Authentic Assessment and Student Performance in Inclusive Schools

M. Bruce King,
Jennifer Schroeder,
University of Wisconsin-Madison
& David Chawszczewski,
National-Louis University

In the current context of school reform, teaching and learning of high intellectual quality (e.g., Newmann & Wehlage, 1995) and teaching for understanding (e.g., Cohen, McLaughlin, & Talbert, 1993) offer compelling alternatives to more traditional forms of instruction focused on basic skills and content. In schools that restructure around a vision of authentic pedagogy and student achievement, students learn more and learning occurs more equitably across student groups (Newmann, Marks, & Gamoran, 1996). At the same time, calls for reform in special education focus on the inclusion of students with disabilities in general education classes (e.g., Lipsky & Gartner, 1996).

In this brief, we investigate the intersection of these reform movements. Specifically, we address two questions:

1. In secondary schools with inclusionary practices, to what extent are teacher-designed assessments authentic?

2. How do students with and without disabilities perform on these assessments?

INSTITUTE MISSION

The mission of the institute is to expand the current knowledge base related to practices and policies in secondary schools that enhance learning, achievement, and postschool outcomes for students with disabilities.

▶▶▶▶ ◀◀◀◀

CORE RESEARCH QUESTIONS

1. What are critical features of instruction, assessment, and support strategies that promote authentic understanding, and achievement (and performance) for all students?

2. How have changes in authentic inclusive learning and schooling practices affected the school and postschool outcomes (and their interaction) for students with disabilities (collectively and disaggregated) using frames of reference focused on equity, value added, and accountability?

3. How do schools accommodate district and state outcome assessments, and how do such accommodations affect the participation in, reporting of, and validity of assessment?

4. In schools evolving toward authentic and inclusive instruction, what are the roles and expectations of stakeholders as they engage in planning for secondary and postsecondary experiences?

5. What contextual factors are required to support and sustain the development of secondary-level learning environments that promote authentic understanding, achievement, and performances for all students?

6. What strategies are effective in providing both information and support to policymakers, school administrators, teachers, human service personnel, and the community so they utilize the findings to create and support learning environments that promote authentic understanding, achievement, and performance for all students?

Data come from high schools that are participating in a 5-year national study conducted by the Research Institute on Secondary Education Reform (RISER) for Youth with Disabilities at the University of Wisconsin–Madison. These schools, selected from a national search, demonstrate varying degrees of schoolwide inclusive and authentic practices.

Authentic and Inclusive Reform

Most recent education reforms have been generated with limited research on or consideration of the implications of the reforms for students with disabilities. But changes in special education do not evolve in isolation from broader national policy interests and issues. Thus, RISER is focused on schools engaged in reform efforts that include students with disabilities and seeks to identify educational practices that benefit *all* students.

RISER is grounded in the model of Schools of Authentic and Inclusive Learning (SAIL; see Hanley-Maxwell, Phelps, Braden, & Warren, 1999). Central to the SAIL model is the concept of authentic achievement and pedagogy. Developed as part of a national study of school restructuring (Newmann & Wehlage, 1995), authentic teaching and learning provide the framework for the study of classroom practices that include both students with and students without disabilities. Authentic pedagogy is consistent with the recent emphasis on constructivist teaching, which has been advocated as a productive alternative to traditional instructional approaches in special education. These traditional approaches have been criticized for operating from a deficit model in which learning expectations for students with disabilities are significantly lowered (Trent, Artiles, & Englert, 1998).

Authentic intellectual work is defined by three general characteristics (Newmann & Wehlage, 1995). The first characteristic is *construction of knowledge*. In the conventional curriculum, students

---

Authentic Intellectual Work

- *Construction of Knowledge*

- *Disciplined Inquiry*

- *Value Beyond School*

---

identify the knowledge that others have produced (e.g., by recognizing the difference between verbs and nouns, labeling parts of a plant, or matching historical events to their dates). In authentic work, however, students go beyond memorizing and repeating facts, information, definitions, or formulas to produce new knowledge or meaning. This kind of work involves higher order thinking in which students analyze, interpret, or evaluate information in a novel way. The mere reproduction of knowledge does not constitute authentic academic achievement.

A second defining feature of authentic achievement is its reliance on a particular type of cognitive work called *disciplined inquiry*. Disciplined inquiry consists of (a) using a knowledge base, (b) striving for in-depth understanding of relevant knowledge and concepts, and (c) expressing conclusions through elaborated communication. By contrast, much of the traditional pedagogy in schools asks students to show only a superficial awareness of a vast number of topics and requires only brief responses from students (e.g., true-false, multiple-choice, or short answers).

A third characteristic of authentic achievement is that it has *value beyond school*—that is, it has meaning or value apart from documenting or certifying the learner's competence. In authentic work, students make connections between what they are learning

and important personal or social issues. Achievements of this sort—whether a performance, exhibition, or written communication—actually influence others and thus have a value that is missing in tasks such as quizzes and standardized tests that only assess an individual student's knowledge or skills.

