences between rigor, relevance, and student work quality in new and comprehensive high schools.[9] This technique also helps us isolate the sources of differences between school types. In comparing assignments and work in new high schools with assignments and work in comprehensive high schools, this technique helps avoid the inadvertent attribution of observed school differences to differences in instructional practices when they are due to some other related factor, like students' educational histories, English-language facility, special learning needs, and demographic characteristics. Analyses also controlled for teachers' experience and whether assignments were typical or challenging.

Following the same logic, student work estimates were adjusted for students' ninth-grade achievement test scores, gender, and race/ethnicity. Estimates were also controlled for the proportions of students receiving free and reduced-price lunch and students from underrepresented minority backgrounds in each school, as well as the assignment type. See the technical appendix for additional detail abut the analyses.

Scoring Criteria for English/Language Arts Assignments

Assignments and other artifacts of instruction gathered in the 24 schools were examined to see whether they provided students with rigorous and relevant learning opportunities. To gauge the rigor of teachers' English/language arts assignments, assignments were scored on the extent to which they required students to go beyond reproduction of knowledge to create or explore new ideas, communicate clearly and well, and use language conventions accurately and effectively. Assignments were scored for relevance on criteria that examined the extent to which assignments called for work with real-world applications and asked students to make choices about what and how they would learn. Table 4 provides additional detail about the scoring criteria for English/language arts.

Table 4. Scoring Criteria for English/Language Arts Assignments

Rigorous assignments ...

- Call for student work that moves beyond the mere reproduction of information to the *construction of knowledge.* Assignments that emphasize construction of knowledge require students to do more than summarize or paraphrase information they have read, heard, or viewed; these assignments require students to take what they already know and use that knowledge to create or explore new ideas through interpretation, analysis, synthesis, or evaluation of information. Some assignments ask students to construct knowledge and then to use this new knowledge to generate additional new understandings.
- Emphasize *elaborated communication,* prompting extended writing and asking students to make assertions and support them with evidence. These tasks ask students to make an assertion by stating a claim, drawing a conclusion, and/or suggesting a generalization, and then to support the assertion with evidence.

Relevant assignments ...

- Emphasize *real-world connections,* prompting students to take on plausible writing roles, go beyond the demonstration of academic competence to achieve real-world purposes, and submit their work to real audiences other than the teacher or other students.
- Call on *students to make choices* about what they will study and how they will demonstrate mastery. This criterion examines the extent to which students partner with faculty in crafting tasks that meet students' instructional goals. Scorers also look for teachers' guidance on how students make choices about topics and methods.

Exhibit 1 helps to illustrate the scoring criteria. It presents an English/language arts assignment with substantial rigor and relevance. The assignment is a student-generated proposal for an internship-based project in science.

Exhibit 1. English/Language Arts Assignment with Substantial Rigor and Relevance

This assignment was designed by a student with input from his advisor and internship mentor. The student wrote a project proposal that addresses the school's learning goals and follows the school's guidelines for developing project proposals.

The Safety of Salmon

Introduction

My project this quarter is to research the alleged health risks associated with eating salmon. I am going to write a paper about salmon. I am also going to teach the servers at [my internship site] about salmon and customers' perceived safety issues. I also am going to put together a little card that talks about the safety of salmon. That card will be passed out to customers if they ask if it is still safe to eat salmon.

It will benefit my internship in different ways. It will help the servers know more about safety of salmon, also they will have some information on the safety to hand out to the customers, and the video will also help new servers when they come in.

Focus
The main question that I hope to answer is, Are there health risks from eating salmon?

Learning goals

a. Communication: I am going to talk to a lot of different people. I am going to go to the place that the restaurants get their seafood. I also am going to do research online, reading, writing, presentation.
b. Social reasoning: I am going to look up different people's opinions and figure.
c. Quantitative Reasoning: statistics
d. Empirical Reasoning: science content knowledge
e. Personal Qualities: Have all my work handed in on time.

Authenticity
I chose this project because it was going to help my mentor. Also I wanted to know why people are saying that salmon is bad for you to eat. This project will challenge me in different ways. It will challenge me to write another paper this quarter. It will expand my learning because I don't know anything about salmon.

Planning
I will do the presentation the week of May 17–21. We have not made a date that I am going to talk to servers just yet.

This assignment is rigorous because it prompts the student to construct knowledge about the health risks posed by salmon consumption. It calls for extended communication by asking the student to draft text for information cards on salmon consumption and write a script for a video on salmon handling. The assignment is relevant because its purpose and audience are authentic—that is, it addresses a real-world problem and goes beyond the classroom to include restaurant personnel and customers. In addition, the student participated in decisions about what he would study and how he would demonstrate mastery.

Scoring Criteria for Mathematics Assignments

Mathematics assignments and instructional artifacts were scored for rigor by using criteria that examined the extent to which they asked for deep understanding of mathematical content, mathematical problem solving and reasoning, and effective communication about problems and solutions. Relevance scores describe the extent to which mathematics assignments had relevant contexts and real-world applications and involved students in decisions about learning. Table 5 summarizes the scoring criteria for mathematics.

Table 5. Scoring Criteria for Mathematics Assignments

Rigorous assignments ...

- Call for student work demonstrating deep, *conceptual understanding of important mathematical content* in one or more of the important ideas in mathematics. These important ideas refer to the large and unifying ideas that help link smaller pieces of mathematics knowledge, that undergird procedural skills, and that connect mathematics within and between content domains. Among the important ideas that 10^{th}-grade assignments are expected to address are chance, dimension, change and growth, transformation, interrelationships, translation of problems from one representation to another, proportionality, and function and recursion. In addition, critical mathematical processes that support the development of these important ideas, such as creating proofs, making and justifying conjectures, and using models and varied representations, are considered essential ideas.
- Require *problem solving or reasoning,* asking students to formulate problems from situations, make generalizations, judge the validity of arguments, make models, and construct valid arguments and proofs. These go beyond assignments that require students to retrieve or reproduce fragments of knowledge or simply apply previously learned algorithms or procedures.
- Explicitly call for *effective communication* of mathematical understanding. Assignments that call for communication ask students not only to "show their work" (i.e., trace the solution path) but also to "explain or justify," providing insight into the clarity of the students' mathematical understanding.

Relevant assignments ...

- Ask students to address mathematical questions, issues, or problems similar to ones encountered in the experience of mathematicians and other professionals who use mathematics to solve problems; in other words, they have a *relevant context and real-world connections.* In addition, scorers examine the extent to which assignments specify an "authentic audience" for student work products.
- Allow *student involvement* in deciding which topics they will investigate, which problems they will study, and how they will tackle these topics and problems. Scorers also examine the extent to which assignments give students guidance in making choices about topics and problems that meet their instructional goals.

Exhibit 2 helps to illustrate these measures by showing a mathematics assignment that scored high on rigor and moderately high on relevance. The mathematics assignment is an in-class, whole-class assignment that calls for individual student products.

Exhibit 2. Mathematics Assignment with Substantial Rigor and Relevance

This is a teacher-constructed assignment that focuses on survey methods and survey bias.

Analyzing Surveys

To demonstrate your understanding of the survey process, you will be given some actual surveys and survey results to analyze and critique. You need to examine the survey process based on the major points covered in Unit 1. Consider yourself a consultant for the company distributing the survey. The company wants to make sure that the data they are getting is accurate so they can make meaningful decisions from the results. You need to give input on the following:

- How were the questions worded?
- Who makes up the population?
- What is an appropriate sample size?
- How was the sample selected (sample method)?
- How was the survey distributed and collected?
- How were survey results presented (if applicable)?
- Was bias avoided at each step?

Your input should consist of an explanation of what they are currently doing and suggestions for improvement. Be sure to support your suggestions with sound reasoning from the unit. If not enough information is provided or an area is not addressed, determine what questions you would need to ask the company in order to address that particular area and then make a recommendation.

Since you are a consultant hired by the company, your report needs to be presented in a professional manner. In other words it should be word-processed and have a cover page.

The Survey Descriptors Are:

1. As part of the redesign process [school] is gathering information from staff, students and the community. Students were given a survey in their classes on [date]. Staff was given the same survey at the staff retreat over the summer (retreat was optional). The survey was available in two formats at curriculum night for parents. For those attending the information session in the [Parent Advisory Committee], surveys were on the seats. Students were also available in the commons to provide the surveys to the parents and explain the process.
2. In order to gather information about people's preferences regarding radio stations, a brief note was sent to the home of some members of the public. The form is to be completed and then mailed in at a certain time.
3. To gather parents' input regarding the school and PTA, the PTA sent home a questionnaire survey with all of the students at the school. The surveys are given to the teachers who then distribute them to the students, and then the students bring them home.
4. At Mervyn's, the sales representative asks customers to fill out a survey about the store and its service. The representative asks the customers when they are at the counter paying for their merchandise and mentions that there will be a drawing from all entries for a $100 gift certificate.

5. A company requesting information from students sends an envelope with multiple choice surveys to teachers throughout the country. They ask the teachers to distribute the surveys to their students and return the completed forms to them for analysis.
6. Puget Sound Consumers' checkbook sends questionnaires through the mail to various households. The questionnaire asks for opinions on a variety of service industries. Participants then mail their results in.

The teacher also included a New York Times article, "Americans on Iraq and on the Economy," with survey results from a New York Times/CBS NEWS Poll.

The assignment in Exhibit 2 received a high rigor score because it requires students to demonstrate conceptual understanding of survey analysis, survey bias, and sample size. Students can not complete this assignment merely by reproducing facts or by applying previously learned algorithms or procedures. It also calls for students to use logical and systematic thinking and reasoning to analyze the six surveys. Students are asked to provide well-developed summaries of the results of their reviews.

The assignment is relevant because it asks students to use mathematics with real-world applications. However, while the assignment provides opportunities for students to choose a survey from a set of six, the assignment specifies the method students will use to demonstrate understanding. Overall, the assignment receives a moderate score on relevance.

Scoring Criteria for English/Language Arts Student Work

The work that students produced in response to teachers' English/language arts assignments was scored by using criteria that described the extent to which students' writing was developed, coherent, and organized; moved beyond the reproduction of knowledge to explore new ideas; and used language conventions accurately and well. Table 6 provides additional detail about the scoring criteria for students' work in English/language arts.

Table 6. Scoring Criteria for Student Work in English/Language Arts

High-quality student work ...

- Demonstrates *elaborated communication* through extended writing that makes an assertion and then supports it with evidence. Student writing should be sufficiently developed, coherent, and well organized.
- Moves beyond the reproduction of information students have read, heard, or viewed, and demonstrates *construction of knowledge,* where students take current knowledge and use that to construct new knowledge, creating and exploring new ideas.
- Shows proficient and *effective use of language conventions and resources.* Student work should have spelling, vocabulary, grammar, and punctuation appropriate for 10th-grade work. Student work should demonstrate artistic use of language resources, including diction, syntax, imagery, and figurative language.

Exhibit 3 provides an example of student work that received high scores on the student work criteria.

Exhibit 3. English/Language Arts Student Work with Substantial Quality

This assignment asks students to write a thesis statement, reach fair and reasonable conclusions, demonstrate a sense of audience, write in paragraphs, and use correct grammar and spelling.

Assignment:

> Comparing similar things is something that we commonly do every day. We may compare two brands of tennis shoes, two compact cars, two universities, two careers, etc. We make these comparisons to decide which items to buy, which candidates to vote for, which restaurants to eat in, etc. The decisions that we make may be important to our purchases, our votes, our businesses, our ways of life.
>
> Choose two things that you have an interest in comparing. Gather information so that you may make a thorough and fair comparison of these items.
>
> Using the information that you have gathered, write an essay comparing the items that you have chosen. Your audience is your advisory. When you come to the conclusion of your essay, make fair statements based upon the information that you presented in your essay. Remember, your conclusion does not have to be totally one-sided.
>
> To organize your essay, use one of the two formats that you have been given.

Student Work:

Renting an Apartment vs. Buying a Home

It's a familiar scene: full luggage sits in the back of an open trunk, Mom bids you a tearful goodbye, Dad stands by, trying to be the "strong one," and you pull out of the driveway, on your way to life on your own. You're going to your new place, and despite your job, age, marital status, and education, you have to choose between an apartment and a house. Which one do you choose? Are you moving into an apartment complex with a nosy landlord and an attractive but loud neighbor? Or do you choose life in suburbia, weeding your rose garden and drinking lemonade on the front porch? There are several advantages and disadvantages of both options.

When someone chooses to rent, they have good and bad things to deal with. One of the major advantages is that a renter doesn't have to worry about maintenance. If something breaks or fails, the only thing a renter has to worry about is a phone call to their landlord. There are also no landscaping and only minimal cleaning responsibilities. But the downside to living in an apartment is that by not having that responsibility and that you are, in essence, only borrowing the house, you don't get to add your personality into your apartment. You can't do any interior design and only some limited landscaping.

Another plus and minus of renting is taxes. You don't have to pay all the outstanding taxes, like property tax, or deal with insurance payments or mortgage commitments. But a renter also has no chance for tax write-offs, and has no opportunity for long-term planning and investing. It is actually suggested that you rent an apartment unless you are planning or staying in a place for over seven years, hence "long-term" planning. So although a renter doesn't have to pay taxes, they don't have as much opportunity to see return investments, or even have tax write-offs.

One of the biggest disadvantages of renting an apartment, and probably the most complained about one, is the people you deal with while renting. Renters have to answer to a landlord and deal with paper-thin walls that allow sound to travel both ways. Landlords can be kind, sweet people, but can also be nosy, rude business people. Either way, you are paying off a building or even house for that person. And in an apartment, a renter doesn't have control over their neighbors. Conversations and arguments can be heard from both ways. The personal space is also limited, and can therefore cause problems easily with your neighbors.

Finally, the prices must be considered. In the [area where this student lives], prices can range from $600 a month to $1,250 a month, and the median price for a two bedroom, one bath ranges around $850. In [a close area], prices have an even wider range – from $700 to over $1800! The median price for a two bedroom, one bathroom apartment is $1000. But if a renter wants to move to an expensive neighborhood, for example [a city nearby], a low end apartment ranges from $945 to $1300, but cost can go as high as $2100 a month! But prices vary depending on the location and the size of an apartment. Overall, the average cost to rent an apartment, and just the rent alone, is around $950 per month.

The other option a person has to buy a home. Again, there are several advantages and disadvantages of buying. It is about the exact opposite of renting. What you don't have to deal with in renting, you do when you buy a home. And what is a problem in renting is almost nonexistent when buying. Once more, maintenance is a good and bad thing. The home owner is responsible for all repairs, painting the outside of the house. Mowing the lawns, and weeding a garden. But they also choose the style of the home decorations, landscaping, and appliances. There is more choice involved, but it is more costly to maintain a home.

One of the biggest downfalls of buying a home is the taxes. A homeowner pays property taxes, home insurance, mortgage payments, and interest. These taxes are what makes a home so expensive – more money goes into a home than into an apartment. But a homeowner can invest in their house and start long-term planning, gaining money back. Credit also tends to be better if you own a home, and pay the bills on time. But when buying a home, the potential buyer must realize that buying a home is a long term commitment. Eventually, it will be cheaper than renting an apartment for long period of time, but is expensive to buy a home.

The problem of neighbors and the neighborhood isn't really an issue when buying a home. One of the decisions that goes into buying should be neighborhood. Neighbors can still be noisy, but you don't live within four inches of your next door neighbor. Privacy can still be an issue, but a lot less so. Choosing your neighborhood and essentially your neighbors is an advantage renting doesn't offer.

Pricing must again be a factor in deciding whether to rent or buy. In the [local] area, houses start around $195,000 for a two-bedroom, one bathroom house. The average price is around $350,000. In [a neighboring city] houses can cost from $150,000 to as much as you are willing to pay, but the median price is $219,800. This is a high figure for a three bedroom, two bathroom house and anything smaller, but even so, there isn't much difference. [A city south of ours] has about the same situation, but prices start higher and eventually end higher, and the average price for a home is $379,200. Again, high for a small home, but still not an extreme difference. A solid average cost would be $310,000 in [our state].

Buying a home and renting an apartment are two very different things. If you total all of the payments for the first five years, you would pay around $11,400 a year for an apartment, and about $57,000 dollars in five years. On a twenty five year plan, just counting regular house payment without interest, someone would pay $12,500 for a year for a $310,000 home, and $62,000 dollars in five years.

In the end, after all is said and done, renting is about the same cost as buying, and buying a home is more worthwhile – you can sell it and get money back, and get to choose your neighbors. But renting an apartment isn't a bad idea, either. It costs less to start out, and a renter doesn't have to worry about unexpected living costs. Whether starting out, starting a family, or starting retirement, there are things to consider about both options. It really is up to the individual. Neither is really better than the other. In an apartment, you don't have to worry about maintenance or taxes, but don't get to put your own flair into your living space and you have to deal with people around you. When buying a house, you do have maintenance and taxes, but can do whatever you want to the house and choose your own neighbors. Both have a lot to offer, and either decision is a good one.

To successfully complete the assignment shown in Exhibit 3, students were required to select an issue and then develop arguments for two opposing vantage points. Students were asked to engage in extended writing. The work that is shown demonstrates elaborated communication, argumentation, and persuasion. The essay is well developed, with few grammatical or technical errors. The work is engaging and compels readers to read to the end. It received high scores on the student work criteria.

Scoring Criteria for Mathematics Student Work

The last set of scoring criteria describes the quality of students' work in mathematics. They measure how well students demonstrate deep understanding of important mathematics content, procedural knowledge about mathematics, mathematical problem-solving and reasoning skills, and effective communication about mathematical thinking. Table 7 describes the criteria in more detail.

Table 7. Scoring Criteria for Student Work in Mathematics

High-quality student work …

- Demonstrates a deep, *conceptual understanding* of important mathematical content in one or more of the important ideas in mathematics. High-quality student work provides evidence that the student could represent and classify mathematical entities; recognize, label, and generate examples and non-examples of concepts; use and interrelate models, diagrams, manipulatives, and varied representations; and identify and apply mathematical principles.
- Shows *procedural knowledge* of mathematical content, including knowledge of key skills and processes in 10^{th}-grade mathematics. Students should demonstrate procedural knowledge by selecting and correctly applying appropriate procedures, verifying or justifying the correctness of a procedure using concrete models or symbolic methods, or extending or modifying procedures to deal with specific factors in problems.
- Demonstrates skill and understanding in *problem solving and reasoning* by providing problem descriptions, determinations of desired outcomes, generation of appropriate models, selection of possible solutions and solution strategy alternatives, testing of trial solutions, evaluation of outcomes, and any needed revisions of solution steps and strategies. Student work that demonstrates mathematical reasoning involves evidence of logical, systematic thinking, which can be intuitive, deductive, or inductive reasoning, in making and justifying conjectures and solving problems. Reasoning often involves hypothesizing, predicting, analyzing, generalizing, synthesizing, or proving.
- Demonstrates *effective communication,* demonstrating organized and consolidated mathematical thinking through written and oral communication, coherent and clear communication of mathematical thinking, and the correct use of notation and terminology.

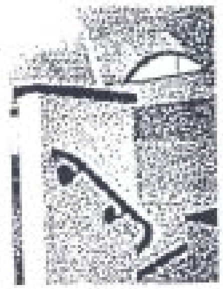
Exhibit 4 illustrates the scoring criteria by highlighting student work that received high marks.

Exhibit 4. Mathematics Student Work with Substantial Quality

The Vaulted Ceiling Problem

In the November 2000 issue of *This Old House Magazine* is an article about building a vaulted ceiling (which means curved) in a bedroom. At the end of the vaulted ceiling would be an "eyebrow window", which is a window that's curved on top with a straight bottom. Most modern eyebrow windows are just a slice off a circle.

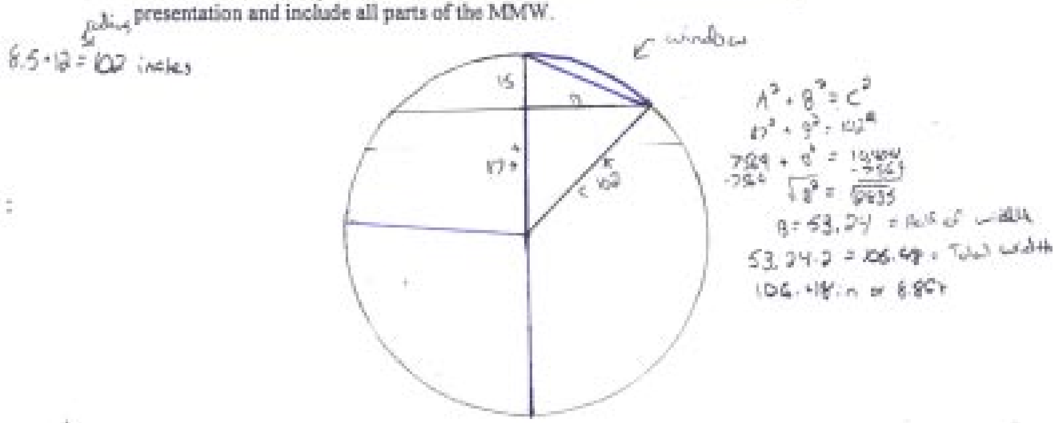
The picture shows the inside view of an eyebrow window at the top of some stairs.