> *Practice, memorization, and drill are necessary . . . . But teachers should provide as much opportunity as possible for ALL students, including those with disabilities, to engage in and become competent in authentic intellectual work.*

These three characteristics are the basis for the standards we are using to assess the intellectual quality of teaching and learning in participating schools—namely, construction of knowledge; disciplined inquiry through elaborated written communication; and value beyond school through connection to students' lives. (See sidebar for examples of standards for scoring teachers' assignments in writing and math. For all standards and scoring criteria used in this study, see the RISER Web site, http://www.wcer.wisc.edu/riser/.) Teachers' lessons, assignments, and student work can score high on some of these characteristics but lower on others, and one would not expect all activities to score high on all three all of the time. Practice, memorization, and drill are necessary to build the knowledge and skills needed for more challenging tasks or to prepare for exams required for promotion or advancement. But teachers should provide as much opportunity as possible for all students, including those with disabilities, to engage in and become competent in challenging intellectual work.

Also central to the SAIL model is the inclusion of special education students in the mainstream of the general education curriculum. Critics point to potentially serious problems with inclusion (see Hanley-Maxwell et al., 1999, for a summary). For example, at the classroom level, there might be negative attitudes of teachers toward students with disabilities and an emphasis on large group instruction that takes little account of individual learning needs. At the institutional level, there might be insufficient time for classroom teachers to collaborate with special educators and a general lack of professional development to prepare teachers to address the demands of inclusive classrooms.

These are important considerations. However, inclusion is prominent in the national reform agenda of special education. Proponents might argue that the above considerations are challenges to be addressed and that with appropriate accommodations for students' disabilities, both special and regular education students should benefit from inclusive environments. Across the U.S., students with a wide range of disabilities are being educated in inclusive settings, although inclusion at the secondary level is still rare (Thousand, Rosenberg, Bishop, & Villa, 1997). In this study of secondary schools that practice inclusion, we explore the degree of authenticity in teacher-designed assessments and the performance of regular and special education students on these assessments.

Research Methodology and Analysis

In addressing the first of the six core research questions investigated by RISER (page 2) —what are critical features of instruction, assessment, and support strategies that promote authentic understanding, achievement, and performance for all students?—we present findings from two sets of data collected during the 1999–2000 school year. The first data set (*whole class*) included assessment tasks and student work for those tasks from 8 teachers in each of two schools. These 16 teachers

<table>
<tr>
<td>

Standards for Teachers' Assignments in Writing

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

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

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

</td>
<td>

Standards for Teachers' Assignments in Math

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

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

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

</td>
</tr>
</table>

had special education students in their classes and emphasized intellectual quality in their teaching. They represented the main academic subject areas of language arts, science, math, and social studies— one teacher in each area from Grades 9–10, and one in each area from Grades 11–12 at each school. The teachers submitted one assessment task that they considered to be an important indicator of what students learned in one of their classes, along with the work the students in that class completed for that task.

The second data set (*matched pairs*) came from 35 teachers in three of the schools (Schroeder, 2000). The teachers represented the main academic subject areas of language arts, science, math, and social studies (8, 7, 10, and 10 teachers, respectively) across Grades 9–12, and they all had special education students in their classes. These teachers also submitted one assessment task that

they considered to be an important indicator of what students learned in one of their classes. However, this set of data differed from the first in that teachers submitted work completed by just two students in the classroom, one student with a disability and one student without a disability, allowing for comparisons between students with and without disabilities on each task. Teachers also submitted a checklist of accommodations they made, if any, for both regular and special education students.

For both data sets, each task was rated on the extent to which the intellectual work it required met each of three standards corresponding to the general characteristics of authentic achievement— construction of knowledge, in-depth understanding through elaborated written communication, and connection to students' lives. For example, a writing task that scored high on construction of knowledge would meet the following criterion: "The task's

dominant expectation is for students to interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information." To score high on elaborated written communication, a mathematics task would need to ask explicitly for generalization and support in students' responses; that is, the task would require students to show through writing their solution paths and to explain the solution paths with evidence such as models or examples. To score high on the third standard—connection to students' lives—a science task would need to present students with a scientific question, issue, or problem that they would have actually encountered or would be likely to encounter in their daily lives; it would ask students to make connections between the topic and real-world situations.

Student work was also evaluated on three standards consistent with the characteristics of authentic intellectual work,[1] but these standards varied somewhat by subject areas. The standards for student work in math, science, and social studies were analysis, disciplinary concepts, and elaborated written communication. The standards for student work in writing were construction of knowledge, forms and conventions, and elaborated written communication.

Ratings for teacher tasks and for student work were achieved through similar processes. For the teacher tasks from the first data set ($n = 16$), six raters were collectively trained in using each standard's rubric. Two raters scrutinized each task description and independently assigned a score. Scores were then compared and any discrepancies negotiated either by discussing application of the rubric or by obtaining a third-party rating. In this manner, all of the tasks were assigned ratings through a consistent process that yielded numerical values. The level of exact

agreement between raters for task authenticity was 77.8%. Percent agreement between raters climbed to 97.2% when adjacent scores (i.e., one off) were used in the analyses. For the teacher tasks from the second data set ($n = 35$), two raters were collectively trained in using each standard's rubric and then independently assigned scores to each task. Percent agreement analyses were run on 20% of these tasks. The overall level of exact agreement between raters for task authenticity was 90.0%. Percent agreement between the raters for task authenticity was 100% when adjacent scores were used in the analyses.

In scoring the student work in the first data set ($n = 244$), 22 teachers from the schools participating in the study served as raters. After collective training and practice sessions, each standard's rubric was applied to the complete body of student work. Students' work was assigned to teacher-raters from schools other than the students' own, with each artifact being scored twice. Once paired scores were compared, a third party adjudicated any discrepancies. As with the teacher tasks, the process yielded a set of numerical scores for each piece of student work. The overall level of exact agreement between raters for work authenticity was 47.1%. However, percent agreement between raters for work authenticity climbed to 88.4% when adjacent scores (i.e., one off) were used in the analyses. For the student work from the second data set ($n = 70$), two raters were trained in using each standard's rubric and then assigned scores to the work. Twenty percent of the work samples in this set were scored twice. The overall level of exact agreement between raters for work authenticity was 75%. Percent agreement between the raters climbed to 100% when adjacent scores were used in the analyses.

For both sets of data, scores assigned to the tasks and student work for each of the three standards of authenticity were added to yield two overall scores, one for authenticity of the task and one for authenticity of work produced by students. The scores for each of the standards and the two overall scores were then compared and statistical analyses run to

---

[1]Student work was not rated on the third general characteristic of authentic achievement, value beyond school, due to logistical limitations of the study.

determine if any differences existed between standards, between academic subjects, or between students with and without disabilities. Correlational analyses were also run on the overall scores to determine if any relationships existed between task authenticity and authenticity of work produced by students with and without disabilities. We report these results below.

Findings (Data Set 1, Whole Class)

Overall degree of authenticity of tasks. Across the 16 classes, the mean rating for task authenticity on all submitted tasks was 6.53 ($SD = 1.33$).[2] Task authenticity scores can range from a low of 3 to a high of 10, which means that the mean score across all tasks fell in the middle of the range of possible scores. Despite this fact, the actual range for the scores on the assessment tasks included in this sample was from 3 to 8. Therefore, no task received the highest score possible for task authenticity, whereas one received the lowest score.

Across the 16 teachers in the four subject areas, the first two standards (construction of knowledge and elaborated written communication) received roughly equal emphasis on the tasks. The mean score for construction of knowledge was 2.24 (out of 3; $SD = 0.75$), and for elaborated written communication, 3.18 (out of 4; $SD = 0.81$). Tasks in social studies, science, and writing scored consistently higher on construction of knowledge and elaborated written communication than did math tasks. Standard 3, connection to students' lives, averaged 1.12 (out of 3; $SD = 0.49$), with all the tasks but one scoring a 1. This result exemplifies the persisting difficulty of developing assignments that ask students to address real-world problems and to explore the connections between topics or concepts and these problems.

---

[2]The standard deviation (SD) is a measure of how much scores deviate from the mean.

Previous research has shown that student performance in math, social studies, and writing is higher in classes with higher levels of authentic pedagogy (Avery, 1999; Newmann & Associates, 1996; Newmann, Lopez, & Bryk, 1998; Newmann et al., 1996). We now explore whether this relationship holds in our study, both for regular and special education students.

Overall degree of authenticity of student work. For the 16 tasks submitted, the mean overall rating for the authenticity of work produced by students was 7.21 ($SD = 2.41$). Overall student work authenticity scores can range from a low of 3 to a high of 12, which means that the mean score across all student work fell close to the middle of the range of possible scores. The range of scores for the student work included in this sample was from 4 to 12. Therefore, some student work did receive the highest score possible for work authenticity, but none received the lowest score.

The authenticity ratings given to student work were further compared by student disability status. The scores given to work produced by students without

> *Students with disabilities who were given more authentic tasks performed considerably better than students with disabilities who were given less demanding tasks.*

disabilities were compared to the scores given to work produced by students with disabilities to determine if there were any significant differences between the work produced by the two groups. Overall, the mean rating of work authenticity for students without disabilities was 7.42 ($SD = 2.47$) and for students with disabilities was 6.54 ($SD = 2.05$). This difference was significant ($p < .05$), indicating that students with disabilities produced work lower in authenticity than that produced by their nondisabled peers.

Relationship between tasks and student achievement. Last, we summarize findings on (a) the relationship between task authenticity and student achievement on the tasks and (b) achievement results for students with and without disabilities. The first important finding is that, consistent with previous research, there was a significant relationship ($r = .62$)[3] between the authenticity of task demands and the authenticity of the work that students produced. That is, task demands that were rated lower in authenticity were associated with student work that was rated lower in authenticity. Conversely, task demands that were higher in authenticity were associated with student work that was also higher in authenticity. This relationship was the same for tasks and work produced by students with and without disabilities.

Categorizing tasks as below average in task authenticity (< 6.5) or above average in task authenticity (≥ 6.5) provides a further illustration of this relationship. The average authenticity score for student work when task demands were *below average* in authenticity was 6.24 ($SD = 2.27$). When task authenticity demands were *above average*, however, the average authenticity score for student work was 8.43 ($SD = 2.01$), a difference of more than two points (see Figure 1).