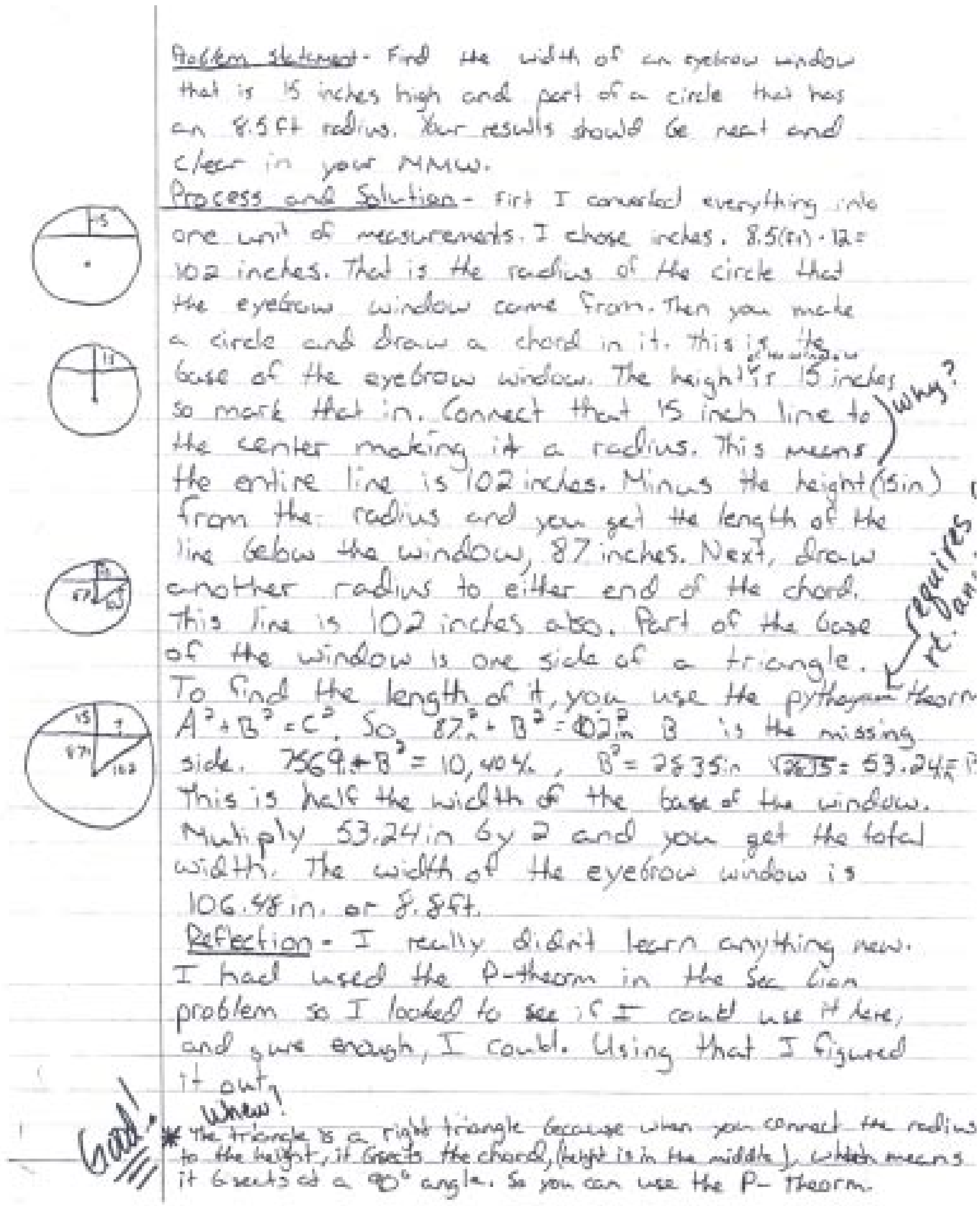


Here's the catch. They didn't know how wide the window was! Here's what they said in the magazine:

> [the contractor] needed to make the curve of the ceiling joists match that of the eyebrow window. When the unit didn't arrive in time, he was forced to rely on the manufacturer's promise that the shape would be equivalent to A 15-inch-deep slice off the top of a circle with an 8.5 foot radius.

Your job is to figure out how wide the base of this window will be. Use your circle facts and a compass and straight edge. Write your solution as a MMW. You do not need to word process the essay but you should be neat and clear in your presentation and include all parts of the MMW.





Problem Statement - Find the width of an eyebrow window that is 15 inches high and part of a circle that has an 8.5 ft radius. Your results should be neat and clear in your MMW.

Process and Solution - First I converted everything into one unit of measurements. I chose inches. 8.5(ft) · 12 = 102 inches. That is the radius of the circle that the eyebrow window come from. Then you make a circle and draw a chord in it. This is the base of the eyebrow window. The height of the window is 15 inches so mark that in. Connect that 15 inch line to the center making it a radius. This means the entire line is 102 inches. Minus the height (15in) from the radius and you get the length of the line below the window, 87 inches. Next, draw another radius to either end of the chord. This line is 102 inches also. Part of the base of the window is one side of a triangle. To find the length of it, you use the pythagorean theorm $A^2+B^2=C^2$. So, $87_{in}^2 + B^2 = 102_{in}^2$. B is the missing side. $7569_{in} + B^2 = 10{,}404_{in}$, $B^2 = 2835_{in}$ $\sqrt{2835} = 53.24_{in}$ = B

This is half the width of the base of the window. Multiply 53.24in by 2 and you get the total width. The width of the eyebrow window is 106.48 in. or 8.8ft.

Reflection - I really didn't learn anything new. I had used the P-theorm in the Sea Lion problem so I looked to see if I could use it here, and sure enough, I could. Using that I figured it out.

* the triangle is a right triangle because when you connect the radius to the height, it bisects the chord, (height is in the middle), which means it bisects at a 90° angle. So you can use the P-theorm.

why?

requires rt. an[gle]

Whew!

Good!

To complete the assignment successfully, students have to demonstrate their understanding of the geometric facts and theorems related to circles and clearly show their solutions. The student work in Exhibit 4 demonstrates clear conceptual understanding and procedural knowledge related to the relevant facts and theorems and is free of misconceptions and procedural mistakes. The student's problem-solving strategies and reasoning are appropriate and lead to the successful completion of the problem. The student's work includes a solution path with a complete and accurate explanation and justification of the student's thinking and conclusions.

Findings

The preceding information on the study design, data collection, and scoring provides the basis for the research questions and results that come next. Again, the evaluation team collected data to answer five questions:

1. How do the rigor and relevance of learning opportunities in new high schools compare with the rigor and relevance of learning opportunities in comprehensive high schools?
2. How does teacher feedback on students' work in new high schools compare with the feedback provided by teachers in comprehensive high schools?
3. How does the quality of students' work in new high schools compare with the quality of students' work in comprehensive high schools?
4. To what extent are rigorous and relevant assignments associated with high-quality student work?
5. To what extent do differences in the quality of student work relate to school differences on jurisdiction-sponsored achievement tests?

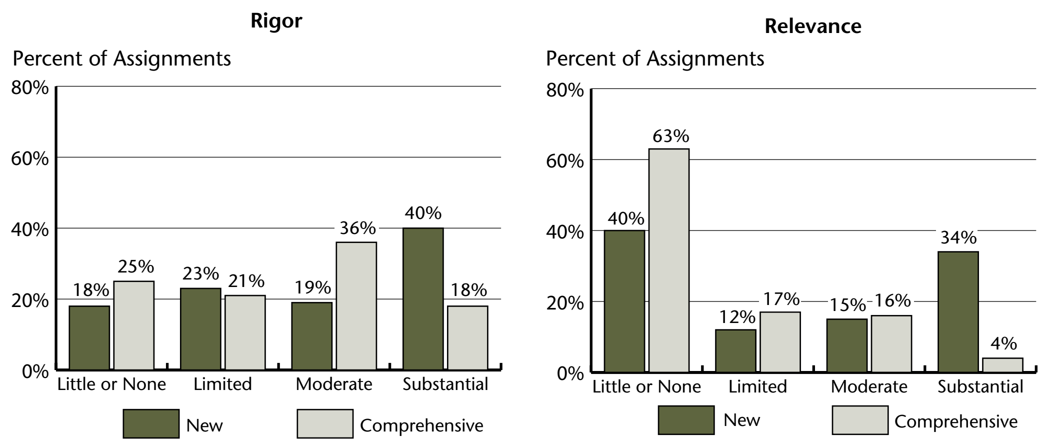
This section of the report answers each question in turn.

Rigor and Relevance in English/Language Arts

English/Language Arts Assignments are More Rigorous and Relevant in New High Schools than Comprehensive High Schools

The following figures contrast the rigor and relevance of assignments in new schools with those of assignments in the more conventional comprehensive high schools. The graphs in Figure 1 show the extent to which 10^{th}-grade students in foundation-affiliated new high schools and comprehensive high schools receive rigorous and relevant assignments in English/language arts. The left-hand graph shows the frequency with which students in new and comprehensive high schools get assignments with *little or no rigor*, *limited rigor*, *moderate rigor*, and *substantial rigor*, while the graph on the right shows the frequency with which students in the two types of high schools get assignments with *little or no relevance*, *limited relevance*, *moderate relevance*, and *substantial relevance*.

Figure 1. Rigor and Relevance of English/Language Arts Assignments in New High Schools and Comprehensive High Schools



The data in Figure 1 show that the 10th-grade English/language arts assignments submitted by teachers in new high schools were both more rigorous and more relevant than those submitted by faculty from comprehensive high schools. The left-hand graph shows that 40% of the assignments submitted by teachers in new high schools were judged by scorers to have substantial rigor, while only 18% of assignments submitted by teachers in comprehensive schools were. The right-hand graph shows that 34% of the assignments from new schools were rated by scorers as having substantial relevance, while only 4% of the assignments from comprehensive schools were. These data suggest that English/language arts assignments in new high schools are more rigorous and more relevant than those in comprehensive high schools; the differences between new and comprehensive schools are statistically significant.[10]

As was mentioned in the research design section of the report, one should exercise caution in attributing differences between the rigor and relevance of learning opportunities in new and comprehensive schools entirely to the reform efforts. It is possible that, in addition to the reform efforts, schools differed systematically in other ways that would have influenced instructional practices. Even with adjustments for differences between schools in student, teacher, and classroom characteristics, there could well be unmeasured differences that are related to differences in instructional practice across school types.

The sidebars in this section highlight some of the differences between rigorous and relevant assignments in 10th-grade English/language arts.

Sidebar: English/Language Arts Assignment with Substantial Rigor

This assignment was designed by a teacher to prompt students to use evidence from a text to support judgments and to explain how evidence supports the claim.

Report Card for *Brave New World*

Use the following "grades" to grade the society in *Brave New World:*

- Exceeding the standard
- Meeting the standard
- Approaching the standard
- Emerging competency
- No evidence

Assign grades to the society in four areas:

- Technology
- Moral and ethical development (sense of right and wrong, fairness, honesty)
- Social development (relations between people)
- Art

Work in groups to complete these tasks:

- Provide an explanation for the grade (3-5 sentences).
- Provide examples from the book (2-4 examples).
- Explain whether this society was more or less advanced than our own for this quality. How? Why?

Work individually to respond to the following questions:

With its advancements, what has this society gained? What has this society lost?

Explanation of scoring

To complete this assignment, it was necessary for students to think about the society in *Brave New World,* assign grades to selected aspects of the society, and make comparisons between the book's and their own societies. To justify the grades they assigned, students were asked to find support within the text. In reflecting on the advancements of modern society compared with those described in the book, students had to discuss the relative merits and/or deficiencies of the two societies.

This assignment scored high on rigor, receiving high scores on elaborated communication and construction of knowledge. It required critical-thinking skills and asked students to support their claims with material from the text. It called for extended writing. The assignment scored poorly on relevance. Although the assignment related the book to students' own society, it had little application outside the classroom and was not written for an authentic audience or purpose. Also, students had no opportunity to make choices about what and how they would study.

Sidebar: English/Language Arts Assignment with Substantial Relevance

This assignment was designed by a student with input from his advisor and other adults.

How to Run a Barbershop

Introduction

My project this quarter is going to be how a person runs a barbershop. The project is going to have a poster of pictures of my long term internship and a paper the money in and money out budgets… I'm interested in this because I want to run a barbershop one day and I choose this project because I want people to know what it takes to run a barbershop.

My project is real because…I'm talking to a person who owns a barbershop and he's my mentor. I'm going to take pictures of the shop when they are cutting hair and listing how many people are coming in. I'm going to make a list how much money comes in… My project is challenging because I have to find the total the barbershop spends in a month and how much money comes in and what supplies it needs and what does it take to run a barbershop. The questions I want to know are

- What does it take to run a barbershop?
- What supplies you really need?
- Is it hard to run a barbershop?
- Do you need to graduate from high school?

To complete my project I got to get the answers to those questions… My mentor will help me by explain to me how he runs a barbershop and help me on my project.

Explanation of scoring

This assignment scored high on relevance. It was relevant to the student's interests and goals. The student had an internship at a barbershop and was interested in opening a business of his own. The work that it prompted had application beyond the classroom.

However, the assignment scored low on rigor. The extent to which the student would construct new knowledge or produce extended writing was unknown. Although it is possible that the student would do rigorous work in response to this assignment, the proposal did not explicitly suggest it.

To understand what helps and stands in the way of rigorous and relevant instruction in reforming high schools, qualitative data were examined for a subset of participating schools. Analyses were structured to suggest some of the conditions and resources that characterize schools where students' learning opportunities in English/language arts are rigorous and relevant and those where learning opportunities are more limited. Interview and observation data were analyzed for three new high schools with high average rigor scores and three new schools with low average rigor scores. Data from interviews with teachers and other school leaders, from student focus groups, and from classroom observations were examined for high- and low-rigor schools. The data for high-rigor schools were compared with those for low-rigor schools to see whether and how they differed. Analyses were repeated for new schools with high- and low-relevance scores. Demographics, free and reduced-price lunch status, and achievement test scores also were reviewed for the six schools.

Conditions for Rigorous Instruction in English/ Language Arts

While the qualitative data suggested some similarities in the emphases of schools with high and low instructional rigor in English/language arts, conversations with faculty and students also suggested interesting differences. Faculty in both groups of schools described the importance of preparing students well for jurisdiction-sponsored tests and for college and university admission requirements. Teachers described the external pressures associated with accountability systems and with jurisdiction- or charter-sponsored tests. Faculty in both groups of schools talked about the need to systematically cover specified content and prepare their students for high school exit examinations and college entrance tests. They talked about the need to satisfy the accountability requirements of the *No Child Left Behind Act* and of their jurisdictions or chartering organizations. One school leader commented that solid performance on state tests would help his “school survive.” Another noted that impressive test scores would bolster her school’s recruiting efforts.

However, informants in high- and low-rigor sites differed in their descriptions of the extent to which their schools stressed four important design features: high academic standards for all students, individualized instruction, frequent assessment, and the provision of supplemental academic supports for struggling students. At one high-rigor school, a faculty member talked about the standards for moving from the 10^{th} to 11^{th} grades, saying:

> We laid it on them really hard…They had to meet very clear expectations… Students had two chances to pass…Advisors spent a lot of time with the 10^{th} graders. And students worked with peer coaches… Most of them stepped up. I was really strict…I said to them, these are the goals and you are held accountable to these goals…I said, “Guys, most of you probably won’t do this [meet standards] the first time around. It’s just a lot. So don’t expect it.”

In this high-rigor school, students who did not meet the 10^{th}-grade standards either repeated the 10^{th} grade, worked under a summer contract to make up the work, worked on 10^{th}-grade standards while beginning 11^{th}-grade work, or were counseled out of school.

At another high-rigor school, students talked about academic standards. One student explained that she and her friends were passionate about their work and approached it with seriousness. A second student added, "A lot of times in a comprehensive school, you can get away with smoke and mirrors. This school takes that away. The work schedule takes away your smoke."

In contrast, the leader of one of the low-rigor schools said:

> We have been good at reaching the students' hearts and preparing them to work harder, but now [we] have to support them and feed them with academic successes as they advance over the four years…Given the right circumstance, students can do so much more and can meet higher expectations, but they need to have it structured in such a way as to not be overwhelmed and give up.

A student described the standards at a second low-rigor school:

> I think that some kids don't take it [their schoolwork] seriously because they know you have to work harder to fail than you do to succeed. It is harder to fail here than it is to succeed. You can get almost an F in a class and still come out of this school. You can always repeat a class. I think that's why kids really don't take it seriously. They know you can always come back.

Many of the respondents from high-rigor schools went on to describe their goals for individualized instruction. In one high-rigor school, a teacher discussed personalized learning, saying, "It is one kid at a time here. It truly is. There is nothing that we do the same for all kids. You look at their interests, their ability, their time, and their home situations." Another explained:

> It is important to know students' strengths and weaknesses and their gaps in learning and their families. I push each student hard to pursue their passions…There has to be content in the learning plan that the student is excited about, as well as, the things that have to be worked on even if they don't want to. My favorite quote from a student is, "You get to work at what you're good at. And you have to work at what you're not good at because you can't hide."

None of the respondents from low-rigor schools discussed individualized learning plans; few discussed individualized instruction.

The faculty and students in high-rigor schools also provided detail on school-based assessment. They talked about frequent informal assessments, student exhibitions, and culminating assessments for grade-to-grade or lower- to upper-grade promotion. Teachers in high-rigor schools also described the academic structures that supported their programs, including ramp-up classes and other supports for students with low literacy skills, seminars that mix Socratic and direct teaching methods, project-based learning components, tutoring services, and after-school academic supports. (Additional detail about project-based learning is provided by a sidebar in this section.) Many of the teachers in high-rigor schools said they supplemented their instructional efforts with the help of outside experts, mentors from business and the community, student teachers, and parent volunteers.

According to informants from low-rigor schools, these structures and outside resources were used infrequently in their programs. Instead, English/language arts teachers from low-rigor schools more often described their instructional programs in traditional terms and lamented the quality of available curriculum materials and instructional supports. At all three low-rigor schools, faculty described plans to implement some or all of these structures moving forward but explained that they lacked the "bandwidth" at this point in their schools' development.

While the qualitative data do not allow us to fully describe the differences between high- and low-rigor schools or fully untangle the reasons for those differences, they do suggest some hypotheses about possible conditions for rigorous instruction in English/language arts. One hypothesis concerns school leadership and another concerns student motivation.

A striking difference between the schools in the high- and low-rigor groups is that all three of the low-rigor schools experienced significant leadership turnover during the first 2 or 3 years. Although the faculty in high-rigor schools had some turnover in their ranks, none had school leader changes and none had substantial teacher turnover. A teacher in one of the low-rigor schools explained that his school strayed from its instructional vision and staff temporarily ignored project-based learning, multidisciplinary instruction, and the establishment of community connections. He described the disruption that came with the departure of his school leader:

> Overall, because of the chaos and "hierarchy vacuum" that existed when we [the current co-leaders] took over school leadership, we have been attending to critical immediate issues out of necessity. Many of these are more tactical than strategic. Now that some of those crises are under control, [we] are now able to come up for air…and begin to think about larger issues of how the school should operate.

Sidebar: What Is Project-Based Learning?

Among the schools in this initiative that reported efforts to implement a common pedagogy across all classes, project-based learning (PBL) is the most commonly cited instructional strategy. Curriculum developers and teachers tend to use the term *project-based learning* to cover a wide range of approaches, including everything from semester-long group efforts resulting in complex products, such as a functioning robot or a museum exhibit, to individual research projects to classroom exercises in which students have some hands-on involvement with materials before writing a report.

Although there is no generally accepted definition of project-based learning, the Buck Institute for Education (n.d.), a leading proponent of the approach, defines PBL as follows:

> A systematic teaching method that engages students in learning knowledge and skills through an extended inquiry process structured around complex, authentic questions and carefully designed products and tasks.

An initiative promoting project-based learning in California schools identified six essential components of an exemplary project, saying they should:

- Be built around instructional objectives in the core curriculum.
- Have real-world connections.
- Be completed over extended time frames.
- Provide opportunities for student decision-making.
- Provide opportunities for student collaboration, but with each student making a unique contribution.
- Include assessment of products and processes.

A related pedagogy is problem-based learning. The Center for Problem-Based Learning at the Illinois Mathematics and Science Academy (n.d.) defines problem-based learning as follows:

> A curriculum development and instructional system which simultaneously develops both problem solving strategies and disciplinary knowledge bases and skills by placing students in the active role of problem-solvers confronted with an ill-structured problem that mirrors real-world problems.

In addition to sharing the same acronym (PBL), project- and problem-based learning share an emphasis on complex, multipart tasks requiring active engagement in solving a realistic problem. Both approaches motivate students' acquisition of knowledge by presenting them with interesting, complex problems or tasks that require knowledge building.

The chief difference between the two versions of PBL lies in the extent to which a problem (as opposed to a topic) is developed by the teacher or by a curriculum developer before students get involved. In project-based learning, students are likely to be involved in negotiating the nature of their project. Students are less likely to negotiate topics in problem-based learning. Given its more advanced structure, problem-based learning tends to occur over shorter time frames than project-based learning. The distinction bet ween the two approaches is fuzzy, however, and many educators will refer to the same activity interchangeably as "project-based" or "problem-based" learning, or simply "PBL."

Describing the consequences of leadership turnover at a second low-rigor school, a teacher said, "Right now, there is uneven instruction with pockets of greatness and pockets of mediocrity and pockets of crappiness."