When task demands and student work were analyzed by student disability status, similar results were found (see Figure 2). On tasks that were below average in authenticity, students without disabilities produced work that received an average score of 6.42 ($SD = 2.39$). Students with disabilities produced work that received an average score of 5.63 ($SD = 1.66$) when

---

[3]Correlation coefficients range in value from –1 to 1 and are a measure of the relationship between two variables. A value of –1 indicates that there is a perfect inverse relationship between two variables, whereas a value of 1 indicates a one-to-one correspondence. A value of 0 indicates that there is no relationship between variables. Values of .6 or above are considered to indicate a strong relationship between variables.

given the same task demands. This score is slightly lower than that produced by their nondisabled peers, but the difference is not statistically significant.

When students were given task demands that were above average in authenticity, students without disabilities produced work that received an average score of 8.62 ($SD = 2.00$). Students with disabilities produced work that received an average score of 7.72 ($SD = 1.92$) when given the same task demands—again, a slightly lower score than that of their nondisabled peers.
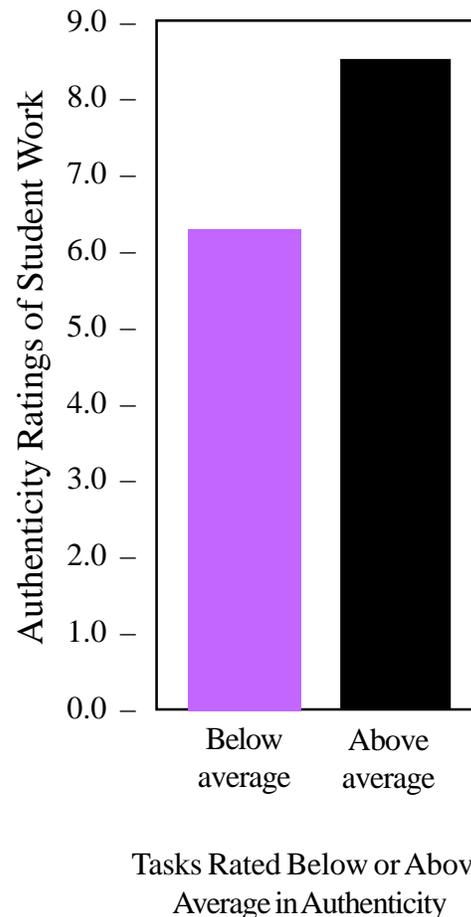

Tasks Rated Below or Above
Average in Authenticity

Figure 1. Mean ratings for authenticity of all student work when task demands are categorized as below or above average (first data set).
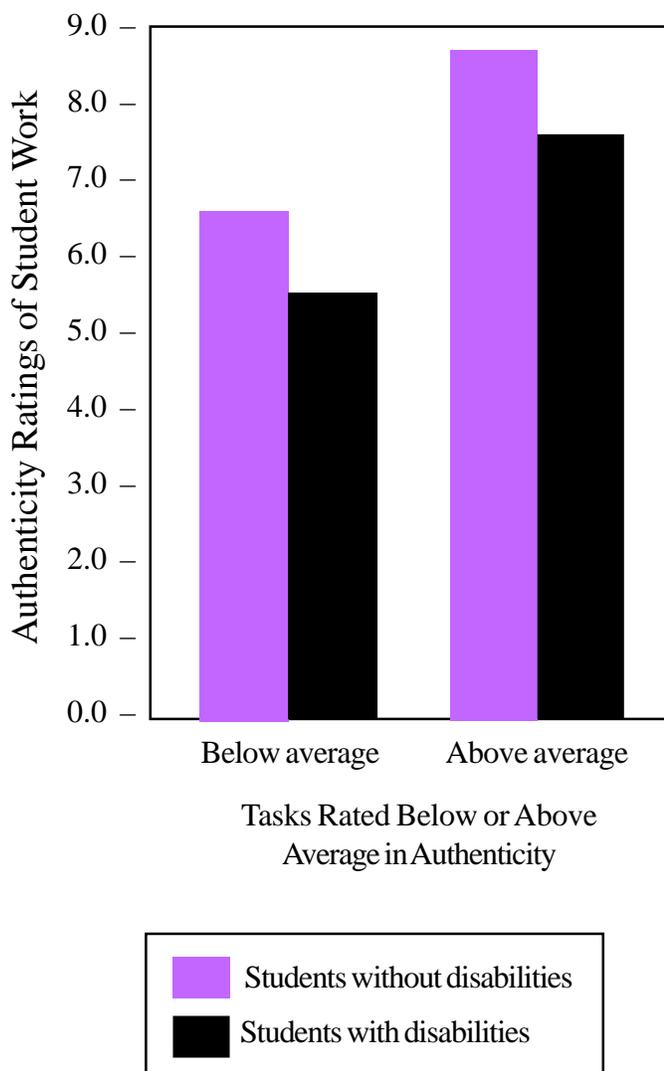
Figure 2. Mean ratings for authenticity of student work for students with and without disabilities relative to tasks rated below or above agerage in authenticity (first data set).

Although students with disabilities did not score, on average, as well as students without disabilities, we note two important trends. First, students with disabilities who were given higher scoring (i.e., above-average) tasks performed considerably better (7.72) than students with disabilities who were given below-average tasks (5.63). That is, special education students in these classes who received tasks with higher

intellectual challenge outperformed those who received tasks with less challenge. However, this difference was not statistically significant ($p = .057$).