The second hypothesis about the conditions for rigorous English/language arts instruction concerns unmeasured differences in the students and the instructional programs of high- and low-rigor schools. A plausible hypothesis for differences in instructional rigor between schools is that high-rigor schools are populated by students with higher motivation levels than students in low-rigor schools. It is possible that these highly-motivated students work to higher standards and with greater autonomy than other students. A counter-explanation is that, in time, students in stronger instructional programs become more facile at taking responsibility for learning and working to high standards.[11] Examination of the demographic and academic data for the two groups of schools shows that students in high-rigor schools have slightly higher ninth-grade test scores and fewer come from underserved minority backgrounds and participate in free- and reduced-price lunch programs. However, also already mentioned, even when analyses control for these characteristics, the differences between high- and low-rigor schools persist and the compositions of the high- and low-rigor groups of schools are unchanged.

Conditions for Relevant Instruction in English/ Language Arts

Interview and observation data were also analyzed to uncover some of the conditions and resources that characterize new high schools where English/language arts assignments have substantial relevance to students' interests and experience, and those where assignments have limited relevance. As they were for the rigor analyses, qualitative data were analyzed for the three new schools with the highest average relevance scores and for the three new schools with the lowest average relevance scores.[12]

Interviews with faculty and students in the two sets of schools suggested three important differences between the instructional programs of high- and low-relevance schools that echo the ideas in the scoring rubrics. The two groups of schools differed in the extent to which students' interests drove the work they did, in the extent to which students were involved in decisions about how they would do their work, and in the extent to which students' work had real-world connections.

Interview and observation data for high-relevance schools provided numerous examples of student-initiated or teacher-guided projects on topics that appealed to students' interests or experience. One teacher explained that an important part of his job is taking students' interests and supporting them so students can turn their interests into "academic endeavors." He said, "If a student comes to me expressing an interest in skateboarding, then it is my job to identify something of value in it."

The data from high-relevance schools also included examples of assignments that provided students with choices about how they would do their work and how they would demonstrate mastery. In these schools students negotiated the means by which they would work, mixing standard research with research in the community, collaboration with experts, work-based projects, and other innovative formats. Students also made decisions about how they would demonstrate mastery. Student products included papers, posters, websites, computer programs, models, video productions, audio-recordings, presentations, performances, and other demonstrations.

Finally, though they were only a subset of the learning opportunities that were described as having real-world applications, student internships provided vivid examples of relevant learning opportunities in high-relevance schools. Teachers and students described students' internship experiences and contributions in elementary and middle school sites, local businesses, community-based organizations, and government. As was already mentioned, many student-initiated projects stressed real-world connections and achieved real-world purposes.

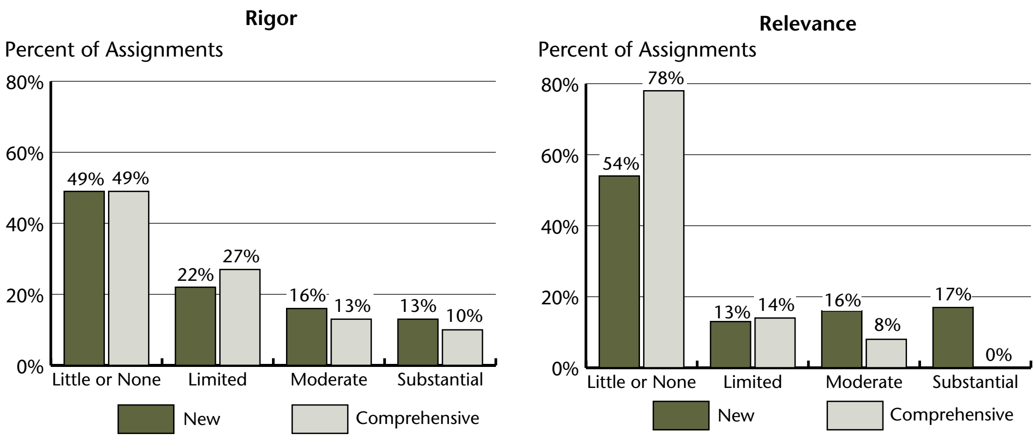
Although the importance of linking English/language arts instruction to students' interests, work styles, and concerns was not ignored by interviewees in low-relevance schools, teachers in these schools described more circumscribed topics and methods and contexts for students' work. The focus of student work and the means by which students worked was more frequently described as teacher specified in low-relevance schools than in high-relevance schools. Faculty in low-relevance schools described internship and other community-based programs as being in their very early stages of development.

Rigor and Relevance in Mathematics

Mathematics Assignments are More Relevant and Slightly More Rigorous in New High Schools than Comprehensive High Schools

Figure 2 provides rigor and relevance data for the mathematics assignments that were submitted. The left-hand graph provides data on the rigor of 10^{th}-grade mathematics assignments in new and comprehensive high schools, showing the frequency with which students received mathematics assignments with little or no rigor, limited rigor, moderate rigor, and substantial rigor. The right-hand graph describes the relevance of tasks in 10^{th}-grade mathematics in the two sets of schools.

Figure 2. Rigor and Relevance of Mathematics Assignments in New High Schools and Comprehensive High Schools



The left-hand graph in Figure 2 shows that almost half of the 10th-grade mathematics assignments in both types of schools had little or no rigor. The mathematics assignments in new high schools show slightly higher levels of moderate or substantial rigor compared with the comprehensive high schools. The right-hand graph in Figure 2 shows that the majority of mathematics assignments collected from 10th-grade teachers in both new and comprehensive schools had little or no relevance. However, more of the mathematics assignments submitted by teachers from new schools were judged by scorers to have substantial relevance; 17% of the assignments collected from teachers in new high schools had substantial relevance, compared with almost none from the comprehensive schools. The average difference between relevance scores for mathematics assignments from the new and comprehensive high schools was statistically significant.[13]

The sidebars in this section help highlight differences between rigorous and relevant instruction in mathematics.

Sidebar: Mathematics Assignment with Substantial Rigor

This assignment, from the Integrated Math Program, was submitted by a teacher.

Count the Pegs

Introduction

Freddie Short has a new shortcut. He has a formula to find the area of any polygon on the geoboard that has no pegs in the interior. His formula is like a rule for an "In-Out" table in which the "In" is the number of pegs on the boundary and the "Out" is the area of the figure.

Sally Shorter says she has a shortcut for any geoboard polygon with exactly four pegs on the boundary. All you have to tell her is how many pegs it has in the interior, and she can use her formula to find the area immediately.

Frashy Shortest says she has the best formula yet. If you make any polygon on the geoboard and tell her both the number of pegs in the interior and the number of pegs on the boundary, her formula will give you the area in a flash!

Problem

1. Your goal is to find Frashy's "super formula," but you might begin with her friends' more specialized formulas. Here are some suggestions about how to proceed:
   Begin by trying to find Freddie's formula and some variations, as described in questions 1A through 1D.
   A. Find a formula for the area of polygons with no pegs in the interior. Your formula should use the number of pegs on the boundary as the "In" and should give you the area as the "Out." Make specific examples on the geoboard to get data for your table.
   B. Find a different formula that works for polygons with exactly one peg in the interior. Again, use the number of pegs on the boundary as the "In" and the area as the "Out."
   C. Pick a number bigger than 1, and find a formula for the area of polygons with that number of pegs in the interior.
   D. Do more cases like Question 1C.

2. Find Sally's formula and others like it, as described in questions 2A through 2C.
   A. Find a formula for the area of polygons with exactly four pegs on the boundary. Your formula should use the number of pegs in the interior as the "In" and should give you the area as the "Out."
   B. Pick a number other than 4, and find a formula for the area of polygons with that number of pegs on the boundary. Again, use the number of pegs in the interior as the "In" and the area as the "Out."
   C. Do more cases like question 2B.
   D. When you have finished work on questions 1 and 2, look for a super formula that works for all figures. Your formula should have two inputs–the number of pegs in the interior and the number of pegs on the boundary–and the output should be the area of the figure.
   E. Try to be as flashy as Frashy!

Write-up

1. Problem statement
2. Process: Explain what methods you used to come up with your formulas.
3. Solution: Give all formulas you found.
4. Evaluation
5. Self-assessment

Explanation of scoring

This assignment scored high on rigor. To complete the assignment, students must demonstrate a conceptual understanding of area and be able to generalize from specific cases, which are two important mathematical ideas. In addition, the assignment required students to engage in fairly substantial problem solving by asking them to generate models, test solutions, and reflect on their problem-solving strategies in writing. Students were asked to show their work and to support their solutions with written explanations.

However, the assignment scored low on relevance because the context for the mathematics was not one typically experienced by adults in the real world, the solution or work product did not satisfy the needs of a real audience, and the student had no involvement in shaping the assignment.

(Used with permission from Key Curriculum Press.)

Sidebar: Mathematics Assignment with Substantial Relevance

This assignment was designed by a student with input from her advisor and internship mentor. The student wrote a project proposal that followed the school's guidelines for developing project proposals and addressed the school's learning goals.

Designing a Website

Introduction

My project is to design a professional website for my internship and also create stationary for them. The website I am going to design is going to have information on what the shop sells and the work they do with pictures of the actual work done. It will also include the business' contact information. The stationary will allow them to have professional handouts, forms and documents that relate to their business.

My finished product will be a professional web page and stationary. At my gateway exhibition, I will show the website I designed and I will show a copy of each stationary I made with all the drafts.

The way my project will benefit my internship site is because it will attract more customers to the show and it will be a better source of information because it's more convenient for a customer to look for an item online than it is to drive to the store.

Focus

- Communication: Interviewing and working regularly with a web designer...Talking to my mentor about what he wants on his webpage and what he thinks of it so far.
- Empirical Reasoning: I will keep an error log where I will write down all of the problems I faced while making the webpage...and testing it to see if I'm correct.

- Social Reasoning: Creating a questionnaire to get feedback from customers about the store, how they learned about the store, and what % of customers has access to the internet.
- Personal Qualities: Meeting my deadlines and being persistent.
- Quantitative Reasoning: I will design a data collection spreadsheet and graph the results from my questionnaire.

Authenticity

The reason I chose this project was because it relates to my interests which are computers and web design. I also plan on starting my own business and I feel that if someone can start a business, then it's helpful to have a website. My resource for this project is my mentor.

Explanation of scoring

This assignment was authored by the student under the guidance of the student's teacher and an adult mentor associated with the student's internship site. The Web site was designed for use in the mentor's place of business. For these reasons, this assignment scored high on both student involvement and real-world connections. The student was exposed to mathematical issues and concepts typically encountered by Web page designers, and the work products were presented to and used by a real audience.

However, the assignment scored low on rigor because the mathematical content and concepts the students was likely to encounter during the completion of the assignment—mean, mode, and standard deviation—were not considered among the important mathematical ideas that 10th graders address. In addition, the level of mathematical problem solving or reasoning needed to complete the tasks was marginal. Finally, outside of a graph and/or table of survey results, the assignment required little or no communication from the student about the mathematics encountered in the assignment or the mathematical knowledge applied.

Challenges for Rigor and Relevance in Mathematics

Analyses of qualitative data on the rigor and relevance of mathematics programs in new high schools were conducted in the same way as analyses for English/language arts. Interview and observation data for the three new high schools with the highest average rigor scores in mathematics were compared with data for the three new high schools with the lowest rigor scores. Analyses were repeated for schools with high and schools with low relevance scores. Analyses were structured to uncover some of the conditions and resources that characterize schools with rigorous and relevant instruction in mathematics.

In contrast to the English/language arts data, our analyses of the qualitative data on mathematics instruction suggested few differences between the practices and conditions of high- and low-rigor schools. Instead, our conversations with school leaders, teachers, and students suggested marked dissatisfaction with mathematics instruction. Teachers in both groups of schools said they were not very satisfied with their instructional programs in mathematics. Faculty reported that mathematics is the most difficult subject to teach well in a way that satisfies their jurisdiction's content standards and the school's instructional vision. Among the barriers to rigorous mathematics instruction that teachers described were the limited availability of useful instructional resources and professional development offerings, the difficulty of integrating mathematics with other content, students' weak numeracy skills, and the limited availability of qualified math teachers.

Teachers in high- and low-rigor schools alike described the limited availability of useful, reform-oriented instructional resources and professional development in mathematics. Teachers explained that good curricular materials in mathematics are hard to find, and multidisciplinary resources are particularly elusive. One teacher explained that she and her colleagues wanted to design an open-ended mathematics curriculum and integrate it with the school's project-based learning components, but were unsuccessful. Another explained:

> It's very hard to find project-based stuff. That's one thing I requested for the upcoming August training. We preach we do it, but I don't think we do it well. So I want to know what a project looks like in geometry and trigonometry. With science it's easier, humanities is okay, but higher-level math is harder unless you start talking about engineering and building bridges, and if that's where we need them to go, then that's what we need to do.

Another teacher explained that he sees a need to give advisors "more of an arsenal of ideas and tools that they can [use to] push the kids and develop these projects more. Many of the student projects that students or teachers described included only basic operations or descriptive statistics as mathematics goals. Teachers in non-project-based schools expressed similar discontent about mathematics curricula.

Faculty in both groups of schools went on to describe the need to balance instruction in procedural mathematical content with their objectives for conceptual understanding of mathematics. Faculty echoed some of the questions in the national debate about mathematics literacy. They posed questions about the proper ordering of instruction for procedural and conceptual understanding and about the optimal mix of direct and constructivist instructional approaches. Faculty worried about what it would take to teach mathematics at the levels of rigor and conceptual understanding they intend. Some new schools have reluctantly adopted software-based mathematics programs, such as *Boxer Math*, *Accelerated Math,* and *Cyber High School* to deliver math content. Teachers stressed that they need professional development in mathematics, but it is not out there.

Teachers in both groups of schools talked about the difficulty of integrating mathematics content with other subjects. They explained that it is common to have a mismatch between grade-appropriate mathematics and grade-appropriate work in other subjects. One teacher explained that the mathematics required by 10^{th}-grade chemistry is not up to 10^{th}-grade mathematics content standards. Another reported that his school tried a joint physics and mathematics course, but it was unsuccessful.

Teachers in high- and low-rigor schools alike said their students came to them unprepared for rigorous instruction in mathematics.[14] Teachers said many of their students lacked basic numeracy skills. Faculty talked about the need for skill building, content review, and structured exercises. One teacher said she's backed away from some of her "big expectations" in mathematics. Teachers viewed their students' weak mathematics backgrounds as barriers to rigorous math instruction.

Another challenge faced by faculty in both groups of schools is the limited availability of qualified mathematics teachers. Interviewees explained that it is difficult to hire and retain qualified math teachers. School leaders described difficulties with teacher turnover, a paucity of teachers with appropriate backgrounds and credentials, and difficulty hiring and retaining mathematics faculty. As a result, like many other schools across the country, some foundation-supported schools had to build programs with teachers who were ill-prepared to teach math. Some of the teachers, students, and parents with whom we spoke described teachers who were uncomfortable with or unenthusiastic about mathematics.

Meeting the Challenges of Rigor and Relevance in Mathematics

Not all of the teachers with whom we spoke gave bleak appraisals of their work in mathematics, however. One teacher explained that he draws mathematics material from a wide range of sources. He said, "I use everything, basically. I'm not locked into a specific curriculum." He said he draws from the Integrated Mathematics Project and Advanced Placement Program materials; he uses resources provided by the intermediary organization through which he receives foundation funding. Other teachers described successful projects with real-life applications of mathematics. One teacher told the story of a student who was interested in crime scene investigation and did an internship at the county coroner's office.

> The student reviewed 600 cases of natural death in the county—the causes of death, gender, age, etc. and put it into a database. The student then did some statistical analyses and made charts that she and the coroner's office will use throughout the county.

> The student created a database and report that is being used by several government offices in the town. She has been a guest speaker at a few service organizations and the local hospital has requested her as a guest speaker for next year.

A third teacher asked students to use trigonometry in their study of sound waves. Another had students use geometry to build models of the Liberty Bell, the Twin Towers, and the Washington Monument.

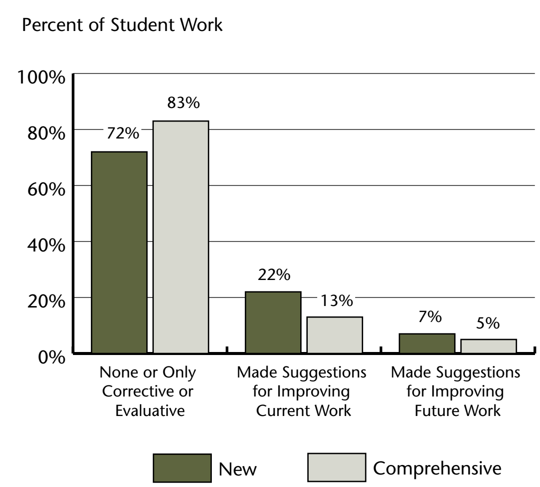
More generally, faculty said they were seeking mathematics expertise in their new hires, supplementing staff capacity with tutors, supplementing their programs with math courses at local colleges and online courses, adding seminars on basic math skills, and continuing to seek or develop problem-based curricula that offer rigorous instruction in mathematics.

Teachers' Feedback to Students

There is Room for Improvement in Teacher Feedback on Students' Work

Scorers also examined the usefulness of teachers' written feedback to students in English/language arts and mathematics. They examined teachers' feedback to see whether it provided information that students could use to refine and improve their current and future work. The scoring criterion distinguished between feedback that merely corrected or evaluated student work and feedback that suggested what the student could do to make content or process improvements in that or future work. Figures 3 and 4 provide feedback scores for new and comprehensive high schools in English/language arts and mathematics, respectively.

Figure 3. Teacher Feedback in English/Language Arts in New High Schools and Comprehensive High Schools

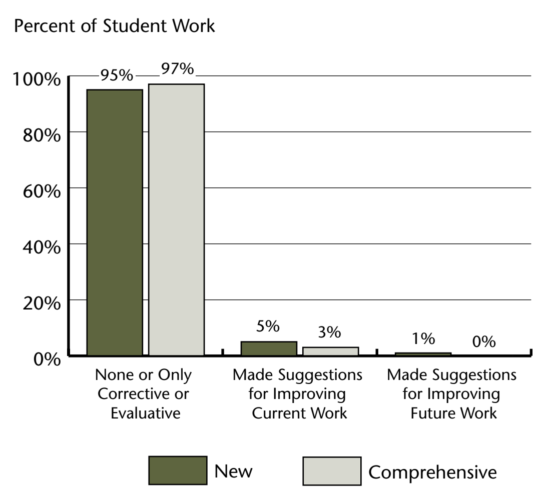


The data indicate that teachers in new high schools provided informative feedback to students more often than teachers in comprehensive high schools. Figure 3 shows that English/language arts teachers in the new high schools in the sample provided slightly more feedback and more informative feedback than teachers in the comprehensive schools. Teachers in the comprehensive high schools provided no feedback or feedback that was only corrective or evaluative on 83% of students' work, compared with feedback on 72% of students' work in new high schools. Teachers in new schools were more likely to provide suggestions for improving students' current work than teachers in comprehensive high schools. Teachers in new schools were also somewhat more likely to take the next step and provide suggestions and guidance for future work. The difference between feedback data in the two types of schools was statistically significant.[15]

It is important to note that these estimates of teachers' feedback to students are conservative. These numbers do not count teachers' verbal feedback to students. Additionally, though teachers were asked to submit all drafts of students' products, it is possible that some drafts were omitted or that teachers provided feedback to students after work was submitted to us.

Figure 4 shows that for both types of schools, mathematics teachers rarely provided written feedback on student work. No feedback or feedback that was merely corrective or evaluative was given to 95% of the work from new high schools and 97% of the work from comprehensive high schools. Although mathematics teachers rarely provided informative feedback, the teachers in new high schools provided corrective or

Figure 4. Teacher Feedback in Mathematics in New High Schools and Comprehensive High Schools



informative feedback slightly more often than teachers in comprehensive high schools. The difference between the feedback offered in new and comprehensive high schools was not statistically significant.[16] As described for the English/language arts data, these are conservative estimates of the frequency with which teachers provide feedback.

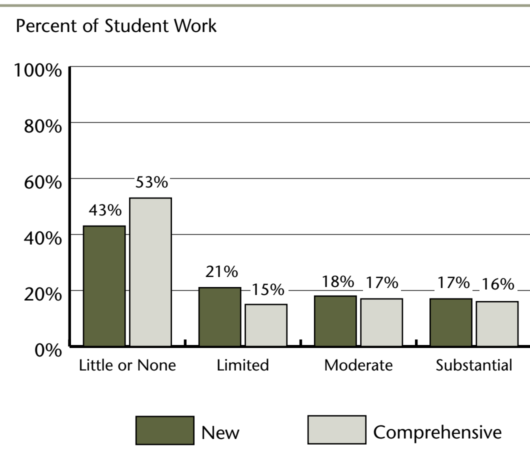
The Quality of Student Work in English/Language Arts

The Quality of Student Work in English/Language Arts is Higher in New High Schools than Comprehensive High Schools

The preceding four figures compared data on the characteristics of assignments and teacher feedback in new and comprehensive high schools. The data showed that students in new high schools had more rigorous and relevant learning opportunities and received more informative feedback than students in comprehensive schools. This section of the report examines the quality of the work that students produced in response. It compares the quality of students' work in new high schools with the quality of work produced in comprehensive high schools. Figure 5 provides data on student work quality in English/language arts.

Figure 5 shows that in both new and comprehensive high schools, a large portion of the 10^{th}-grade English/language arts work was judged by scorers to have little or no quality. In the new high schools, 43% of the work received the lowest scores and in comprehensive high schools over half of the work did. On average, 10^{th}-grade English/language arts students in new high schools produced higher-quality work than students in comprehensive high schools, although the difference was not large. The difference between the two types of schools was statistically significant only after controlling for background characteristics.[17]

Figure 5. English/Language Arts Student Work Quality in New High Schools and Comprehensive High Schools



Though far from positive, these results are best understood in the context of information about the brief tenure of these schools and the entering characteristics of their students. In addition to the fact that many faculty in new high schools are relatively new teachers, as a group the faculty in more than half of the schools were teaching their first 10th-grade class at the time these data were collected. Their students came to them at high educational risk. High percentages of students received free- and reduced-price lunch, had special learning or language needs, came from underserved minority backgrounds, and had lackluster scores on standardized achievement tests. On average, the students in these analyses performed at the 38th percentile in language arts on nationally normed ninth-grade achievement tests.

The sidebars in this section provide additional detail about the characteristics of high- quality student work in English/language arts. One sidebar shows high-quality student work produced in a high school with a student-centered learning model. The work was completed in response to a student-initiated project. The other shows high-quality student work completed in a traditional academic program.

Sidebar: English/Language Arts Student Work with Substantial Quality from a Project-Based Learning School

This work was generated by a student with input from his advisor. The student described the work he intended to do in a project proposal that addressed the school's learning goals and followed the school's guidelines for developing project proposals.

Business Plan for Porkey Entertainment

The information contained herein is believed to be reliable, but the management team makes no representations or warranties with respect to this information. The financial projections that are part of this plan represent estimates based on extensive research and on assumptions considered reasonable, but they are, of course, not guaranteed. The contents of this plan are confidential and are not to be reproduced without express consent.

EXECUTIVE SUMMARY

Introduction

Porkey Entertainment is dedicated solely to the promotion of local music, businesses, and organizations. Within the next two years, Porkey Entertainment plans to serve over 200 bands and sponsor events all over [our local location].

Event Description

In this endeavor, we are trying to establish some credit with the public, make money for future endeavors, and advertise our new sites and other local companies. Over 15 bands will play back to back with interruptions only for raffles, announcements, and free food. Bands from all over [our region] will be playing and bringing fans. We will be advertising with banners, stickers, notebooks, and CD's. Our main goal is for everybody to have fun, creating a positive image for Porkey Entertainment.

Marketing and Sales
For this event we will advertise mainly through bands, with flyers and advertisements on local web sites. Tickets will be $6.00 in advance, and $8.00 at the door and complimentary tickets will be given away at other local events. Our total costs of the event will be $943.00 including the Stage, Venue, PA System, Staff, and food. Local bands will not be paid for playing.

MARKETING AND SALES

Market Analysis
Introduction
The money we make during this event will be mainly from ticket sales. Food will be free, and raffle tickets will be practically free. For our first event, we don't plan on making an abundant amount of profit. This is why the ticket sales and advertising of the event are vital to the success of our event.

The Market
Mainly we are selling to high school students. So to attract high school students, we must use high school themes and make the music selection attractive to high school students. There are hundreds of thousands of high school students in [our state] alone, creating an easy target for sales. The problem is that almost all current advertisements are directed in some way towards teens. But we have an advantage. Because our product is music, and the music is being performed by mainly high school students, word of mouth will be our biggest form of publicity.

The Bling
In the beginning we will rely mainly on small donations and loans. This will take care of the ticket production, deposits, and other cheap beginning expenses. The main cost will be closer to the concert, giving us time to sell tickets and make the money we need. Once we pay for the venue, PA [system], and stage, we will pay off the loans and try to reimburse staff. All profit will be put in the bank until our next event.

MANAGEMENT

Introduction
Porkey Entertainment will be managed by people who have already had a lot to do with local events and have had experience in this type of business. We want people to be involved and care about what happens to the company. The people working for us need to care more about what happens to the company than how much they are getting paid.

Company Organization
At the top of the company there will be a president and two vice presidents. Under them there will be artists, web designers, and promoters. At each event we will have ticket takers, bouncers, tech people, and a set up and take down [crew].

Management Compensation
Everybody employed by Porkey Entertainment will be paid as they finish tasks and for what the Porkey Entertainment decides is reasonable. Because our main goal is to promote local music, paying our workers will be our second priority.

Sidebar: English/Language Arts Student Work with Substantial Quality from a Traditional Academic Program

This paper was written by a student in response to a whole-class assignment on Holden Caufield from *Catcher in the Rye.* The assignment asked students to write the paper as if they were Holden's psychiatrist. They were to choose three things about Holden's personality to discuss, then analyze those three things, and write what they think about them. Students were asked to relate to Holden and identify what is wrong with his thinking.

The Catcher in the Rye

"Good evening Holden! Im very glad that you could make this last meeting of ours, sense you do know Im leaving for retirement. I have been your psychiatrist for sometime now Holden and I think its time to discuss some important matters. I think I should start off by saying Holden that you are a very courageous boy for saying and doing things that you feel are right for you and not for anyone else. While others never say what they feel or never do anything for themselves. That takes guts. But I would also like to discuss some certain issues I've noticed about you. Just a few suggestions I have to help you in your situation and to help you understand where Im coming from. To make this easy for you and me I picked three words that I think best describe you, and with these three words Im going to define each one to help you better understand what exactly they mean. Okay? Lets give it a whirl.

First word…

In isolation: This is the very first word I picked because it's the first vibe that I got from you. Holden I've noticed you always say a word called phony, and I see that you apply this to manly people. You always state how someone is phony, or their doing something phony, and their looking phony, and well basically just phony. But Holden, did it ever occur to you that maybe their trying to be nice? That just maybe their trying to get to know you more, maybe understand you a little better? It seems that your so bitter to allot of people. Maybe if you just tried to understand them you would understand where their coming from, and you wouldn't think their so what you call phony. But don't get me wrong. Im not saying that you are a very secluded person cause also from what I see there are certain people that you do care about. For example your brother Allie. It seems you care a whole lot for him, and when he passed away you kind of died with him. It hurt you. You stated how intelligent he was and how he never got mad at anyone. You also care a whole lot about your sister Phoebe. You say how she brings in straight A's and is just great to talk to. But Holden these people are young. Could it be that maybe you still have a mind of a child? Maybe your not ready to be on your own and make decisions that regard your life. You might think you are ready to face the real world but if you keep isolating people its going to be very hard out there. Its going to seem that the only sane person in the world is you. So maybe after this meeting you could try not being so harsh and maybe not having a judgment right away.

Second word…

Intimidated: Now I don't think your intimidated by allot of people but certain ones. For instance you told me that on your date with Sally you guys went to go see her. You stated that I should have seen the way he hugged her and the way they said Hello to each other. Now Holden, was it really as bad as it seemed or perhaps someone was a little jealous? I think you really liked Sally considering the fact of how good you told me she looked. But this isn't the only time I recall you getting jealous. I happen to remember the time you told me when your roommate at Pencey Prep named Stradlater was gone you kept thinking about what they were doing and what was happening, and when Stradlater came back you were really eager to know what had happen. So eager that you and Stradlater got in a fight. I think this shows that this girl was very important to you. See Holden by being in isolation people will never know who you like. See if Stradlater would have know you liked her maybe the whole date wouldn't have never happened. But also maybe if you would have told Jane how you felt she would have never gone on the date either. My point Holden is that people can't read your mind. By letting people in even if they are phony they'll care. Being intimidated is actually a good thing. Its means you do have feelings and care about more people then just your brother Allie and sister Phoebe.

Finally, my third word...

Confident: Yes Holden hard to believe I picked this word after all the crap I have just given you. But its true and its you. You seem pretty sure of yourself even after getting kicked out of school and experiencing certain things. Like your first encounter with a prostitute. That doesn't happen in everyday life Holden. But that's not why I picked this word. You see when you got kicked out of Pency Prep you decided to go see your teacher Mr. Spencer and to listen to what he had to say. A teacher who even flunked you. You see to me that shows you have confidence in yourself even after everything that's happen to you. Not allot of people have that. Some would say Damn, what do I do now? Some might even cry. But you, you just delt with it, and that Holden is a gift.

So to wrap up our last meeting I would like to tell you I hope you take my words I have given you and use them wisely. My biggest fear in life are people, cause their liable of anything. But I know no matter what, I have to deal with them. Just like you Holden. Im sure you can see how my words I picked for you conjoin altogether. See Holden being in isolation is not helping very much in your life. Your always negative. Its okay to smile once in awhile. But its also okay to have a crappy day once in awhile too. Look Holden your young. Enjoy it while you can, cause other wise you'll grow up to be a sad and grumpy old man! So take care of yourself, Holden. Oh and Holden you can see Mrs. Cadle on your way out if your interested in finding a new psychiatrist. It was a pleasure having you. Hopefully all goes well. Bye Holden."

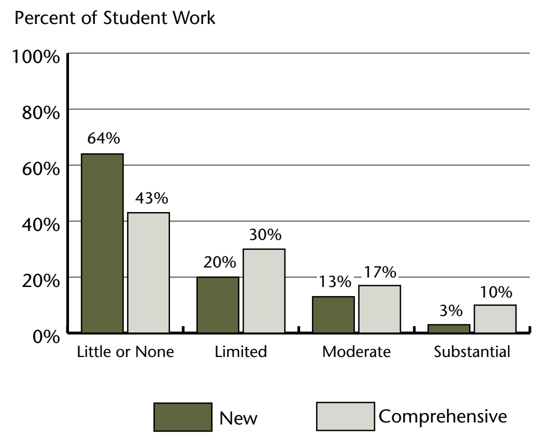
The Quality of Student Work in Mathematics

The Quality of Student Work in Mathematics is Higher in Comprehensive High Schools than New High Schools

This section provides analogous data on the quality of student work in mathematics. Figure 6 compares the quality of student work produced in new high schools with that produced in comprehensive high schools.

Like the English/language arts data in Figure 5, the data in Figure 6 show that scorers judged a large percentage of students' mathematics work as having little or no quality. The pattern of results differs from the English/language arts results in an important way; in mathematics the students in comprehensive high schools outperformed the students from new high schools. In new high schools, 64% of the student work received the lowest scores and in comprehensive high schools 43% of students' work did. The difference between the average quality of student work in mathematics produced at new and comprehensive high schools was statistically significant.[18]

Figure 6. Mathematics Student Work Quality in New High Schools and Comprehensive High Schools



For a host of reasons, these results are not surprising. It is important to remember the early developmental stage of the new high schools and the modest achievement test scores of entering students. On average, students in these new high schools performed at the 42nd percentile in mathematics on nationally normed ninth-grade achievement tests. Also important is the fact that almost half of the assignments given by teachers in these new high schools were judged by scorers as having little or no rigor. These assignments are unlikely to generate intellectually complex work from the students who respond to them. Finally, the results should be examined in the context of teachers' earlier-described dissatisfaction with the mathematics programs in new high schools. With faculty working to balance direct and student-centered instruction in mathematics and looking for satisfying mathematics curricula and models, the very modest quality of students' work in new high schools is not unexpected.

Relationships between the Rigor and Relevance of Assignments and the Quality of Student Work

Rigor and Relevance Matter

Figures 1 through 6 provided data on the characteristics of students' learning opportunities in new and comprehensive high schools and on the quality of their work. Recall that the conceptual model for this research (and common sense) suggests that rigorous and relevant learning opportunities are more likely to prompt high-quality student work than assignments that make fewer demands and offer students fewer real-world connections.

This section of the report not only examines whether more rigorous and relevant assignments lead to higher-quality work, but also looks at which of the two measures matters more in prompting high-quality student work and whether the relationship between assignments and work differ for English/language arts and mathematics.

The data show strong positive correlations between the assignment measures and student work quality in both subjects and mirror the correlations found by Newmann, et al. (1998) in their work with Annenberg Challenge Schools in Chicago. From the research, Newmann and his colleagues concluded that "the quality of the assignments teachers assign to their students is virtually deterministic of the quality of work that students produce on average" (p. 50).[19] The correlations between rigor and relevance and student work quality in the schools in this study range from 0.65 to 0.98 (see the technical appendix for additional detail on these data.)

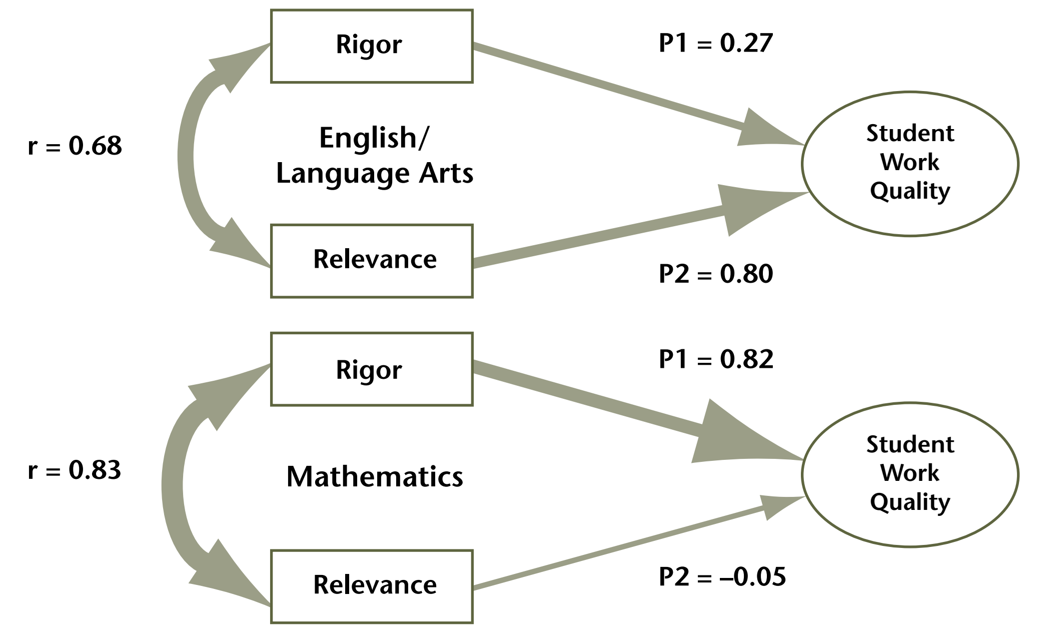
The correlations between rigor and relevance in each of the two subject areas are also strong (0.68 for English/language arts and 0.83 for mathematics).[20] These correlations suggest that rigor and relevance are not incompatible. Many rigorous assignments are also relevant and many relevant assignments are rigorous. Given the high correlations between rigor and relevance, it is important to tease out the individual effects of these characteristics of assignments on the quality of students' work. The path diagram in Figure 7 isolates the effects of rigor and relevance on student work quality and shows which of the two has the most direct relationship with the quality of students' efforts.[21]

In English/language arts, on the left-hand side is the correlation between rigor and relevance mentioned above, 0.68. The path diagram shows that the relationship between rigor and student work quality is not as high as the simple correlation suggests once the influence of assignment relevance is removed. For English/language arts, the relationship between relevance and student work quality (0.80) is actually stronger than the relationship between rigor and student work quality (0.27). So for English/language arts, relevance is more highly related to the quality of students' efforts than is the rigor of assignments.

However, for mathematics, the story is quite different. Here, there is practically no relationship between relevance and student work quality (–0.05). In mathematics, it is rigor that makes the difference in the quality of student work (0.82).

To summarize the findings so far, in English/language arts, assignments in the new high schools in the sample are more rigorous and more relevant than those from the comprehensive schools. Both rigor and relevance

Figure 7. Path Diagram of the Relationships between Rigor, Relevance, and Student Work Quality in English/Language Arts and Mathematics



have positive effects on the quality of the work that students produce, but relevance matters most. Additionally, students at the new schools are producing higher-quality work.

In mathematics, however, the relationships are less clear. Assignments given to students in the new high schools in the sample are more relevant and, on average, slightly more rigorous. In mathematics it is rigor, not relevance, that is more highly related to the quality of students' work. Given the similarity of the mathematics rigor data in new and comprehensive high schools, it is not totally incongruous that the quality of students' math work is somewhat higher in comprehensive schools than new high schools. The qualitative data on mathematics instruction suggest additional explanations for the modest quality of student work in mathematics.

The Relationship between Student Work Quality and Standardized Achievement Test Scores

Students Who Do High-Quality Work in Class Do Better on Standardized Achievement Tests

The final research question focuses on the relationships between the quality of students' work in class and their performance on standardized achievement tests. These relationships are important because, as already discussed, high stakes are associated with standardized test results for both students and schools. In many jurisdictions, students cannot graduate if they fail their high school exit exams. Further, for all but a very few institutions, students are less attractive in the college admissions process if they score poorly on college entrance tests. At the school level, schools can be sanctioned under the *No Child Left Behind Act* if their students make insufficient progress on jurisdiction-sponsored standardized tests, and with lackluster test scores, new schools may be unable to recruit good students and faculty. Understanding the links between high-quality student work and capable performance on standardized achievement tests, therefore, is of more than casual importance. The expectation is that good work in class will translate into solid performance on standardized achievement tests.

Data for students in six of the schools in this study (five new schools and one comprehensive school) were used to examine the association between students' performance in class and their scores on 10^{th}-grade achievement tests. These schools were selected for analysis because students in these sites took nationally normed standardized tests in English/language arts and mathematics (i.e., the CAT-6 or SAT-9). Since both of the tests are nationally normed, they can be put on the same metric. Scores can be converted to a normal curve equivalent or NCE, allowing for estimation of the relationships using data from schools that take different tests.[22,23] As with other analyses described in this report, estimates were adjusted for student demographics and prior achievement scores, class composition, teacher characteristics, and type of assignment (typical vs. challenging).

Even after controlling for student, class, and school characteristics, the data showed that an increase of one standard deviation in students' English/language arts or mathematics work quality scores was associated with an increase of approximately a third of a standard deviation in students' achievement test. These relationships between student work and standardized test data were statistically significant. The data suggested that the quality of students' work in class is related to students' performance on standardized tests.[24,25]

Caution should be exercised in interpreting these results, however. Similar analyses were completed on student work and test score information for a sample of students attending comprehensive high schools in Washington state; these analyses show only a weak relationship between student work and standardized test data. Because national norming data are not available for the test taken by Washington high school students, the Washington scores could not be combined with the national dataset. The Washington relationship was smaller and not significant. See the technical appendix for more details on these analyses.

Implications for the Initiative

Our previous evaluation report, *Creating Cultures for Learning: Supportive Relationships in New and Redesigned High Schools* (AIR/SRI, 2005), examined the progress of reform in new and redesigned high schools, giving special attention to the development of relationships between and among students and teachers. The current report investigates whether the school-level changes described in that report have corollaries in the classroom. This report looks inside the classroom, describing students' learning opportunities and their results. It examines teaching and learning to determine whether schoolwide change lays an adequate foundation for classroom innovation and high-quality student work.

Summary of Findings

The data examined by this report suggest that schoolwide changes in foundation-affiliated sites have, indeed, reached into the classroom. These data support several important inferences about the nature of learning opportunities and student work in foundation-affiliated schools. We have found the following:

- ***Assignments given in the new high schools are more rigorous than the assignments given in the comprehensive high schools.*** English/language arts assignments in new schools are more likely to entail the construction of knowledge and elaborated communication. Mathematics assignments in new high schools also tend to be more rigorous than those in the comprehensive high schools, but the difference is very small.

- ***Assignments in the new high schools place a strong emphasis on embedding learning opportunities in real-world settings and giving students a voice in shaping these opportunities.*** English/language arts and mathematics assignments in these schools are more likely to have real-world connections and to incorporate elements of student choice, compared with assignments given by teachers in the comprehensive high schools.

- ***Rigor and relevance are not incompatible.*** Most rigorous assignments are also relevant.

- ***Teacher feedback on students' work in English/language arts is more informative in new high schools than in comprehensive high schools though there is room for improvement in both.*** There is much room for improvement in the quantity and quality of mathematics teachers' feedback in both types of schools.

- ***Students in the new high schools do higher-quality student work in English/ language arts, compared with students in the comprehensive high schools.*** In the new high schools, student work in English/language arts is more likely to demonstrate a deep conceptual understanding of content, clear communication, facility with language, and the construction of new knowledge.

- ***Students in the comprehensive schools do higher-quality work in mathematics, compared with students in the new high schools.*** In their class work, students in the comprehensive schools are more likely to show that they know and understand important math content, are facile with mathematics conventions, and possess skill and understanding in problem solving and reasoning.

- ***In English/language arts, both rigor and relevance in assignments lead to higher-quality student work,*** and relevance is the more important of the two. Assignments with a strong emphasis on embedding learning opportunities in real-world settings and giving students a voice in shaping their assignments lead to better-quality student work.

- ***In mathematics, rigor matters for student work quality, but relevance does not.*** Relevance in mathematics assignments is not correlated with the quality of student work.

- ***Students who do higher-quality work in school do better on standardized achievement tests,*** although the results are not strong. We see some support for this relationship, but the relationship hinges on what the tests measure and the content of the courses.

We do not yet have samples of assignments from the same teachers over multiple years. Hence, we cannot draw conclusions about trends in the rigor and relevance of teachers' assignments over time as faculty spend more time in innovative new schools. We are currently collecting additional assignments and student work to address this question.

Implications

From these findings, we derive several implications.

- ***Professional development around teaching practices that incorporate both rigor and relevance is critically important for faculty in reforming schools.*** In almost every school, teachers asked for help in developing and honing their practice. They lamented the limited availability of useful professional development materials, offerings, and coaching. Especially important is professional development on the implementation of innovative practices within the context of current federal and state accountability requirements.