Second, students with disabilities who were given higher scoring (i.e., above-average) tasks performed better (7.72) than students *without* disabilities who were given below-average tasks (6.42). This difference was statistically significant ($p < .05$). Special education students in these classes who received tasks with higher intellectual challenge outperformed their nondisabled peers who received tasks with less challenge. We consider some implications of these findings in the last section.

> *Special education students who received tasks with higher intellectual challenge outperformed their nondisabled peers who received tasks with less challenge.*

Findings (Data Set 2, Matched Pairs)

The matching of pairs of students in the second set of data allows for much of the same information to be gathered about tasks and student work. However, differences in the information gathered during data collection also allow for comparisons within pairs of students. This additional information is reported below.

Overall degree of authenticity of tasks. Across the 35 teachers in the second data set, the mean rating for task authenticity on all tasks was 7.30 ($SD = 2.09$). This average fell just above the middle of the range of possible scores (slightly higher than the first data set, which had a mean of 6.53). The actual range for the scores on the assessment tasks included in this data set was from 3 to 10. Therefore, some tasks in this data set, unlike those in the first data set, did receive the highest score possible for task authenticity.

These data yield an additional comparison. Ratings of task authenticity were compared for the tasks given to students with and without disabilities to determine whether the accommodations given to students changed the intellectual demands of the tasks. For example, an accommodation that involved eliminating certain parts of a task could lower task authenticity if the parts eliminated were those requiring students to analyze information (construction of knowledge), elaborate on their explanations through extended writing (elaborated written communication), or connect the topic to their lives (connection to students' lives). Accommodations could conceivably increase the authenticity of a task, although none did so in this set of data.