- ***Not every intermediary organization and district has a clear vision of effective instruction and appropriate curricular materials for high-need student populations.*** The organizations receiving foundation funds vary markedly in their histories and the nature of their expertise (AIR/SRI, 2003, 2004). Some of the organizations specialize in supporting a grassroots planning process rather than in providing instruction. Some have a commitment to a given instructional approach (typically project-based learning) but believe that teachers should develop the curriculum materials to use with this approach based on their particular students' interests and needs. Schools that do not receive curriculum resources from their grantee organization have the added burden of developing curriculum while trying to start or redesign a school; some resort to packaged software or traditional district textbooks.

- ***Teachers need compelling illustrations of the kinds of rigorous, relevant assignments that students with backgrounds similar to those of their own students can do.*** Although teachers have concerns about their students' academic preparation, there is proof from some of the classrooms in this study that low-income, historically underserved students can rise to the challenge of highly relevant, rigorous assignments. Good models of high-quality, relevant assignments and student work are needed to support teachers' work and stimulate students' efforts. The presence of tools per se, however, is rarely sufficient to bring about changes in instructional practice. Teachers need supported time for interacting with the resources and, ideally, with other teachers who have used and are currently using them.

- ***Innovative schools could learn from examples of mathematics assignments that are both rigorous and relevant.*** Given many teachers' concerns about the quality of their efforts in mathematics, the foundation might want to consider funding an online library of mathematics assignments that have real-world connections and are linked to high standards, supplemented with professional development. There are lessons to be learned from the new schools where mathematics assignments are both rigorous and relevant. One possibility would be to train master teachers at each school using the rubrics developed for the teacher assignment and student work scoring as a starting point for discussing real assignments given by teachers. Many of the schools started under this initiative stress a project- or problem-based approach that is theoretically compatible with teaching mathematics concepts and skills but difficult to reconcile with the specifics of district and state standards for algebra and geometry courses. Rather than expecting individual grantees or schools to solve this problem for themselves, the foundation could fund an organization with instructional development expertise in secondary mathematics to develop materials and facilitate the sharing of information among teachers.

- ***Foundation-affiliated schools should offer structured mathematics classes, rather than relying primarily on mathematics learning through internships or unstructured projects.*** Incorporating mathematics into internships and student-designed projects is difficult because of the specific content that students are expected to learn in high school algebra and geometry. Students' college prospects will be better served by mastering the required high-level content than by relying entirely on internships or unstructured projects that incorporate mathematics at a more basic level (such as basic operations, percentages, and averages, as opposed to linear equations and rates of change). Some schools are successfully integrating high-level mathematics content with science in their project-based curriculum. These projects are structured, have been refined over several years in the classroom, and are supported with instructional materials and guides for implementation.

- ***Schools serving high-need students should provide supplementary academic supports,*** including academic coaching during the day and after-hours homework support. School staff cited lack of tutoring services as a barrier for many of the students in foundation-affiliated schools. In some sites, business and community partners provide mentors or tutors. Federal and state programs supporting after-school academic activities are a potential source of funding. Private nonprofit and community-based organizations are potential partners in providing such supports.

References

AIR/SRI. (2003). *High time for high school reform: Early findings from the evaluation of the National School District and Network Grants Program.* Menlo Park, CA: SRI International.

AIR/SRI. (2004a). *The National School District and Network Grants Program: Year 2 evaluation report.* Menlo Park, CA: SRI International.

AIR/SRI. (2004b). *Exploring assignments, student work, and teacher feedback in reforming high schools: 2002–03 data from Washington state.* Menlo Park, CA: SRI International.

AIR/SRI. (2005). *Creating cultures for learning: Supportive relationships in new and redesigned high schools.* Menlo Park, CA: SRI International.

Berends, M., Bodilly, S., & Kirby, S. (2002). *Facing the challenges of whole-school reform* (MR-1498-EDU). Santa Monica, CA: RAND.

Bill & Melinda Gates Foundation. (2004). *Creating a portfolio of great high schools*. Retrieved July 20, 2005, from http://www.gatesfoundation.org/nr/downloads/ed/GreatHighSchools.pdf.

Bill & Melinda Gates Foundation. (2005). *High schools for the new millennium: Imagine the possibilities*. Retrieved July 20, 2005, from http://www.gatesfoundation.org/nr/downloads/ed/EdWhitePaper.pdf.

Bodilly, S. J. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools* (MR-942-NAS). Santa Monica, CA: RAND.

Bottoms, G., Presson, A., & Han, Lingling. (2004). *High School Reform Works—When Implemented: A Comparative Study of High- and Low-implementation Schools.* Atlanta, GA: Southern Regional Education Board.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Bryk, A. S., Nagaoka, J. K., & Newmann, F. M. (2000). *Chicago classroom demands for authentic intellectual work.* Chicago, IL: Consortium on Chicago School Research.

Buck Institute for Education. (n.d.). *PBL handbook.* Retrieved April 3, 2004, from http://www.bie.org/pbl/pblhandbook/contents.php

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Doran, H. & Drury D. (2002). *Evaluating success: KIPP educational program evaluation.* Retrieved July 20, 2005, from http://www.naschools.org/uploadedfiles/Microsoft%20Word%20-%20KIPP%20Final%20Technical%20Report%2010.21.02%20_no%20embargo_.pdf

Green, J., & Winters, M. (2005). *Public high school graduation and college readiness rates: 1991–2002.* (Education Working Paper No. 8). New York: Center for Civic Innovation, Manhattan Institute.

Illinois Mathematics and Science Academy. (n.d.). *Center for Problem-Based Learning.* Retrieved August 8, 2005, from http://www2.imsa.edu/programs/pbln/overview/history/php

Kemple, J. J. (2001). *Career academies: Impacts on students' initial transitions to post-secondary education and employment.* New York: Manpower Demonstration Research Corporation. Retrieved July 20, 2005, from http://www.mdrc.org/Reports2002/CA_StudentsImpacts/CA_StudentImpactwTech.pdf

Newmann, F., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report.* Chicago, IL: Consortium on Chicago Schools Research.

Newman, F. M. & Wehlage, G. G. (1995). Successful School Restructuring: A Report to the Public and Educators by the Center on Organization and Restructuring of Schools. Madison, WI: University of Wisconsin.

Smith, L., Ross, S., McNelis, M., Squires, M., Wasson, R., Maxwell, S., Weddle, K., Nath, L., Grehan, A., & Buggey, T. (1998). The Memphis Restructuring Initiative: Analyses of activities and outcomes that impact implementation success. *Education and Urban Society,* 30(3), 296-325.

Stringfield, S., Millsap, M., & Hermann, R. (1997). *Special strategies for educating disadvantaged children: Findings and policy implications of a longitudinal study.* Washington, DC: U.S. Department of Education.

Endnotes

[1] The views and recommendations expressed in this report are those of the authors and do not necessarily reflect the opinions of the Bill & Melinda Gates Foundation.

[2] The attributes are common focus, high expectations, personalization, a climate of respect and responsibility, time for staff to collaborate, performance-based decisions, and the use of technology as a tool. Powerful teaching and learning are characterized by active inquiry, in-depth learning, and performance assessment.

[3] When the foundation revised its national education strategy in the summer of 2004, it articulated its goals in terms of high performance for all students and improvements in the proportion of students graduating from high school ready for college, without specifying the means by which schools and systems of schools would achieve these ends. Thus, both school structures and instructional approaches were deemphasized in order to stress the importance of a performance orientation.

[4] Seven comparison sites for the new schools will be included in future analyses and reports.

[5] Instructional characteristics are fully or partially in place in the school.

[6] Under the assumption that readers are familiar with these, Table 3 does not provide data on the characteristics of comprehensive high schools.

[7] Many of the assignments and work were scored by two scorers, allowing for an analysis of interrater reliability.

[8] Analysts typically use MFRM to score student essays on standardized tests and in other situations where many raters are involved. MFRM is used in these cases and in this analysis because raters may vary in how they apply the scoring standards, and MFRM takes into account differences in rater severity.

[9] The analyses in this study use the more specialized type of regression analysis called hierarchical linear modeling (HLM) because of the nature of the data. The assignments are not randomly selected from the set of all English/language arts and mathematics assignments but are selected from within the 24 sampled schools, and inside those schools they are selected from the sampled teachers. In other words, the data are hierarchically structured: assignments are nested within teachers, and teachers are nested within schools. It would not be practical to randomly select assignments without first picking schools and teachers, nor would it be as useful. The HLM technique adjusts for the fact that the eight assignments given by a teacher are more similar to each other than eight randomly selected assignments from different teachers would be.

[10] The graphed results do not control for background factors. Using a regression analysis (HLM) to hold constant other variables that might explain the differences between new and comprehensive schools, the assignments in the new schools were still more rigorous and more relevant than in the comprehensive schools. These regressions controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging). The differences for rigor and for relevance were statistically significant ($p < 0.05$).

[11] Rigor scores were estimated with and without controls for student (including 9th-grade achievement test results), teacher, and school (including free- and reduced-price lunch data) characteristics, and the membership of the high-rigor and low-rigor groups was unchanged.

[12] The three schools with the highest average relevance scores were the same as those with the highest average rigor scores in English/language arts. The same is true for the three low-relevance schools.

[13] Using HLM, whether or not the data were adjusted to control for background factors such as student demographics, average classroom student achievement, class composition, teacher characteristics, and assignment type (typical vs. challenging), the results were the same: new schools scored slightly higher on rigor, and they scored higher on relevance. The difference between school types in the rigor of mathematics assignments was not statistically significant ($p > 0.10$); the difference between the relevance of mathematics assignments at new schools and comprehensive high schools was statistically significant ($p < 0.01$).

[14] Teachers also voiced concern about students' low literacy levels. Both this year and last, teachers reported that students' low literacy levels prompted them to focus on skill building in reading (AIR/SRI, 2004a).

[15] Whether or not we controlled for assignment type, teaching experience, class prior achievement, classroom composition, and the school risk index, the difference was statistically significant ($p < 0.05$).

[16] Whether or not we controlled for the same background variables as in English/language arts, the difference between school types was not statistically significant ($p > 0.10$).

[17] The difference between the average scores in the two types of schools was marginally statistically significant ($p < 0.10$). When student work scores were adjusted for differences between schools in student demographics and prior achievement, class composition, teacher characteristics, and assignment type (typical vs. challenging), on average students at new high schools outperformed their peers at comprehensive schools by a notable margin; the difference between school types was highly statistically significant ($p < 0.01$).

[18] Whether or not we controlled for background factors mentioned above, students at comprehensive schools produced higher-quality work in mathematics than did students in new schools; in both cases, the difference between school types is statistically significant ($p < 0.01$).

[19] It should be emphasized that the correlations presented above are correlations at the assignment level; therefore, the strong correlations we found do not mean that different students respond in the same way to the same assignment, but that assignments with higher levels of rigor and relevance tend to result in student work of better quality.

[20] In English/language arts, the correlations between rigor and student work are quite high (0.81). The correlations between relevance and student work are even higher (0.98). We also see that the relationship between rigor and relevance in assignments is very high, (0.68), indicating that, in general, English/language arts assignments that are rigorous also tend to be relevant. Similarly, we see very high correlations between assignment rigor and student work in mathematics (0.83) and between assignment relevance and student work (0.65). The correlation between rigor and relevance in mathematics is also very strong (0.83). These findings are consistent with those obtained by Newmann, et al. (1998), who found that the correlations between the quality of teacher assignments and the quality of student work in English/language arts and mathematics in grades 3, 6, and 8 ranged from 0.71 to 1.00, with most of the correlations in the high 0.80s.

[21] As with the correlations mentioned above, these numbers show assignment-level relationships.

[22] It is important to note that differences between the measured constructs and measurement models used by different test batteries make the analysis of score information from different tests a controversial issue.

[23] Students at 10 schools in the sample either took achievement tests that did not have national norms or did not provide 10th-grade test score data.

[24] The relationships between rigor and relevance of assignments and student achievement scores were also examined, but the effects seem to flow through the quality of student work.

[25] This relationship can be considered a conservative estimate of the correlation between student work quality and test scores. The HLM accounts for measurement error in the measure of student work quality, but standard errors for test scores were not available.

Technical Appendix

I. Sample and Data Collection

Teacher Assignments and Student Work

The data on rigor, relevance, and student work quality in this report come from teacher assignments and student work collected in 12 new high schools and 12 large comprehensive high schools. Tenth-grade English/language arts and mathematics teachers were asked to provide eight assignments over the course of the school year, along with instructions given to the students and relevant pages of textbooks, answer keys, and grading rubrics. We asked teachers to provide four assignments that were typical of the assignments they give to their students on a day-to-day basis and four challenging assignments that show what students know and can do at high levels. For three of the eight assignments (two typical and one challenging assignment), we asked for the student work from 10 randomly selected students. In some schools where a project-based learning curriculum was implemented, we revised the data collection to three collections per year, where we collected project proposals (assignments) and completed projects (student work) from 10 randomly selected students for all three data collections.

Assignments and work were collected in 2002–03 and 2003–04. In the first year, we collected assignments from eight large comprehensive high schools in Washington state in the year before these schools converted into several smaller schools or learning communities. Assignments were collected from 47 teachers, 24 in English/language arts and 23 in math (one math teacher initially agreed and then later declined to participate). The second year, we collected assignments and work from 12 new schools and 4 additional comprehensive high schools from across the nation. The assignments and student work came from 58 English/language arts classes and 53 math classes. At the comprehensive schools, we aimed for three English/language arts and three math teachers from each school. At the new schools, there was often only one or two of each type of teacher.

Teachers were eligible for inclusion in the study if they taught English/language arts or mathematics to sophomore students, had a class that consisted of mostly sophomore students, and if at least 25% of all sophomores took that level of coursework.

Participating teachers sent a letter home with their sophomore students describing the study, providing students and their parents the opportunity not to participate in the study (i.e., passive consent). Teachers submitted a class list of sophomores to the AIR/SRI data collector, after removing the names of students who opted out. In the first year of data collection, up to 12 students were selected per class; in the second year, 10 students were selected per class (fewer if there were not enough students in the class).

Site Visits

In spring of 2004, site visits were conducted in 13 of the schools in which teacher assignments and student work were collected. Nine new high schools were selected to represent the 12 new schools in the study sample. All four comprehensive high schools in the national sample were visited. The eight schools in the Washington state sample were not visited in 2004.

Two-person teams visited each school over a period of 2 to 4 days, as needed. School site visits included interviews with school principals and other leaders considered key to the success of reform activities, interviews with five teachers, observations of five classrooms, and focus groups with two groups of students with approximately six students in each group. Interviews and focus groups were audio taped to support the completeness and accuracy of the data records.

Principals and Lead Staff

Site visit teams began and ended school visits with principal interviews (where possible). Site visitors also interviewed reform facilitators, coaches, design team leaders, and curriculum leaders. These interviews covered topics such as the conception of the school's mission, academic organization, curriculum and instruction, and supports offered by the grantee organization.

Teachers

Site visit teams interviewed five teachers at each school. Teachers to be interviewed were selected to meet the following criteria at schools where the criteria were consonant with the structure of the school:

- A 10th-grade English/language arts teacher (if the school did not have a 10th-grade English/language arts teacher, we interviewed a 9th-, 11th-, or 12th-grade English/language arts teacher).
- A 10th-grade mathematics teacher (if the school did not have a 10th-grade mathematics teacher, we interviewed a 9th-, 11th-, or 12th-grade mathematics teacher).

- A teacher of any subject at the 9th-grade level (if the school did not have a 9th grade, we selected a teacher in the lowest grade above 8th grade) and at the 11th- or 12th-grade level (if the school did not have an 11th or 12th grade, we selected a teacher at the school's highest grade).
- Someone who taught an innovative class, preferably at a higher grade level in the school.

These categories were incongruent with the school structures of some of the schools. For example, some schools do not have discrete English/language arts or mathematics classes; in these cases, we asked leaders to identify teachers of classes where mathematics and English/language arts were substantial parts of instruction. Some of these schools do not group students by grade level in mathematics and language arts. In these cases, we selected teachers so that their five classes represented a range of student levels. Site visit teams tried to schedule teacher interviews so that the same teachers could also be observed as part of our classroom observations (see below). Among the topics discussed with teachers were teaching and learning, relationships between teachers and students, and the school's ability to serve all students well.

Students

Site visit teams completed two student focus groups per school. Students were members of the classrooms of the teacher interviewed, when possible, with one six-member group coming from one of the lowest-grade classes in the school and the other from one of the highest-grade classes. Schools were asked to select from among the more heterogeneous of these classes. Students from selected classes were asked to take parent consent forms home for parent signature, and focus group students were selected from among those who returned signed forms. School coordinators were asked to select a mix of students by gender, racial/ethnic group, and native language status for each group. In focus groups, students were asked to describe the nature of their schoolwork, the nature of relationships between students and teachers at the school, and their assessment of how well the school was preparing them for life after graduation.

Classrooms

The site visit teams conducted 50-minute observations in the classrooms of interviewed teachers. Observers provided a narrative description of the activities they had observed; they described the structure of the instructional activity and the roles played by teachers and students. Observers also noted the instructional resources used and aspects of classroom management. After the observations, visitors met with teachers to discuss what they had seen. Observers asked teachers whether the work they had observed was part of a long-term product and, if so, whether students were using rubrics to examine their work, whether students would have opportunities to revise their work, and whether students would have opportunities to apply what they had learned to real-world contexts.

Buildings and Structures

At the conclusion of the school visit, site visitors completed an Implementation and School Environment Inventory. The inventory described the physical environment of the school, catalogued the school design components that were in planning or in place, noted the correspondence between the school model and school environment, and described the school location and neighborhood.

II. Coding and Analyses of Site Visit Data

Data Coding

After returning from visits to schools, site visitors organized the data they had collected into a standard set of sections within data-capture forms. For each type of interview, there was a form with a set of headings, organizing the data in a structure parallel to the flow of the interview protocol. In addition, a school summary form was used to capture more general or synthetic impressions. Site visitors completed the data-capture forms on the basis of their notes, checking interview tapes when appropriate for clarification or to obtain exact wording for quotations. Conventions were used to indicate the source for each piece of information, to designate the speaker's exact words as opposed to paraphrases, and to separate data that came directly from the interview from inferences or clarifications provided by the site visitor. Senior analysts reviewed the data-capture forms and requested clarifications and additions as needed.

In preparation for data coding, we developed a manual of codes, definitions, and procedures. Codes were developed for the constructs in the foundation's theory of change and for additional constructs in the conceptual framework. Codes described capacity issues, key school attributes, characteristics of curriculum and instruction, learning outcomes, other student and school outcomes, and many other topic areas. Each of these broad coding categories included codes for subtopics. Codes were designed to allow parsing of data-capture forms by topic. There were more than 130 codes in all.

Data coding began with test coding, moved on to reliability and validity coding, and concluded with operational coding. After the coding structure used with 2003 data was refined, coders were trained to use the new draft coding manual and worked in pairs on a sample set of data-capture forms to test the codes. Throughout the test coding process, weekly meetings among the coders and several analysts offered an opportunity for joint review of coding results and discussion of potentially ambiguous codes or other needed revisions to the coding manual.

Once the coding structure was tested and refined, subsets of five data-capture forms were selected to cover a wide variety of form types and content areas. These data forms were used to conduct reliability and validity trials. The trials were designed to promote common uses of codes across coders and to ensure that segments of text were coded as analysts would expect. Coder pairs coded the text segments individually and then negotiated an agreed-upon set of codes. These codes were then reviewed and refined by senior analysts. The resulting set of codes, agreed upon by coders and analysts, was taken as the standard against which coders' original individual responses were compared in order to examine the reliability of coding decisions. Agreements and disagreements with the standard codes for each paragraph were tallied by code, and agreement scores were calculated as agreements/(agreements + disagreements). The reasons for any low agreement scores were explored and other outstanding issues were resolved. The coding definitions were then updated to improve clarity where necessary, and the process was repeated with the new set of definitions.

We then conducted a second reliability round on school data, in which 74% of codes that were used more than five times in the coding sample had estimated reliabilities ranging from 70% to 100%. Codes below that threshold corresponded to concepts that generally were difficult to separate from related topics in the narratives. For example, school personnel often talked about schoolwide policies that promote close teacher-student relationships in the same breath as practices that promote personalization of learning within the classroom. Such interrelated constructs made coding distinctions challenging. In cases like these, we computed reliability estimates for two interrelated codes together and encouraged the use of narrative data by paired codes.

Once we moved from reliability to operational coding, weekly meetings continued for the resolution of any new issues that arose. Pair coding on selected data-capture forms was used on an ongoing basis for calibration.

Qualitative Data Analysis

Many of the analysts of teaching and learning data began their work by reviewing samples of data-capture forms for schools in their analysis group. These reviews helped analysts get a more comprehensive view of the school contexts and schoolwide issues.

Analysts then queried the ATLAS.ti database to review coded data. They consulted the coded data on teaching and learning in English/language arts and mathematics, generated an initial set of themes to pursue, and developed matrices and other supporting documents to track whether or not, and in what way, a particular issue was in evidence at each school. To vet and refine the emerging themes, analysts worked in small teams by topic area and iteratively reviewed and discussed data until they reached consensus on the supported themes. A larger team of qualitative and quantitative analysts met weekly to evaluate the qualitative themes and examine the consistency of findings across the qualitative and survey data and to decide on areas that warranted further analysis.

For the qualitative analysis, we selected schools with the highest scores on each of the assignment and student work measures and compared them with schools with the lowest scores on each metric. The research team looked at both the adjusted and unadjusted scores, that is, with and without controlling for background factors. The adjustments did change the rankings somewhat, but fortunately we found that the same schools would be selected either way.

Once the research team had selected sets of schools, the entire collection of qualitative data was searched, including data from interviews with teachers and principals and from focus groups with students, for all mentions of rigor and relevance and student outcomes, positive or negative.[1] Analysts paid special attention to response information about English/language arts teaching and learning from English/language arts faculty in analyses on the rigor and relevance of assignments and the quality of student work in English/language arts. They followed the same logic for mathematics.

III. Measuring the Rigor and Relevance of Teacher Assignments and the Quality of Student Work

Many-Facet Rasch Measurement

This project builds on the work of Newmann and others and their study of assignments and student work in the Chicago Public Schools (Newmann, Lopez, & Bryk, 1998; Bryk, Nagaoka, & Newmann, 2000; Newmann, Bryk, & Nagoaka, 2001). The estimation procedures described here largely parallel the procedures used by the Chicago researchers. There are two parts to these analyses. First, a Many-Facet Rasch Measurement (MFRM) analysis is used to combine the scores for the individual rubrics for each assignment into a rigor score and a relevance score for that assignment.[2] The second part of the analyses uses hierarchical linear modeling (HLM) conducted at the classroom level to examine the rigor and relevance of assignments across school types.

The measurement of the rigor and relevance of teacher assignments poses two challenges. The first challenge relates to differences in the severity of the scorers: although they were trained together using the same materials, some scorers may have higher standards than others. If *all* the assignments were rated by *all* the raters on *all the rubrics*, then the simple average of the ratings would balance out differences in rater severity. However, such a massive scoring activity was not feasible. Thus, we need to adjust statistically for the differences in the severity of the scorers.

The second challenge is associated with differences in the stringency of the rubrics for the criteria. For example, it might be harder for English/language arts assignments to achieve top scores for Construction of Knowledge than Elaborated Communication.[3] We used the Many-Facet Rasch Measurement (Linacre, 1989a) to combine the individual raw scores for each criterion to develop measures of the rigor and relevance of assignments. Scales were developed separately for English/language arts (ELA) and mathematics.

The Many-Facet Rasch Measurement model that was used with the assignment data is:

$$\log\left(\frac{P_{n\,i\,j\,k}}{P_{n\,i\,j\,(k-1)}}\right) = B_n - C_i - D_j - F_{ik}$$

where

$P_{n\,i\,j\,k}$ is the probability of assignment *n* getting a rating of *k* on rubric *i* by scorer *j*

$P_{n\,i\,j\,(k-1)}$ is the probability of assignment *n* getting a score of *k-1* on rubric *i* by scorer *j*

B_n is the parameter for assignment *n* (quality of the assignment: rigor or relevance)

C_i is the parameter for rubric *i* (stringency of the rubric)

D_j is the parameter for scorer *j* (severity of the scorer)

F_{ik} is the parameter for receiving a rating of *k* relative to *k-1* on rubric *i* (step difficulty).

Using this expression of the relationships among the parameters and evaluating it against our dataset using maximum likelihood, we derived estimates for the parameters. The product of the analysis was the measure of each element of three facets: the assignment rigor (or relevance), B_n; the rubric stringency, C_i; and the scorer severity, D_j. The model also outputs a measure of step difficulty, F_{ik}. Of these, the estimate of interest is B_n, the quality of teacher assignments, yet other estimates are also important, since they serve as statistical controls. In other words, the Many-Facet Rasch Measurement model corrects the estimates of assignment rigor and relevance for scorer severity and rubric difficulty.

Rescaling

The logit measure produced by the Rasch model can theoretically range from negative infinity to positive infinity. For reporting purposes, we rescaled the logit measure to a 0 to 10 scale. The transformation formula is:

Assignment measure = 10 x (logit measure – min) / (max – min)

Where the logit measure is the original measure for the assignment, min is the minimum value for any assignment, and max is the maximum value for any assignment. By the same operation, the estimate for the standard error is also transformed to the same scale by:

Standard error = 10 x original standard error / (max – min)

This rescaling was done separately for English/language arts and mathematics.

Diagnostics for MFRM Scoring Results

We examined a series of the MFRM diagnostics to assess the psychometric properties of each of the facets. Diagnostics were examined with a goal of improving the overall reliabilities of the teacher assignment and student work scores in the future.

In this section, we report on the statistical properties of the MFRM scores by examining each of the facets: quality of the assignment or work, stringency of the rubric, and severity of the scorers. For the first facet, we report the reliability of assignment and student work scores. For the second facet, we report the ranking of stringency (or item difficulty) derived for each of the criteria. For the third facet, we report interrater reliability. Finally, because the training of raters and the scoring occurred over two summers and half the raters in the second year were new, the standards that the raters applied could drift. That is, raters from 2004 might have been more severe or more lenient than the 2003 raters.

Facet 1: Reliability of Scores

Using MFRM, we derived two types of scores for each teacher assignment, a rigor score and a relevance score, and one score for each piece of student work. We obtained these scores for English/ language arts and mathematics, creating six measures in total. Table A-1 provides reliability statistics for the measures. Here reliability is analogous to Cronbach's alpha. It refers to the estimate of the replicability of the measures that can be expected if the quality scores for the same assignments or student works were to be measured again. The low reliability of our ELA relevance score was a result of low variance among the scores. The majority of assignments received the lowest scores on the rubrics for the two relevance criteria. Researchers generally set a threshold for acceptable reliability at 0.65 to 0.70. Since the reliability of our measures approached this threshold, a statistical technique called the latent variable HLM analysis was required. This technique decreases the influence of unreliable scores by giving them lower weights in the analysis.

Table A-1. Reliability of Measures

Criterion	*Reliability*
English/Language Arts	
Assignment Rigor	0.74
Assignment Relevance	0.41
Student Work Quality	0.68
Mathematics	
Assignment Rigor	0.66
Assignment Relevance	0.69
Student Work Quality	0.62

Facet 2: Stringency of Criteria

Table A-2 examines the relative difficulty of the different scoring criteria for teacher assignments and student work. For this analysis, we included the items for rigor and relevance together, instead of treating the categories separately. This method keeps all the items on the same metric, allowing us to compare across criteria. A high measure of criterion difficulty indicates a more stringent (difficult) criterion.

For English/language arts assignments, elaborated communication was the criterion most likely to produce high scores (i.e., the "easiest" criterion, with the lowest difficulty score of –1.86), whereas real world connections was the most difficult criterion on which to achieve a high score. Interestingly, all rigor items scored lower in item difficulty than the relevance items. This pattern indicates that qualities associated with high rigor were observed more frequently in the teacher assignments we collected than the qualities associated with high relevance.

Table A-2. Ranking of Criterion Difficulty

Criterion	*Measure*	*Criterion Difficulty*
English/Language arts		
Teacher assignments		
Elaborated communication	Rigor	–1.86
Construction of knowledge	Rigor	–0.17
Student involvement	Relevance	0.96
Real-world connections	Relevance	1.07
Student work		
Language conventions and resources	Student work quality	–0.45
Elaborated communication	Student work quality	–0.43
Construction of knowledge	Student work quality	0.88
Mathematics		
Teacher assignments		
Effective communication	Rigor	–0.32
Problem solving and reasoning	Rigor	–0.26
Important math content	Rigor	0.05
Real-world connections	Relevance	0.13
Student involvement	Relevance	0.40
Student work		
Procedural knowledge	Student work quality	–1.66
Conceptual understanding	Student work quality	0.31
Effective communication	Student work quality	0.61
Problem solving and reasoning	Student work quality	0.73

Another psychometric property we examined was the distribution of difficulty scores. A larger variance in item difficulties leads to better-quality MFRM scores. The item difficulties for our measure of assignment relevance for ELA were 0.96 for Student Involvement and 1.07 for Real-World Connections. The logit difference between the two is so small that teacher assignments got either high scores or low scores on both criteria. Lack of variance in item difficulty measures for ELA assignment relevance scores may explain the low reliability of the measure reported earlier.

As a rule, the more variation in item difficulty, the better the quality of the Rasch scores, since the spread helps to differentiate student work of different levels of quality. For mathematics student work, there is a relatively reasonable spread of criterion difficulties, which help obtain higher reliability for student work quality measures. For English/language arts, however, two of the three items were almost equal in their levels of difficulty. For student work measures, this lack of variation may explain why the reliability statistic for English/language arts is lower than the one for mathematics.

Facet 3: Rater Severity

The goal of this analysis is to see whether the scorers applied the same levels of severity for a set of rubrics when evaluating student work. If scorers vary in their severity, we need to control for the differences when deriving the measures of assignment rigor, assignment relevance, or student work quality. Too many discrepancies among the scorers suggest that the training given to the scorers may not have been adequate; however, MFRM is used in cases where scorers vary in how they apply the scoring standards, and it takes into account differences in rater severity. Therefore, because some level of variation among raters is expected, the use of MFRM is justified.

We examined the reliability of rating by checking how often raters agreed when they scored the same criterion on the same assignment. The first column in Table A-3 lists the criteria on which each assignment was scored. On a given assignment, each criterion was scored by a different rater. To examine the reliability of raters, 75 of the English/language arts assignments and 150 of the mathematics assignments were scored by two different raters. Raters double-scored 75 student work papers in English/language arts and 300 in mathematics on two of the criteria. The scoring was designed so that at some point each rater would score the same criterion on the same assignment as the other 11 raters in the scoring. The first column shows the criteria, and the second shows the number of score points possible in each criterion. The score scales for the different criteria range from 3 to 6 points.[4]

The third column shows the percentage of observations that received the same rating from both scorers (i.e., perfect agreement), and the fourth column shows the percentage of observations that scored within one point. The English/language arts scorers had perfect agreement of 58% or more on all four of the assignments rubrics. They had at least 88% agreement within one point for all four rubrics. There was more variation among the mathematics scorers, with a range from 50% perfect agreement on Important Mathematical Content to 85% perfect agreement on the Student Involvement criterion.

Differences among ratings could be due to a variety of factors, such as differences in scorers' understanding of the scoring rubrics, differences in the effectiveness of scorer training, subjective differences among raters, and differences among the assignments and pieces of student work. Our goal is to achieve perfect agreement rates of 60% or higher. For those criteria for which the rates fall below 60%, we review our training procedures and refine our training to improve agreement rates during subsequent scoring sessions. Given that there is some variation in rater severity, we conclude that the use of the Many-Facet Rasch model was useful in controlling for these differences.

Table A-3. Agreement Rates on Assignments and Student Work Scored by Two Scorers

Criterion	*No. of Points Possible*	*Perfect Agreement*	*Agreement within 1 Point*
English/Language arts			
Teacher assignments			
Elaborated communication	4	58%	88%
Construction of knowledge	4	58%	88%
Real-world connections	4	74%	88%
Student involvement in crafting assignments	4	78%	97%
Student work			
Construction of knowledge	4	63%	94%
Elaborated communication	4	60%	90%
Language conventions and resources	6	69%	91%
Mathematics			
Teacher assignments			
Important mathematical content	4	50%	94%
Problem solving and reasoning	4	55%	90%
Effective communication	3	73%	95%
Real-world connections	4	67%	93%
Student involvement in crafting assignments	4	85%	95%
Student work			
Conceptual Understanding	4	68%	89%
Procedural knowledge	4	45%	79%
Problem solving and reasoning	4	83%	93%
Effective communication	4	73%	99%

Rater Drift

Each summer, we scored the data collected during the school year just completed. Since this report presents data collected during two successive school years, it is possible that there could be systematic differences in the values of the scores associated with the year the scoring session was conducted. This could occur if one year's raters were more or less severe in their ratings.

To compare how raters from year 1 and year 2 rated the same assignments and student work products, a subsample of assignments and student work that were collected and scored in the first year was rescored in the second year. Using this subsample, we conducted a statistical test for rater drift with MFRM, treating the scoring year as one of the facets in the Rasch equation. The scoring year variable is coded as a dummy variable (0 = first year, 1 = second year), so its effect can be interpreted as the size of drift, and the significance level can be determined. The results in Table A-4 show that the differences were statistically significant only for student work in mathematics, where the raters were more lenient in year 2. In other words, for the same pieces of student work, on average, scorers gave higher ratings in year 2.

Table A-4. Logit Estimates for Scoring Year Parameters

	Difference from Year 1 to Year 2	*Chi-square*	*P-value*
Teacher assignments			
English/language arts	0.24	3.00	0.08
Mathematics	0.02	0.00	0.89
Student work			
English/language arts	0.18	1.10	0.29
Mathematics	0.32	10.00	< 0.01

However, the MFRM adjusts for drift by controlling for rater severity, and raters who participated in each scoring session were treated as separate raters in the analysis.

We also examined the level of rater drift occurring in the scoring of student work products for each criterion. Drift toward more lenient standards from year 1 to 2 was noted in two math criteria, procedural knowledge and effective communication. The average scores student work received from raters increased from 2.4 to 2.7 for procedural knowledge and from 1.4 to 1.6 for effective communication; the differences in scores are statistically significant. The rest of the criteria for student work also moved in the same direction, but differences were not statistically significant.

IV. Creating Cut Scores for Categories

The three outcomes of interest—assignment rigor, assignment relevance, and student work quality—were each divided into four categories for this report: *little or none*, *limited*, *moderate*, and *substantial*. To do this, it was necessary to first determine what raw score from each criterion would correspond to each category. For most of the criteria it was easy, because most were on a four-point scale. For the ones that were on a three-point or six-point scale, the scoring leaders were consulted to see which score corresponded with which category. For example, the math scoring leader indicated that for criterion 3, a 1 would indicate little or no rigor, a 2 would indicate limited or moderate rigor, and a 3 would indicate substantial rigor.

The next step involved using the Rasch output to predict the Rasch scores that would correspond to the raw scores. Translating from the raw scores to the Rasch score is not an exact science, since the scores are also adjusted for rater severity.

For example the output for math rigor is shown below:

	Raw Score	*Logit Score*	*Max*	*Min*	*0-10 Measure*
Criterion 1	1	-2.85	3.61	-5.58	2.97
	2	-2.01	3.61	-5.58	3.88
	3	-0.38	3.61	-5.58	5.66
	4	0.75	3.61	-5.58	6.89
Criterion 2	1	-2.69	3.61	-5.58	3.14
	2	-1.29	3.61	-5.58	4.67
	3	-0.12	3.61	-5.58	5.94
	4	0.92	3.61	-5.58	7.07
Criterion 3	1	-2.28	3.61	-5.58	3.59
	2	-0.67	3.61	-5.58	5.34
	3	0.58	3.61	-5.58	6.70

	Crit 1	*Crit 2*	*Crit 3*	*Average*	*Cutoffs*
Little or no	1	1	1	3.24	0.00 to < 3.93
Limited	2	2	2	4.63	3.93 to < 5.14
Moderate	3	3	2	5.65	5.14 to < 6.27
Substantial	4	4	3	6.89	6.27 to 10.00

The average for little or no rigor is found by averaging together the 0-0 Rasch scores for assignments that received a 1 on criterion 1 (2.97), a 1 on criterion 2 (3.14), and a 1 on criterion 3 (3.59). The cutoffs were calculated by taking the midpoint between the average scores for pairs of successive categories, (e.g., the cutoff between limited and moderate, 5.14, is the midpoint between 4.63 and 5.65).

V. HLM Analyses Comparing the Rigor and Relevance of Teacher Assignments in New and Comprehensive High Schools

Given the nested structure of the teacher assignment data (i.e., assignments are nested within teachers), we used the hierarchical linear modeling (HLM) technique as the primary analytic method (Raudenbush & Bryk, 2002). In addition, since the quality of the teacher assignment is not directly observable and hence inevitably measured with error, we treated assignment rigor and relevance scores generated from the Many-Facet Rasch analyses as latent variables, taking advantage of the fact that the Rasch analysis provides both measures of the rigor and relevance of the assignment and their respective standard errors.

Specifically, we constructed a three-level HLM latent variable model, where level 1 is the measurement model, level 2 is the assignment level, and level 3 is the teacher level. The purpose of the measurement model at level 1 is to explicitly take into account the measurement errors in the Rasch scores of rigor and relevance in our analyses. It partitions the scores for each teacher assignment into a true score of the underlying latent measure and a measurement error. The measure of rigor and the measure of relevance were distinguished by two dummy indicator variables. The true values of the latent variables estimated at level 1 were then used as the outcomes at level 2 (assignment level) and modeled as a function of assignment type. The level-2 intercepts, which represent the average levels of rigor and relevance for each teacher, adjusted for assignment type, were further modeled as random effects predicted by a set of teacher and school characteristics in the teacher-level model at level 3.[5] Separate analyses were conducted for English/language arts and mathematics. The specification of the three-level model is as follows:

Level-1 Model (Measurement)

$$Y_{ijk} = \pi_{1jk}\,(\text{DUM\_RIGOR}) + \pi_{2jk}\,(\text{DUM\_RELEVANCE}) + \varepsilon_{ijk}$$

where

- The outcome Y_{ijk} is the observed Rasch score of latent variable i for assignment j given by teacher k, the two latent variables being the rigor and relevance of teacher assignments (i: 1 = rigor, 2 = relevance). The Rasch scores were weighted by the inverse of the standard errors of the measurement derived from the Many-Facet Rasch analysis.
- DUM_RIGOR and DUM_RELEVANCE are two dummy variables indicating the specific latent variables that generated the observed scores. The two dummies were un-centered at level 1 and also weighted by the inverse of the standard errors of the observed outcome.

- π_{1jk} and π_{2jk} are the true scores of the two latent variables for assignment j given by teacher k respectively.

- ε_{ijk} is a measurement error embedded in the observed score; $\varepsilon_{ijk} \sim N(0,1)$, given the weighting of both the dependent and independent variables.

Level-2 Model (Assignments)

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

Where

- CHALLENGING is a dummy variable indicating the type of teacher assignment (1= challenging, 0 = typical). It was grand mean centered and fixed at level 3.

- β_{10k} and β_{20k} are the mean rigor score and relevance score, respectively, for teacher k, adjusted for the type of assignment.

- r_{1jk} and r_{2jk} are the random effects associated with assignment j given by teacher k on the rigor and relevance of assignments, respectively.

Level-3 Model (Teachers)

$$\beta_{10k} = \gamma_{100} + \gamma_{101}(\text{YEARS}) + \gamma_{102}(\text{PRIORACH}) + \gamma_{103}(\text{PCT10}) + \gamma_{104}(\text{PCTEL}) + \gamma_{105}(\text{PCTSE}) + \gamma_{106}(\text{ZRISK}) + \gamma_{107}(\text{NEW}) + u_{10k}$$
$$\beta_{11k} = \gamma_{110}$$
$$\beta_{20k} = \gamma_{200} + \gamma_{201}(\text{YEARS}) + \gamma_{202}(\text{PRIORACH}) + \gamma_{203}(\text{PCT10}) + \gamma_{204}(\text{PCTEL}) + \gamma_{205}(\text{PCTSE}) + \gamma_{206}(\text{ZRISK}) + \gamma_{207}(\text{NEW}) + u_{20k}$$
$$\beta_{21k} = \gamma_{210}$$

where

- YEARS, PRIORACH, PCT10, PCTEL, and PCTSE were teacher-level control variables. YEARS represents teacher's number of years of teaching experience. PRIORACH is a measure of average ninth-grade achievement in ELA or math of the class taught by the teacher. PCT10, PCTEL, and PCTSE are measures of classroom composition, which represent the percentage of students in 10^{th} grade, percentage of students who are English learners, and percentage of students with special educational needs in the class taught by the teacher. All these teacher-level control variables were grand mean centered.

- ZRISK and NEW are measures of school characteristics. All teachers in the same school share the same value on these two measures. ZRISK is a school risk index composed of the percentage of students receiving free or reduced-price lunch and the percentage of underrepresented minority students, centered around its grand mean. NEW is an uncentered dummy variable indicating school type (1 = new, 0 = comprehensive).

- γ_{100} and γ_{200} are the grand means of the rigor and relevance of teacher assignments respectively, across all teachers, adjusted for assignment type, teaching experience, prior achievement of the class taught, classroom composition, school risk index, and school type (new vs. comprehensive).
- γ_{110} and γ_{210} are the mean differences in rigor and relevance, respectively, between challenging teacher assignments and typical assignments across all teachers.
- γ_{107} and γ_{207} represent the school type effects, which are the differences in rigor and relevance, respectively, between teachers in new schools and teachers in comprehensive schools, adjusted for control variables in the model.
- u_{10k} and u_{20k} are the random effects associated with teacher k on the rigor and relevance of assignment respectively adjusting for control variables in the model.

By removing the school type variable (NEW) from the above model, we were able to obtain empirical Bayes (EB) estimates of rigor and relevance for each teacher, adjusted for assignment type, years of teaching experience, class prior achievement, classroom composition, and the school risk index (not adjusted for school type). These teacher-level estimates were then aggregated to the school level to produce the rigor and relevance measures for individual schools.

Students' prior achievement was an important covariate in this model; it was necessary to control for the overall student prior achievement level of the class taught by the teacher when analyzing teacher assignments. However, the creation of a consistent achievement measure was hampered by two problems. The first was that the standardized tests often differ among schools, and the second was the prevalence of missing test scores. We addressed the first problem by converting scores to normal curve equivalents. All major commercial tests provide norming documents that translate scale scores to percentile ranks or normal curve equivalents of the original norming sample. Although the norming samples for different tests are not identical, we assume that nationally representative norming samples are similar enough to provide a consistent metric for using prior achievement scores based on different tests as a covariate in the HLM models.

The prevalence of missing test scores was handled via imputation. Imputed scores were generated via best-subsets regression using student-level eighth-grade test scores (where available), ethnicity, sex, English learner status, and free or reduced-price lunch status. In some cases, the scores for entire schools (i.e., those not reporting a nationally normed test score) required imputation. The imputed scores were then combined with the observed scores as a covariate in the HLM model.