Although the task was generally the same for each pair of students in the second data set, some differences were found in task authenticity. Because of accommodations, students without disabilities received tasks with an overall mean rating of 7.43 ($SD = 2.12$), whereas students with disabilities received tasks with an overall mean rating of 7.17 ($SD = 2.06$). This difference, though small, is statistically significant ($p < .05$). Whether this difference matters in the classroom is unclear. However, the evidence that indicates that task authenticity and the authenticity of student work are related suggests that changes in task demands due to accommodations may be important in determining what students produce. We note, however, that for the vast majority of tasks, accommodations made no difference in the degree of intellectual demands. Figure 3 shows the percentage of tasks given to students with disabilities that received the same, lower, or higher authenticity ratings due to accommodations when compared to the tasks given to students without disabilities.
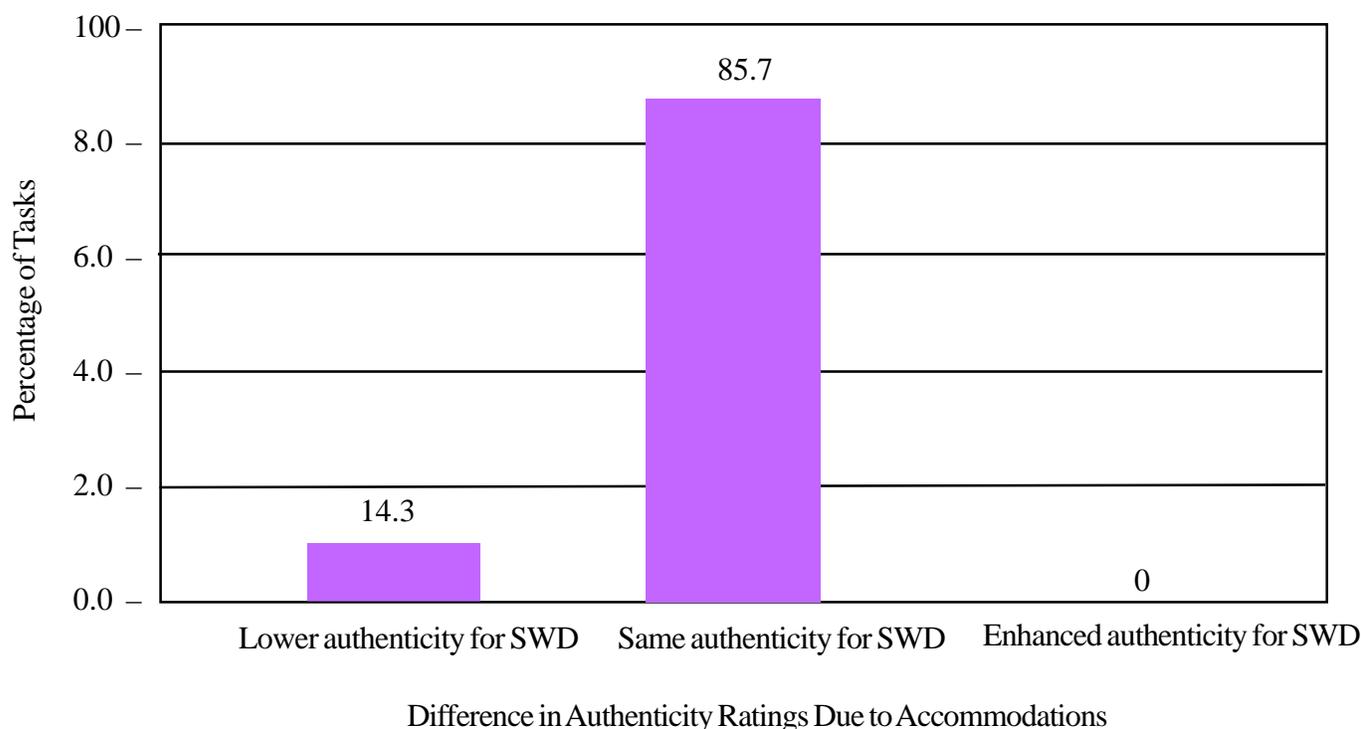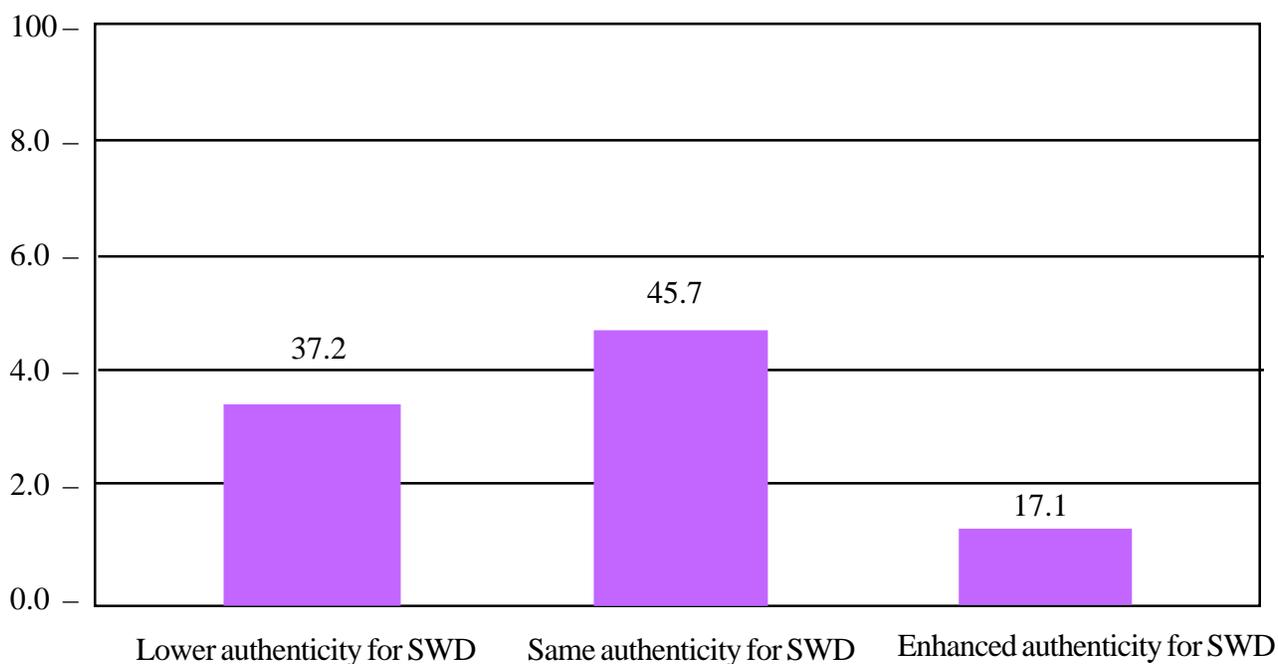

Figure 3. Percentage of tasks given to students with disabilities (SWD) receiving authenticity ratings lower or higher than, or the same as, tasks given to students without disabilities due to accommodations (second data set).

10

Overall degree of authenticity of student work. For the 35 tasks submitted, the mean overall rating for the authenticity of work produced by students was 7.47 (*SD* = 2.64). The mean score across all student work fell in the middle of the range of possible scores. The range of scores for the student work included in this sample was from 3 to 12.

The authenticity ratings given to student work in the second data set, as in the first, were compared by student disability status. The mean rating of work authenticity for students without disabilities was 8.03 (*SD* = 2.64), and for students with disabilities it was 6.91 (*SD* = 2.65). This difference was significant ($p <$ .01), indicating that students with disabilities produced work that was rated lower in authenticity than that produced by their nondisabled peers. However, despite this overall difference, it is interesting to note that whereas 37% of the students with disabilities produced work that was lower in authenticity than that produced by their matched nondisabled peer, 62% produced work that was the same, or *higher*, in authenticity than that produced by their matched peer (see Figure 4). Additionally, of the 37% (13 students) who produced work lower in authenticity than their peer, 4 had received tasks that were lower in authenticity as well.

Authenticity Ratings for Work by Students with Disabilities Compared to Work by Students without Disabilities

Figure 4. Percentage of work produced by students with disabilities (SWD) receiving authenticity rating slower or higher than, or the same as, the work produced by students without disabilities (second data set).