For the eight Washington state schools, ninth-grade test scores were missing for an average of 24% of students (school ranged from 9% to 83% missing data). For the 16 schools outside of Washington state, ninth-grade scores were missing for 83% of the students in the ELA sample (schools ranged from 28% to 100% missing) and 69% in the math sample (schools ranged from 16% to 100% missing). Of the 16 non-Washington state schools, 9 were missing usable ninth-grade reading and math scores, mainly because 6 were using state-developed tests that could not be put on the same metric with nationally normed tests.

Estimates of fixed effects and variance components based on the above model are presented in Table A-5 and Table A-6, respectively.

Table A-5A. Estimates of Fixed Effects Based on an HLM Comparing the Rigor of Teacher Assignments in New and Comprehensive High Schools–English/Language Arts

	English/Language Arts Rigor					
	Model 1		*Model 2*		*Model 3*	
Measure	***Coefficient***	***SE***	***Coefficient***	***SE***	***Coefficient***	***SE***
Mean rigor (intercept)	4.18***	0.16	4.81***	0.15	4.75***	0.15
New school (vs. comprehensive)	**0.11**	**0.29**	**0.59***	**0.26**	**0.74***	**0.31**
Challenging assignment			1.06***	0.16	1.05***	0.16
Average 9th-grade test score					0.02	0.01
Years teaching experience					–0.002	0.01
Percent 10th graders in class					–1.02	10.01
Percent English learners					–0.20	0.32
Percent special education					0.23	1.51
School risk index					0.11	0.18
N of assignments	404					
N of teachers	57					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-5B. Estimates of Fixed Effects Based on an HLM Comparing the Relevance of Teacher Assignments in New and Comprehensive High Schools—English/Language Arts

	English/Language Arts Relevance					
	Model 1		*Model 2*		*Model 3*	
Measure	***Coefficient***	***SE***	***Coefficient***	***SE***	***Coefficient***	***SE***
Mean relevance (intercept)	2.86***	0.12	2.89***	0.09	2.88***	0.12
New school (vs. comprehensive)	**1.35*****	**0.39**	**1.20*****	**0.25**	**1.25*****	**0.28**
Challenging assignment			0.74***	0.16	0.75***	0.15
Average 9th-grade test score					–0.007	0.01
Years teaching experience					–0.006	0.01
Percent 10th graders in class					1.26	0.79
Percent English learners					–0.31	0.52
Percent special education					1.62	1.20
School risk index					–0.20	0.14
N of assignments	404					
N of teachers	57					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-5C. Estimates of Fixed Effects Based on an HLM Comparing the Rigor of Teacher Assignments in New and Comprehensive High Schools–Mathematics

	Mathematics Rigor					
	Model 1		*Model 2*		*Model 3*	
Measure	*Coefficient*	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Mean rigor (intercept)	4.81***	0.15	4.17***	0.15	3.95***	0.16
New school (vs. comprehensive)	**0.60***	**0.25**	**0.10**	**0.25**	**0.58**	**0.35**
Challenging assignment			1.03***	0.18	1.04***	0.18
Average 9th-grade test score					–0.02~	0.01
Years teaching experience					–0.005	0.01
Percent 10th graders in class					–1.13*	0.56
Percent English learners					–0.33	0.47
Percent special education					–2.66**	0.55
School risk index					–0.35~	0.17
N of assignments	413					
N of teachers	52					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-5D. Estimates of Fixed Effects Based on an HLM Comparing the Relevance of Teacher Assignments in New and Comprehensive High Schools–Mathematics

	Mathematics Relevance					
	Model 1		*Model 2*		*Model 3*	
Measure	*Coefficient*	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Mean relevance (intercept)	2.93***	0.09	2.86***	0.12	2.77***	0.11
New school (vs. comprehensive)	**1.17*****	**0.25**	**1.35*****	**0.35**	**1.58*****	**0.41**
Challenging assignment			1.21***	0.20	1.22***	0.20
Average 9th-grade test score					–0.02*	0.01
Years teaching experience					0.0001	0.02
Percent 10th graders in class					0.52	0.58
Percent English learners					–0.85	0.60
Percent special education					0.12	0.60
School risk index					–-0.36~	0.19
N of assignments	413					
N of teachers	52					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-6A. Estimates of Variance Components Based on Alternative Model Specifications – Rigor

Variance Components	*Fully Unconditional*	*Full Model (Model 3)*
English/Language arts		
Within student	2.25 (82%)	1.87 (81%)
Between students	0.48 (18%)	0.43 (19%)
Total	2.73 (100%)	2.30 (100%)
Mathematics		
Within student	1.16 (63%)	0.92 (70%)
Between students	0.68 (37%)	0.40 (30%)
Total	1.84 (100%)	1.32 (100%)

Table A-6B. Estimates of Variance Components Based on Alternative Model Specifications – Relevance

Variance Components	*Fully Unconditional*	*Full Model (Model 3)*
English/Language arts		
Within student	0.29 (28%)	0.21 (38%)
Between students	0.74 (72%)	0.35 (63%)
Total	1.03 (100%)	0.56 (100%)
Mathematics		
Within student	1.38 (50%)	1.02 (59%)
Between students	1.39 (50%)	0.70 (41%)
Total	2.77 (100%)	1.72 (100%)

VI. HLM Analyses Comparing the Quality of Teacher Feedback on Student Work in New and Comprehensive High Schools

In addition to the rigor and relevance of teacher assignments, we also examined the quality of teacher feedback on student work in new and comprehensive high schools, using a three-level HLM model. The model conceives the feedback teachers gave to each piece of student work as nested within teachers, who were in turn nested within schools. The specification of the model is as follows:

Level-1 Model (student work)

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}\,(\text{CHALLENGING}) + \varepsilon_{ijk}$$

where

- The outcome Y_{ijk} is the feedback quality score for student work i associated with assignment given by teacher j at school k.

- CHALLENGING is a dummy variable indicating the type of teacher assignment that is associated with student work i within teacher j at school k (1 = challenging, 0 = typical). It was grand mean centered and fixed at level 2.
- π_{0jk} is the average feedback quality score for teacher j at school k, adjusted for assignment type.
- π_{1jk} is the effect of assignment type on the quality of feedback on student work i associated with assignment given by teacher j at school k.
- ε_{ijk} is an error term associated with each piece of student work, adjusted for assignment type.

Level-2 Model (teachers)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{YEARS}) + \beta_{02k}(\text{PCT10}) + \beta_{03k}(\text{PCTEL}) + \beta_{04k}(\text{PCTSE}) + \beta_{05k}(\text{PRIORACH}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

where

- YEARS, PCT10, PCTEL, PCTSE, and PRIORACH were teacher-level control variables. YEARS represents teacher's number of years of teaching experience. PRIORACH is a measure of average ninth-grade achievement in ELA or math of the class taught by the teacher. PCT10, PCTEL, and PCTSE are measures of classroom composition, which represent the percentage of students in 10^{th} grade, percentage of students who are English learners, and percentage of students with special educational needs in the class taught by the teacher. All these teacher-level control variables were grand mean centered and fixed at level 3.
- β_{00k} is the average feedback quality score for school k, adjusted for assignment type and teacher/classroom characteristics.
- β_{01k}, β_{02k}, β_{03k}, β_{04k}, and β_{05k} represent the effects of teacher/classroom characteristics on the average feedback score for teachers at school k, adjusted for assignment type.
- β_{10k} is the average effect of assignment type on feedback quality across all teachers at school k.
- r_{0jk} is a random effect on feedback quality associated with teacher j at school k, adjusted for assignment type and teacher/classroom characteristics.

Level-3 Model (schools)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(NEW) + \gamma_{002}(ZRISK) + u_{00k}$$
$$\beta_{01k} = \gamma_{010}$$
$$\beta_{02k} = \gamma_{020}$$
$$\beta_{03k} = \gamma_{030}$$
$$\beta_{04k} = \gamma_{040}$$
$$\beta_{05k} = \gamma_{050}$$

where

- NEW and ZRISK are measures of school characteristics. NEW is an uncentered dummy variable indicating school type (1 = new, 0 = comprehensive). ZRISK is a school risk index composed of the percentage of students receiving free or reduced-price lunch and the percentage of underrepresented minority students, centered around its grand mean.
- γ_{000} is the grand mean of feedback quality across all teachers at all schools adjusted for assignment type, teaching experience, prior achievement of the class taught, classroom composition, school risk index, and school type (new vs. comprehensive).
- γ_{001} represents the difference between new and comprehensive high schools in the quality of teacher feedback, adjusted for other control variables in the model.
- γ_{002} represents the effect of the school risk index on the quality of teacher feedback, adjusted for other control variables in the model.
- γ_{010}, γ_{020}, γ_{030}, γ_{040}, and γ_{050} are the average effects of teacher/classroom characteristics on school-level average feedback quality across all schools, each adjusted for the other control variables in the model.
- u_{00k} is a random effect on school average feedback quality associated with school k, adjusted for the control variables in the model.

Table A7. Estimates of Fixed Effects Based on an HLM Model Comparing the Usefulness of Teacher Feedback in New and Comprehensive High Schools - English/Language Arts and Mathematics

	English/Language Arts		*Mathematics*	
Measure	*Coefficient*	*SE*	*Coefficient*	*SE*
Mean (intercept)	1.69***	0.10	1.34***	0.06
New school (vs. comprehensive)	**0.43***	**0.19**	**0.14***	**0.11**
Challenging assignment	0.31*	0.13	0.16	0.07
Average 9th-grade test score	–0.001	0.01	–0.004	0.003
Years teaching experience	0.01**	0.005	–0.003	0.01
Percent 10th graders in class	–0.49	0.47	–0.48**	0.17
Percent English learners	0.28~	0.16	0.18	0.14
Percent special education	–0.76*	0.38	0.09	0.20
School risk index	–0.13	0.10	–0.005	0.07
N of pieces of student work	831	821		
N of teachers	57	52		
N of new schools	12	12		
N of comprehensive schools	12	12		

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

VII. HLM Analyses Comparing the Quality of Student Work in New and Comprehensive High Schools

Like the teacher assignment data, student work data also have a nested structure, that is, multiple pieces of student work are nested within students. Moreover, the quality of student work is not directly observable and hence inevitably measured with error. Therefore, we analyzed the student work data by using an HLM latent variable model similar to the one used for the teacher assignment analyses, treating student work scores generated from the Many-Facet Rasch analyses as a latent variable.

Specifically, in the three-level HLM latent variable model for student work analyses, level 1 is the measurement model, level 2 is the student work level, and level 3 is the student level. The purpose of the measurement model at level 1 is to explicitly take into account the measurement errors in the Rasch scores of student work quality by conceiving the Rasch score of each piece of student work as comprising the true score of the underlying latent measure of student work quality and a measurement error. The true scores of the latent variable, student work quality, estimated at level 1 were then used as the outcomes at level 2 (student work level) and modeled as a function of assignment type. The level-2 intercepts, which represent the average level of work quality for each student, adjusted for assignment type, were further modeled as random effects predicted by a set of student and school characteristics in the student-level model at level 3.[6] Separate analyses were conducted for English/language arts and mathematics. The specification of the three-level model is as follows:

Level-1 Model (Measurement):

$$Y_{jk} = \pi_{jk}(\text{DUMSW\_WT}) + \varepsilon_{jk}$$

Where

- The outcome Y_{jk} is the observed Rasch score of the quality of student work j submitted by student k. The Rasch scores were weighted by the inverse of the standard errors of the measurement derived from the Many-Facet Rasch model.
- DUMSW_WT is a dummy variable representing the latent variable that generated the observed scores of student work quality. It was un-centered at level 1 and weighted by the inverse of the standard errors of the Rasch scores.
- π_{jk} is the true score of the latent variable for student work j submitted by student k.
- ε_{jk} is an measurement error embedded in the observed score; $\varepsilon_{jk} \sim N(0,1)$, given the weighting of both the dependent and independent variables.

Level-2 Model (Student work):

$$\pi_{jk} = \beta_{0k} + \beta_{1k}(\text{CHALLENGING}) + r_{jk}$$

where

- CHALLENGING is a dummy variable indicating the type of teacher assignment that is associated with student work j from student k (1 = challenging, 0 = typical). It was grand mean centered and fixed at level 3.
- β_{0k} is the mean student work score for student k, adjusted for the type of assignment.
- r_{jk} is a random effect associated with student work j submitted by student k on the quality of the work.

Level-3 Model (Students):

$$\beta_{0k} = \gamma_{00} + \gamma_{01}(\text{PRIORACH}) + \gamma_{02}(\text{BLACK}) + \gamma_{03}(\text{ASIAN}) + \gamma_{04}(\text{HISPANIC}) + \gamma_{05}(\text{INDIAN}) + \gamma_{06}(\text{MALE}) + \gamma_{07}(\text{NEW}) + \gamma_{08}(\text{ZRISK}) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

where

- PRIORACH is a grand-mean-centered measure of students' ninth-grade achievement. Since students in different schools took different types of standardized tests in ninth grade, we converted their ninth-grade scores to normal curve equivalents based on the norming tables provided by test developers.[7]

- BLACK, ASIAN, HISPANIC, and INDIAN are dummy variables for student race, with WHITE as the reference group. MALE is a dummy variable for student gender. All these student-level control variables were grand mean centered.
- NEW and ZRISK are measures of school characteristics. All students in the same school share the same value on these two measures. ZRISK is a school risk index based on the percentage of minority students and the percentage of students eligible for free or reduced-price lunch. It was centered around its grand mean. NEW is an uncentered dummy variable indicating school type (1 = new, 0 = comprehensive).
- γ_{00} is the grand mean of the quality of student work across all students, adjusted for assignment type, student characteristics, the school risk index, and school type.
- γ_{10} is the average effect of assignment type (i.e., challenging vs. typical) on the quality of student work across all students.
- γ_{07} represents the school type effect, which is the difference in the quality of student work between new and comprehensive high schools, adjusted for assignment type, student characteristics, and the school risk index.
- u_{0k} is a random effect associated with student k on the quality of student work.

Estimates of fixed effects and variance components based on the above model are presented in Table A-7 and Table A-8, respectively.

Table A-8A. Estimates of Fixed Effects Based on an HLM Comparing the Quality of Student Work Between New and Comprehensive High Schools—English/Language Arts

	English/Language Arts					
	Model 1		*Model 2*		*Model 3*	
Measure	***Coefficient***	***SE***	***Coefficient***	***SE***	***Coefficient***	***SE***
Mean Student work (intercept)	4.13***	0.09	4.10***	0.09	3.87***	0.10
New school (vs. comprehensive)	**0.34***	**0.15**	**0.30~**	**0.15**	**0.88*****	**0.16**
Challenging assignment			1.35***	0.12	1.31***	0.12
9th-grade test score – Z score					0.19*	0.09
Black					0.17	0.21
Asian					0.00	0.27
Hispanic					–0.06	0.24
Indian					–0.16	0.45
Male					–0.08	0.14
School risk index					–0.49***	0.08
N of pieces of student work	776					
N of students	408					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-8B. Estimates of Fixed Effects Based on an HLM Comparing the Quality of Student Work Between New and Comprehensive High Schools—Mathematics

	Mathematics					
	Model 1		*Model 2*		*Model 3*	
Measure	***Coefficient***	***SE***	***Coefficient***	***SE***	***Coefficient***	***SE***
Mean Student work (intercept)	4.09***	0.05	4.09***	0.05	4.00***	0.05
New school (vs. comprehensive)	**–0.50*****	**0.09**	**–0.50*****	**0.86**	**–0.29****	**0.10**
Challenging assignment			0.54***	0.08	0.53***	0.08
9th-grade test score – Z score					0.16**	0.05
Black					0.02	0.14
Asian					0.07	0.15
Hispanic					0.15	0.13
Indian					0.22~	0.13
Male					–0.14~	0.08
School risk index					–0.16**	0.05
N of pieces of student work	783					
N of students	387					
N of new schools	12					
N of comprehensive schools	12					

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-9. Estimates of Variance Components Based on Alternative Model Specifications

Variance Components	*Fully Unconditional*	*Full Model (Model 3)*
English/Language arts		
Within student	2.07 (83%)	1.41 (73%)
Between students	0.43 (17%)	0.53 (27%)
Total	2.50 (100%)	1.94 (100%)
Mathematics		
Within student	0.58 (73%)	0.50 (81%)
Between students	0.21 (27%)	0.115 (19%)
Total	0.79 (100%)	0.62 (100%)

VIII. HLM Analyses Assessing the Relationships between the Rigor and Relevance of Teacher Assignments and the Quality of Student Work

We employed a two-step approach to assess the relationships between the rigor and relevance of teacher assignment and the quality of student work. First, we constructed a three-level HLM latent variable model (Model 1) to derive estimates of assignment-level student work quality while taking into account the measurement error in our Rasch measure of student work. This assignment-level measure of student work was then linked to measures of the rigor and relevance of teacher assignments in a second three-level HLM latent variable model (Model 2), which computes the correlations among the three measures based on estimates of variance components at both the assignment level and the teacher level.

Model 1: A Three-Level HLM Latent Variable Model for Deriving Estimates of an Assignment-Level Measure of Student Work Quality

Specifically, Model 1 is specified as follows:

Level-1 Model (Measurement):

$$Y_{jk} = \pi_{jk}(\text{DUMSW\_WT}) + \varepsilon_{jk}$$

Level-2 Model (Student work):

$$\pi_{jk} = \beta_{0k} + r_{jk}$$

Level-3 Model (Assignment):

$$\beta_{0k} = \gamma_{00} + \gamma_{01}(\text{CHALLENGING}) + u_{0k}$$

Similar to the model for comparing new and comprehensive high schools in the quality of student work (see Section VII), the above model specifies the Rasch score of a given piece of student work (i.e., the outcome) as composed of the true score of the latent measure and a measurement error at level 1. Both the Rasch scores of student work and the dummy indicator for the latent measure of student work quality, DUMSW_WT, were weighted by the inverse of the standard error of the measurement derived from the Many-Facet Rasch analysis, such that the residual at level 1 has a mean of 0 and a variance of 1.

In the student work model at level 2, the true score of the latent measure of student work quality estimated at level 1 (π_{jk}) was modeled as a random effect, which varies randomly across different pieces of student work linked to a given teacher assignment. The intercept of the level-2 model (β_{0k}), which represents the average level of student work quality for all pieces of student work linked to a given assignment, was further modeled at level 3 (assignment level) as a random effect and as a function of assignment type (challenging vs. typical). We computed the level of student work quality for each assignment, adjusted for assignment type, as the sum of empirical Bayes estimates of the level-3 intercept and the level-3 residual (i.e., $\gamma_{00} + u_{0k}$) based on the above model. This assignment-level measure of student work quality was subsequently linked to the rigor and relevance of teacher assignments in Model 2 detailed below.

Model 2: A Three-Level HLM Latent Variable Model for Relationships between the Rigor and Relevance of Teacher Assignments and Student Work Quality

Like Model 1, Model 2 is also a three-level latent variable model that takes into account the measurement error of the latent outcomes. However, unlike Model 1, which has only one latent outcome, Model 2 incorporates three latent outcomes: the rigor and relevance of teacher assignment and the quality of student work. The first two latent outcomes were measured by the Rasch scores of rigor and relevance of teacher assignments, and the third latent outcome was measured by the assignment-level estimates of student work quality derived from Model 1. To enable simultaneous estimation of the three latent outcomes, we stacked assignment data and student work data in such a way that each assignment occupies three rows in the dataset, one for each outcome (rigor, relevance, and assignment-level student work quality). The specification of Model 2 at each level is as follows:

Level-1 Model (Measurement):

$$Y_{ijk} = \pi_{1jk}(\text{DUMRIGOR\_WT}) + \pi_{2jk}(\text{DUMREL\_WT}) + \pi_{3jk}(\text{DUMSW\_WT}) + \varepsilon_{ijk}$$

Level-2 Model (Assignment):

$$\pi_{1jk} = \beta_{10k} + \beta_{11k} \text{(CHALLENGING)} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + \beta_{21k} \text{(CHALLENGING)} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k} + r_{3jk}$$

Level-3 Model (Teacher):

$$\beta_{10k} = \gamma_{100} + \gamma_{101} \text{(ZRISK)} + \gamma_{102} \text{(YRS\_TCH)}$$
$$+ \gamma_{103} \text{(MEAN\_PRIORACH)} + u_{10k}$$
$$\beta_{11k} = \gamma_{110}$$
$$\beta_{20k} = \gamma_{200} + \gamma_{201} \text{(ZRISK)} + \gamma_{202} \text{(YRS\_TCH)}$$
$$+ \gamma_{203} \text{(MEAN\_PRIORACH)} + u_{20k}$$
$$\beta_{21k} = \gamma_{210}$$
$$\beta_{30k} = \gamma_{300} + \gamma_{301} \text{(ZRISK)} + \gamma_{302} \text{(YRS\_TCH)}$$
$$+ \gamma_{303} \text{(MEAN\_PRIORACH)} + u_{30k}$$

In the measurement model at level 1, the three latent outcomes are distinguished by three dummy indicator variables: DUMRIGOR_WT, DUMREL_WT, and DUMSW_WT. Both the outcome and the dummy indicator variables were weighted by the inverse of the standard error of the measurement such that the level-1 residuals had a mean of 0 and a variance of 1. At level 2, the true scores of the rigor and relevance of teacher assignments (π_{1jk} and π_{2jk}) and the quality of student work (π_{3jk}) were set to vary randomly across assignments within teachers. Assignment type, CHALLENGING, was used as a predictor for both the rigor and relevance measures.[8] The level-2 intercepts (β_{10k}, β_{20k}, and β_{30k}), which represent the average levels of rigor and relevance of teacher assignments and the quality of student work for each teacher, were further modeled as random effects and as a function of teaching experience (YRS_TCH), classroom average prior achievement (MEAN_PRIORACH), and the school risk index (ZRISK) at level 3.