> *Sixty-two percent of the students with disabilities produced work that was the same, or higher, in authenticity than that produced by their nondisabled peer.*

Relationship between tasks and student achievement. Consistent with previous research and the data provided by the first data set, there was a significant relationship ($r = .68$)[4] between the authenticity of task demands and the authenticity of the work that students produced. That is, task demands that were rated lower in authenticity were associated with student work that was rated lower in authenticity. Conversely, task demands that were higher in authenticity were associated with student work that was also higher in authenticity.

Accommodations

The classrooms that we investigated included both students with and students without disabilities. Although current legislation (e.g., the Individuals with Disabilities Education Act Amendments of 1997) calls for the inclusion of students with disabilities in the least restrictive environment, which is often considered to be the general education classroom, simply putting students with disabilities into the general education classroom is not enough to guarantee their access to the general education curriculum. To benefit from the general education setting and to be able to complete the same tasks as their peers, students with disabilities often require accommodations (McGee, Mutch, & Leyland, 1993). Thus, we collected information in the second data set from teachers about the changes, or accommodations, that they made for their students (both with and without disabilities).

[4]See #4 above.

Twenty-five students without disabilities (71%) received accommodations for the given assessment tasks. These students received an average of 6 accommodations. In contrast, all 35 students with disabilities (100%) received accommodations. These students received an average of 18 accommodations. Accommodations ranged from giving encouragement to complete the task to changing the requirements of the task. Figure 5 shows the 10 most common accommodations given to students with and without disabilities.

As mentioned previously, accommodations may change the authenticity of the tasks students are asked to complete. For the tasks collected in the second data set, accommodations did change task authenticity, effectively lowering the authenticity ratings of 14%. However, even given this effect on task authenticity, it is important to note that accommodations are intended to allow students with disabilities to successfully complete tasks that they would otherwise not be able to access. Therefore, it would be a mistake to conclude that because of their potential to lower authenticity, accommodations are detrimental. Rather, accommodations, if used appropriately, should be viewed as helping students to access complex, authentic tasks.[5]

Conclusions

Teachers who use more authentic assessments elicit more authentic work from students with and without disabilities. As these data demonstrate, teachers who design and give assessment tasks that call for various forms of higher order thinking, requiring analysis or interpretation, in-depth understanding, direct connections to the field under study, and an appeal to an audience beyond the classroom, will enable

[5]Braden, Schroeder, & Buckley (2000) present a framework for implementing assessment accommodations. Significantly, they assert, "Assessments should retain authenticity, even if they are modified to a simpler skill level" (p. 8).
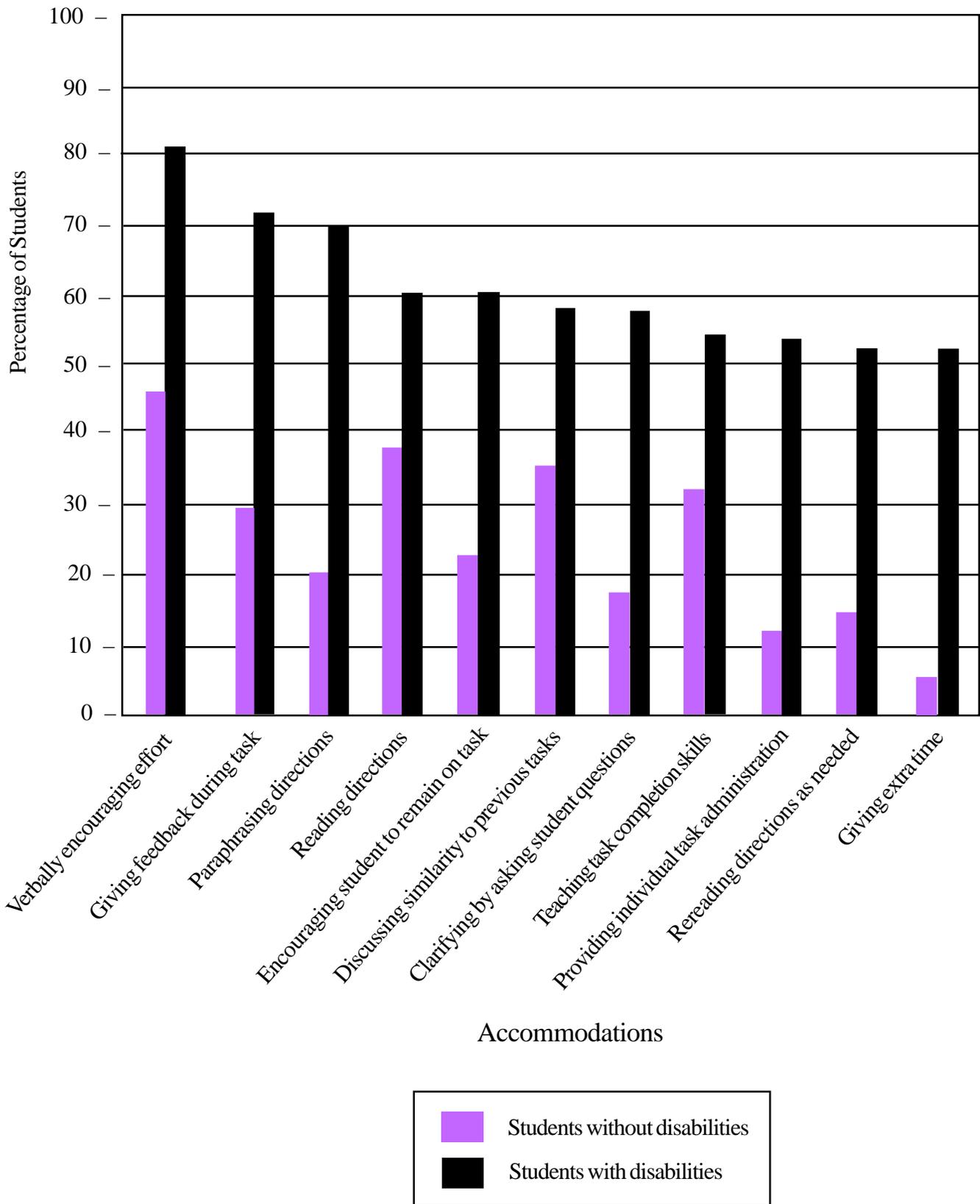
Figure 5. The 10 most common accommodations made for students with and without disabilities.

students to respond in a more sophisticated manner. Students are encouraged to demonstrate their understanding through the construction of knowledge rather than the mere reproduction of facts. Assessments that call for students to respond constructively create opportunities for them to achieve in a manner not captured through a variety of traditional assessment procedures.

These findings suggest that students with disabilities can respond well to more authentic tasks. Although students with disabilities did not score as well on more authentic tasks as their nondisabled peers, the gains for them suggest that such tasks enabled improved demonstrations of learning and a simultaneous improvement in achievement when compared to less authentic tasks. *With more challenging tasks, students with disabilities performed better than students with and without disabilities who received less challenging tasks.* Student achievement generally seems to benefit from the use of more authentic forms of assessment, and the achievement of students with disabilities, who are typically unaccounted for at the secondary level, is no exception.

Although accommodations were used extensively in Data Set 2, they altered the authenticity of only 14% of the 35 tasks. This result demonstrates that teachers are able to adapt assessments for special education students while maintaining the level of intellectual challenge. Significantly, teachers can sustain high expectations of students in inclusive classrooms. At the same time, the result suggests that challenging tasks can be given to mixed groups of students, including students with disabilities, with relatively minor accommodations.

That said, some explanations are needed for the continuing differences between the scores of disabled and nondisabled students, regardless of the level of a task's authenticity. For one, the assessments included here demand a certain level of literacy, in both reading and writing, which may make tasks more difficult for certain students because of their disabilities. A broad definition of *elaborated communication* would allow students to show in-depth understanding through a variety of media, not simply through writing as was required for this study. Alternative student products such as demonstrations or exhibitions may provide a solution for this particular problem but are still atypical in schools. A second explanation arises from the pedagogical context in which the assessments are administered. Although not considered in this study, the curriculum and instruction employed before a given assessment may have an impact on disabled students' ability to respond, given the nature of their disabilities and classroom accommodations. Put simply, the instruction provided to students will affect their ability to access and successfully complete an assessment task.

There is more work to be done with regard to these issues. We are collecting additional assessment data (teacher tasks and student work) from all four high schools participating in the study. We are also visiting the schools to conduct observations of teachers' lessons in the four main subject areas. The lessons are rated according to criteria for authentic instruction. These data will provide further insight into the promise of authentic and inclusive reforms for students with disabilities.

References

Avery, P. (1999). Authentic instruction and assessment. *Social Education, 63*(6), 368-373.

Braden, J. P., Schroeder, J. L., & Buckley, J. A. (2000). *Secondary school reform, inclusion, and authentic assessment* (Brief #3). Madison, WI: Research Institute on Secondary Education Reform for Youth with Disabilities.

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (Eds.). (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Hanley-Maxwell, C., Phelps, L. A., Braden, J., & Warren, V. D. (1999). *Schools of authentic and inclusive learning* (Brief #1). Madison, WI: Research Institute on Secondary Education Reform for Youth with Disabilities.

Lipsky, D. K., & Gartner, A. (1996). Inclusive education and school restructuring. In W. Stainback & S. Stainback (Eds.), *Controversial issues confronting special education* (pp. 3-15). Boston: Allyn & Bacon.

McGee, A. M., Mutch, L. M., & Leyland, A. (1993). Assessing children who cannot be "tested." *Educational Psychology, 13*(1), 43-48.

Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.

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

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*, 280-312.

Newmann, F. M., & Wehlage, G. G. (1995). *Successful school restructuring: A report to the public and educators.* Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Schroeder, J. L. (2000). *Authentic learning and accommodations for students with disabilities and without disabilities in restructuring secondary schools*. Unpublished master's thesis, University of Wisconsin–Madison.

Thousand, J., Rosenberg, R. L., Bishop, K. D., & Villa, R. A. (1997). The evolution of secondary inclusion. *Remedial and Special Education 18*, 270-84.

Trent, S. C., Artiles, A. J., & Englert, C. S. (1998). From deficit thinking to social constructivism: A review of theory, research, and practice in special education. *Review of Research in Education, 23,* 277-307.

Research Institute on Secondary
Education Reform (RISER)
for Youth with Disabilities
1025 W. Johnson St., Suite 461
University of Wisconsin—Madison
Madison, WI 53706

Address Correction Requested