The variance components for the three latent outcomes estimated on the basis of the above model are provided in the Table A-9. It shows that most of the variance in rigor and student work quality lies between assignments within teachers (e.g., 77% and 83%, respectively in ELA), rather than between teachers for both ELA and math. Both rigor and student work quality show greater variation at both the assignment level and the teacher level in ELA compared with math. The measure of relevance, however, exhibits a very different pattern: it varies more widely between teachers than between assignments within teachers in ELA and varies to a similar extent at both the teacher and the assignment levels in math.

Table A-10. Estimates of Variance Components of Rigor and Relevance of Teacher Assignments and the Quality of Student Work at the Assignment Level and the Teacher Level

	Variance Components		
	Assignment Level	*Teacher Level*	*Total*
English/Language Arts			
N	414	57	
Rigor	1.89 (77%)	0.56 (23%)	2.45 (100%)
Relevance	0.24 (24%)	0.74 (76%)	0.98 (100%)
Student work	1.47 (83%)	0.30 (17%)	1.77 (100%)
Mathematics			
N	438	52	
Rigor	0.91 (65%)	0.50 (35%)	1.41 (100%)
Relevance	1.02 (50%)	1.03 (50%)	2.05 (100%)
Student work	0.21 (91%)	0.02 (9%)	0.23 (100%)

In addition to estimates of variance components, the HLM program also generated correlations among the three latent outcomes at both the assignment level and the teacher level, which are listed in Table A-10.

Table A-11. Correlations between the Rigor and Relevance of Teacher Assignments and the Quality of Student Work at the Assignment Level and the Teacher Level

	Assignment Level			*Teacher Level*		
	Rigor	*Relevance*	*Student Work*	*Rigor*	*Relevance*	*Student Work*
English/Language arts						
Rigor	1.00			1.00		
Relevance	0.68	1.00		0.70	1.00	
Student work	0.81	0.98	1.00	0.94	0.89	1.00
Mathematics						
Rigor	1.00			1.00		
Relevance	0.83	1.00		0.58	1.00	
Student work	0.80	0.65	1.00	0.70	-0.01	1.00

Building on the parameter estimates of the teacher-level model in Model 2, we further formulated a final latent variable model as follows, where the covariate-adjusted teacher-level estimates of the rigor and relevance of assignments (β_{10k} and β_{20k}) were used as predictors for the teacher-level estimate of the quality of student work (β_{30k}):

$$\beta_{30k} = \gamma_{300} + \gamma_{301}\text{(ZRISK)} + \gamma_{302}\text{(YRS\_TCH)} + \gamma_{303}\text{(MEAN\_PRIORACH)} + \gamma_{304}(\beta_{10k}) + \gamma_{305}(\beta_{20k}) + u_{30k}$$

In the above model, the coefficients for the rigor and relevance measures (γ_{304} and γ_{305}) represent the effects of the rigor and relevance of teacher assignments on the quality of student work, respectively, controlling for other covariates in the model. Estimates of these effects, as well as the effects of other covariates, are presented in Table A-11. Note that the final latent variable model is a teacher-level model and the results in Table A-11 are teacher-level direct effects, which correspond to the teacher-level correlations, not the assignment-level correlations presented in Table A-10.

Table A-12. HLM Estimates of Teacher Characteristics on Student Work Quality (Teacher-level effects based on final latent variable model formulation of Model 2)

	English/Language arts		*Mathematics*	
Measure	*Coefficient*	*Standard Error*	*Coefficient*	*Standard Error*
Rigor	**0.45~**	**0.23**	**0.22***	**0.09**
Relevance	**0.30**	**0.18**	**–0.10**	**0.06**
Class 9th-grade test score (Z-score)	0.25*	0.12	0.10~	0.05
Years of teaching	–0.01	0.01	–0.002	0.01
School risk index	–0.14	0.10	–0.06	0.06

Note: ~ $p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

IX. HLM Analyses Assessing the Relationships between the Quality of Student Work and Student 10th-Grade Achievement

Our working hypothesis is that high-quality student work translates into higher achievement. Thus, the goal of our analysis was to assess the effect of quality of student work on achievement test scores. The unit of analysis was students, and the dependent variable was 10th grade test scores. The independent variable was the quality of student work, derived by the Many-Facet Rasch model, and the statistical controls included a prior-year achievement score, class average prior-year achievement score, minority status, and school level risk index. To account for the fact that our data has a nested structure (Students are nested within classrooms) and that the residuals therefore are not independent, we used two-level hierarchical linear models (HLM). The equation below is a general representation of our model. Level 1 is the student level and level 2 is the teacher level; thus, the teacher-specific intercepts are set to vary as random effects (i.e., u_{0j}), while the predictor's coefficients are fixed. For ease of interpretation, all continuous variables were standardized with a grand mean of 0 and standard deviation of 1 (i.e., Z-score).

Level-1 Model (Students)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{PRIORACH}) + \beta_{2j}(\text{STUDENT WORK})$$
$$+ \beta_{3j}(\text{RIGOR}) + \beta_{4j}(\text{RELEVANCE}) + \beta_{5j}(\text{MINORITY}) + r_{ij}$$

Level-2 Model (Teachers)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{PRIORACH}) + \gamma_{02}(\text{ZRISK}) + u_{0j}$$
$$\beta_{gj} = \gamma_{g0}$$

The dependent variable in our analysis is 10th grade test score. The tests were administered in the spring; thus, they represent the level of achievement students exhibited toward the end of the 10th grade. We obtained the test scores in English/language arts and mathematics from existing datasets. For students in the eight schools in the state of Washington, we used the 10th grade scores from WASL tests. The rest of the students had different types of tests. In six of the schools, the tests were normable–that is, when the test publishers provided information as to how their scores convert into national normal curve equivalent (NCE) scores.[9] The national tests for which norming information was available were the CAT-6 (California) and the SAT-9 (Minnesota and Rhode Island). By using the national norms to convert the test scores on a common metric, we assume that different tests captured the same construct and the norming samples were from populations that have the same distribution of mathematics achievement. The rest of the schools either were unable to provide test score data or were able to provide only data for tests that could not be pooled into a common metric with the data at other schools. Thus, we used two types of test scores in our analysis: WASL test scores and national test scores.

The independent variable in this analysis is the quality of student work. As detailed earlier in Section III, we used Many-Facet Rasch Measurement to derive the measures of quality for student work. For this analysis of student achievement, however, the unit of analysis is not pieces of student work but students. One simple approach to getting a student-level value for student work (each student may have up to three pieces of student work) would have been to create a simple average of student work scores for each student. This approach, however, would ignore the fact that some scores are more reliable than others. To benefit from the fact that our Rasch scores for student work come with standard errors, we took a latent variable approach. We used HLM to derive student-level measures using the inverse of the standard errors as a weighting factor. In this way, unreliable measures are pulled toward the grand mean of the measures, thus preventing outlier measures from having undue influence on the results of the analysis. Section VII of this appendix has further details on this approach.

The other independent variables are the rigor and relevance of teacher assignments. Originally, these scores were measured at the assignment level. For this analysis, we derived teacher-level measures by using a procedure similar to that described above for student work scores.

We used HLM to derive the teacher-level scores that were adjusted for measurement errors, as well as factors associated with assignments, classrooms, and teachers. Section V describes this approach.

The statistical controls include ninth-grade test scores (prior-year scores) and their classroom average. As detailed earlier, for the six schools in the achievement analysis that had achievement scores that could be put on the same metric, we created a classroom average of prior-year scores to control for contextual effects. We group mean centered both individual-level scores and classroom-average scores with a standard deviation of 1. Other statistical controls were minority status and a school risk index. Minority is defined as being African-American, Hispanic, or American Indian, while non-minority was defined as either white or Asian. The school risk index was based on the percentage of minority students and percentage of students eligible for free or reduced-price lunch.

Missing ninth-grade test scores were imputed by using a best-subsets regression model, as described earlier. Missing data were also an issue for 10^{th} grade test scores, but since the 10^{th} grade scores are the dependent variable, they could not be imputed. This resulted in the achievement analyses being conducted with achievement data from two subsets of schools. For the six schools using nationally normed achievement tests, the percentage of missing test score data averaged 54% of the student sample for English/language arts (schools ranged from 32% to 68% missing) and 21% for math (schools ranged from 5% to 32% missing). For the Washington state schools, 12% of the eligible students were missing 10^{th} grade test scores in both reading and math (schools ranged from 0% to 24% missing), and those students were not included in the achievement analyses.

Student achievement has been shown to be associated with a multitude of factors, including the pedagogy employed in the classroom, students' home backgrounds, and the composition of students in the classroom. The strongest predictors of student achievement are typically the students' prior-year test scores and the classroom average prior-year test scores. In most of our analyses, these control variables helped explain a large proportion of the variance in the dependent variables. Thus, within our analyses, we are able remove most of the influence of prior achievement, as well as factors that may be associated with it, such as students' social and family background characteristics. In the current model, we also control for teacher characteristics with a measure of teachers' years of teaching experience. However, we believe our analysis would have been more robust if we had measures of quality and nature of the classroom instruction.

Analytical Samples and Stepwise Analytic Approach

We have four analytical samples defined by test type (WASL score and national score) and by subject (ELA and mathematics). For WASL test score analysis, mathematics data included 154 students who were taught by 23 teachers in 8 schools, while ELA data included 149 students taught by 24 teachers in the same 8 schools. For national test score analysis, mathematics data included 66 students who were taught by 11 teachers in 6 schools, while ELA data included 76 students taught by 14 teachers in the same 6 schools.

For each analytical sample described above, we run the following 10 models of student achievement. Model 1 is an intercept-only model from which we learn the size of variance at student level and teacher level. Model 2 tests how covariates are related to the outcome, student achievement. Models 3, 4, and 5, evaluate the strength of the association between student achievement and rigor, relevance, and student work, respectively. The same evaluation is conducted in the following models, 6, 7, and 8, but this time with a full set of covariates. Controlling for covariates, Model 9 evaluates how rigor and relevance jointly affect the outcome; finally, Model 10 includes all three—rigor, relevance, and student work—for an understanding of the system of relationships among the predictors. Tables A-12 and A-13 show the ELA results for the national sample and the Washington state sample. Tables A-14 and A-15 show the results for mathematics.

Table A-13. Estimates of Fixed Effects and Random Components–English/Language Arts–10th Grade National Tests (SAT-9 and CAT-6)

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	0.11	0.21	0.43**	0.14	–0.03	0.23	–0.14	0.20	0.08	0.20
Rigor	.	.	.	.	0.31	0.20	.	.	.	.
Relevance	.	.	.	.	.	.	0.53*	0.19	.	.
Student work	.	.	.	.	.	.	.	.	0.29*	0.13
9th-grade test	.	.	0.49**	0.12	.	.	.	.	.	.
Avg. 9th-gr. test	.	.	0.21	0.16	.	.	.	.	.	.
Minority	.	.	0.27	0.20	.	.	.	.	.	.
School risk index	.	.	-0.15	0.16	.	.	.	.	.	

	Model 6		*Model 7*		*Model 8*		*Model 9*		*Model 10*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	0.33~	0.16	0.18	0.21	0.28~	0.14	0.18	0.25	0.20	0.27
Rigor	0.14	0.12	.	.	.	.	0.01	0.20	–0.02	0.22
Relevance	.	.	0.27	0.19	.	.	0.27	0.34	0.10	0.37
Student work	.	.	.	.	0.34**	0.12	.	.	0.33**	0.12
9th-grade test	0.49**	0.12	0.49**	0.12	0.53**	0.12	0.49**	0.12	0.53**	0.12
Avg. 9th-gr. test	0.21	0.16	0.18	0.16	0.23	0.15	0.18	0.16	0.22	0.17
Minority	0.25	0.20	0.23	0.20	0.22	0.19	0.23	0.20	0.20	0.19
School risk index	–0.08	0.17	0.05	0.21	0.10	0.18	0.06	0.25	0.18	0.27
N of schools	6									
N of teachers	14									
N of students	76									

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-14. Estimates of Fixed Effects and Random Components-English/Language Arts-10th Grade WASL Test

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	–0.02	0.14	0.67**	0.12	0.00	0.14	0.03	0.21	–0.05	0.13
Rigor	.	.	.	.	0.10	0.20	.	.	.	.
Relevance	.	.	.	.	.	.	0.11	0.33	.	.
Student work	.	.	.	.	.	.	.	.	0.32**	0.09
9th-grade test	.	.	0.48**	0.08	.	.	.	.	.	.
Avg. 9th-gr. test	.	.	0.46**	0.12	.	.	.	.	.	.
Minority	.	.	0.15	0.16	.	.	.	.	.	.
School risk index	.	.	0.45**	0.13	.	.	.	.	.	

	Model 6		*Model 7*		*Model 8*		*Model 9*		*Model 10*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	0.67**	0.12	0.66**	0.15	0.63**	0.12	0.64**	0.16	0.63**	0.17
Rigor	–0.03	0.10	.	.	.	.	-0.05	0.15	–0.09	0.16
Relevance	.	.	0.00	0.17	.	.	0.06	0.25	0.04	0.26
Student work	.	.	.	.	0.09	0.07	.	.	0.11	0.08
9th-grade test	0.48**	0.08	0.48**	0.08	0.46**	0.08	0.48**	0.08	0.46**	0.08
Avg. 9th-gr. test	0.46**	0.13	0.46**	0.13	0.42**	0.13	0.46**	0.13	0.43**	0.14
Minority	0.15	0.16	0.15	0.16	0.15	0.16	0.16	0.16	0.15	0.16
School risk index	0.45**	0.13	0.45**	0.14	0.43**	0.14	0.44**	0.14	–0.42*	0.15
N of schools	8									
N of teachers	23									
N of students	154									

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table A-15. Estimates of Fixed Effects and Random Components-Mathematics-10^{th} Grade National Tests (SAT-9 and CAT-6)

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	0.02	0.22	0.53**	0.17	–0.02	0.24	0.02	0.23	0.13	0.19
Rigor	.	.	.	.	–0.08	0.21	.	.	.	.
Relevance	.	.	.	.	.	.	–0.06	0.33	.	.
Student work	.	.	.	.	.	.	.	.	0.44**	0.16
9^{th}-grade test	.	.	0.44**	0.13	.	.	.	.	.	.
Avg. 9^{th}-gr. test	.	.	0.58~	0.27	.	.	.	.	.	.
Minority	.	.	0.00	0.33	.	.	.	.	.	.
School risk index	.	.	0.07	0.19	.	.	.	.	.	

	Model 6		*Model 7*		*Model 8*		*Model 9*		*Model 10*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	0.50*	0.21	0.54*	0.18	0.55*	0.17	0.49~	0.23	0.47~	0.21
Rigor	–0.07	0.18	.	.	.	.	–0.09	0.21	–0.17	0.20
Relevance	.	.	0.01	0.21	.	.	0.06	0.25	0.14	0.23
Student work	.	.	.	.	0.32*	0.16	.	.	0.35*	0.16
9^{th}-grade test	0.45**	0.13	0.44**	0.13	0.40**	0.13	0.45**	0.13	0.40**	0.13
Avg. 9^{th}-gr. test	0.62~	0.29	0.58~	0.28	0.44	0.27	0.63~	0.31	0.51	0.30
Minority	–0.01	0.33	0.00	0.33	0.17	0.34	0.00	0.34	0.19	0.34
School risk index	0.14	0.27	0.06	0.22	–0.09	0.20	0.13	0.28	0.01	0.27
N of schools	6									
N of teachers	11									
N of students	66									

Note: ~ $p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table A-16. Estimates of Fixed Effects and Random Components – Mathematics – 10th Grade WASL Test

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	–0.04	0.15	0.49**	0.17	0.00	0.15	–0.26	0.17	–0.05	0.15
Rigor	.	.	.	.	0.29~	0.17	.	.	.	.
Relevance	.	.	.	.	.	.	–0.56*	0.24	.	.
Student work	.	.	.	.	.	.	.	.	0.04	0.07
9th-grade test	.	.	0.49**	0.08	.	.	.	.	.	.
Avg. 9th-gr. test	.	.	0.24	0.16	.	.	.	.	.	.
Minority	.	.	–0.08	0.16	.	.	.	.	.	.
School risk index	.	.	0.04	0.19	.	.	.	.	.	
	Model 6		*Model 7*		*Model 8*		*Model 9*		*Model 10*	
Measures	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***	***Coeff***	***SE***
Intercept	–0.47*	0.17	0.60**	0.19	0.48*	0.17	–0.56*	0.20	–0.54*	0.21
Rigor	–0.19	0.16	.	.	.	.	–0.14	0.18	–0.18	0.18
Relevance	.	.	–0.30	0.28	.	.	–0.22	0.30	–0.22	0.31
Student work	.	.	.	.	0.07	0.06	.	.	0.08	0.06
9th-grade test	0.50**	0.08	0.49**	0.08	0.51**	0.09	0.49**	0.08	0.51**	0.09
Avg. 9th-gr. test	0.16	0.17	0.13	0.19	0.21	0.17	0.10	0.20	0.05	0.20
Minority	–0.08	0.16	–0.08	0.16	–0.07	0.16	–0.08	0.16	–0.06	0.16
School risk index	–0.11	0.23	–0.12	0.25	0.08	0.20	–0.20	0.27	–0.19	0.28
N of schools	8									
N of teachers	24									
N of students	149									

Note: ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

X. Data Collections for Future Analysis

The schedule for future data collections is shown in Table A-16. The analyses shown in this report are from the first 2 years of data collection. As described earlier, in 2002–03, we collected data from 8 large comprehensive high schools in Washington state that were planning for a transition into small learning communities. In 2003–04, we collected assignments and student work from 12 new high schools and 4 additional comprehensive schools planning to transition into small learning communities. For the 2004–05 school year, we have collected data from these same 12 new high schools so we can examine changes over time in these schools. In addition, we have collected data from 7 large traditional high schools that serve as comparisons for the 12 new schools. The 8 Washington comprehensive schools from 2002–03 have been redesigned into small learning communities and were in their second year of redesign in 2004–05. We have collected assignments and student work from at least

2 new learning communities for each of the redesigned schools. Where possible, we have solicited work from the same teachers as in our initial data collections in these schools.

In the final year of data collection (2005–06), we will follow up on the 4 comprehensive high schools from Year 2; by then, they will have been operating as small learning communities for more than a year. As with the Washington schools, we will collect data from at least 2 of the small learning communities created from each of the 4 originally large comprehensive high schools, and attempt to get assignments from the same teachers.

Table A-17: Data Collection Schedule – School Type and Number of Schools

School Type	*Year 1 (2002–03)*	*Year 2 (2003–04)*	*Year 3 (2004–05)*	*Year 4 (2005–06)*
New schools		12	12 (Same as Y2)	
Comprehensive comparisons for new			7	
Comprehensive before transition to small learning communities	8 (Washington)	4 (National)		
Small learning communities			16 (Washington)[10]	8 (National)[11]

REFERENCES

Linacre, J. M. (1989). *Many-Facet Rasch Measurement*. Chicago, IL: Mesa Press.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. (Vol. 1). Thousand Oaks, CA: Sage Publications.

Endnotes

Appendix Endnotes

[1] Many of the student outcomes related to student attitudes, but student attitude outcomes are not included in this report.

[2] This section describes the development of the measures for rigor and relevance of teacher assignments, and the same techniques were used to measure the quality of student work.

[3] All of the year 1 assignments and some of the year 2 assignments were scored independently by two raters and would have two scores for each criterion. Some of the year 1assignments were rescored in year 2, so those assignments would have three scores for each criterion on an assignment.

[4] The numbers in the table can be compared with what would be expected to occur by chance. For a criterion with 4 possible points, even if the raters assigned the scores completely at random, there would still be a 25% chance of perfect agreement (4 in 16) and a 63% chance of agreement within one point (10 in 16). Similarly, for a criterion with 3 possible points, there would be a 33% chance of perfect agreement (3 in 9) and a 78% chance of agreement within one point (7 in 9). For a criterion with 6 possible points, there would be a 17% chance of perfect agreement (6 in 36) and a 44% chance of agreement within one point (16 in 36).

[5] Ideally, we would further nest teachers within schools. However, the HLM software used could accommodate only three levels of analysis. Therefore, we added school-level measures to the teacher-level model at level 3. The SAS procedure, PROC MIXED, does allow us to construct a four-level model (i.e., measurement model, student work, student, school); however, the program had difficulty converging and was unable to produce a reliable estimate of student-level variance.

[6] Ideally, we would further nest students within schools. However, the HLM software program can accommodate a maximum of three levels. Therefore, we added school-level measures to the student-level model at level 3.

[7] Although the norming samples for different standardized tests are not identical, we assume that they are similar enough to provide a consistent measure for prior achievement as a student-level control variable in the HLM model.

[8] Assignment type was not used as a predictor for the student work measure in the assignment-level model, because the student work measure was already adjusted for assignment type in Model 1.

[9] For example, the SAT-9 and the CAT-6 are given to a sample of students that is representative of students in the United States and can therefore be translated into national norms.

[10] Data will be collected from two or more small learning communities for each of the eight Washington schools.

[11] Data will be collected from two or more small learning communities for each of the four redesigned schools in the national sample.

National

Evaluation of

High School

Transformation