WCER Highlights

WISCONSIN CENTER FOR EDUCATION RESEARCH • SCHOOL OF EDUCATION • UNIVERSITY OF WISCONSIN–MADISON

Some closer looks at testing accommodations

Four recent studies shed some light on the effects of accommodations on students' test scores.

Recent education reform efforts require high standards for all students. The inclusion of students with disabilities in statewide assessment systems is now required by law and considered to represent a key aspect of good testing practices. From social and accountability perspectives, this inclusion of students with disabilities is highly valued, especially when their test scores are known to be valid. Unfortunately, many students with disabilities receive testing accommodations of unknown validity.

Including students with disabilities in assessment is important for two reasons:

1. it is critical to improving the quality of the educational opportunities available to these students, and
2. it provides meaningful and useful information about students' performance to the schools and communities that educate them.

Inclusion raises important questions, however. How appropriate are common performance standards for students with disabilities? What accommodations should be used? What are the effects of testing accommodations on the validity of assessment? How should scores be reported when accommodations have been used? *Continued next page*

WCER Highlights

Summer 2002
Vol. 14, No. 2

4 Better teacher evaluation

6 Comprehensive Centers offer 'silent service'

7 Educational equity for deaf students

FROM THE DIRECTOR

The roles of standards

Standards continue to play an important role in K–12 education. In this issue of *Highlights*, you will read how standards are being used in the education of students with and without disabilities and in teacher evaluation.

In a dissertation study, graduate student Stephanie Cawthon investigated inclusion policies, such as those mandated by the Individuals with Disabilities Education Act (IDEA), that require educators to evaluate deaf students' participation in curricular content and assessment. Cawthon's results suggest that inclusive placement in the context of standards-based reform appears to result in equitable exposure to curricular content—that is, the opportunity to learn.

In a related article, Stephen Elliott, Tom Kratochwill, and their graduate students report on the effects of accommodations on test scores of students with and without disabilities across several studies. Their general conclusion is that the accommodations investigated are not resulting in valid information about student learning.

In another dissertation study, Steven Kimball studied how three school districts applied new standards for teaching—specifically, the model contained in *Enhancing Professional Practice: A Framework for Teaching*. Kimball concludes that new teacher evaluation systems are having positive effects on teaching practice. Nevertheless some problems remain. Principal among them is the need for the evaluator to have subject-matter competence.

For more information, visit the WCER Web site at www.wcer.wisc.edu.

Andy Porter

Accommodations are usually recommended in packages, rather than individually.

Education leaders at state and district levels struggle with these issues as they work to create policies for testing students with disabilities. In the past, many students with disabilities have been excluded from large-scale achievement tests. Reasons for the exclusion of students with disabilities are varied, but the most common are confusion about the use of test accommodations, concern over causing undue stress from testing, and fear that district test scores will go down.

Four recent experimental studies conducted by UW–Madison Education Professors Stephen N. Elliott and Thomas R. Kratochwill and their graduate students, with funding from the U.S. Department of Education's Office of Special Education and Rehabilitative Services, shed some light on current questions about the effects of accommodations on test scores of students with and without disabilities.

Packaged accommodations

Elliott, Kratochwill, and student Brian McKevitt conducted a study (2001) designed to

1. describe the nature of information on testing accommodations listed in students' individualized education programs (IEPs),
2. document the testing accommodations educators actually use when administering performance assessments to students, and
3. examine the effects accommodations have on the test results of students with and without disabilities.

The study found that accommodations were recommended in packages rather than individually – for example, the instructor might offer a combination of verbal encouragement, reading the directions aloud, simplifying some language, and rereading subtask directions. These accommodation packages were found to have moderate to large effects on performance assessment scores for most students with disabilities and for some students without disabilities. This increase in scores for students without disabilities raises questions about the validity of the accommodations. If changes in testing procedure affect students without disabilities in the same direction and degree as they affect students with disabilities, the changes are not truly acting as accommodations.

Constructed-response questions

Graduate student Aleta Gilbertson Schulte conducted a study supervised by Elliott and Kratochwill (2001) to determine whether accommodations on standardized tests affect students with disabilities differently than students without disabilities. The researchers predicted that accommo-

dations would significantly improve the test scores of fourth-grade students with disabilities but not those of students without disabilities.

In fact, both groups of students improved significantly in the accommodated condition as compared with the nonaccommodated condition. Although students with disabilities benefited more than students without disabilities from accommodations on multiple-choice questions, the two groups benefited equally from accommodations on constructed-response questions.

The finding that both groups of students experienced benefits from testing accommodations on constructed-response questions indicates that the changes in test procedure may have affected both construct-relevant and construct-irrelevant variance. The interaction between accommodation group and question type could indicate that constructed-response questions are more difficult for all students and that accommodations remove barriers to these questions that are not present in multiple-choice questions.

These findings reinforce the principle that research on testing accommodations must take an individual perspective, and that all students must take the tests in both accommodated and nonaccommodated conditions if researchers are to determine whether accommodations provide valid information on student learning.

Extended time for tasks

In a dissertation study (2000), Ann Marquart examined the use of an extended-time accommodation on student scores on an eighth-grade mathematics test. Marquart found that giving students extended time to complete the test (40 minutes, rather than 20 minutes) did not have significant results for students with or without disabilities.

However, in a survey Marquart administered to students and their parents and teachers, most students reported feeling more comfortable, more motivated, and less frustrated under the extended-time condition. They thought they performed better, reported the test seemed easier, and preferred taking the test under the extended-time condition. Most teachers, but few parents indicated that a score from an accommodated test is as valid as a score from the same test administered without accommodations. Many parents, but no teachers believed that the score from an accommodated test is less valid than the score from a nonaccommodated test, and some members from both groups were uncertain. Most parents and teachers believed that if accommodations are used during test administration, those accommodations should be reported along with the test results.

Reading test accommodations

Graduate student Brian McKevitt (pictured above) studied the effects of testing accommodations on the scores of eighth-grade students on standardized reading tests, on score validity, and on teacher and student attitudes about testing (2001).

McKevitt found that the accommodations recommended by teachers did not significantly affect the test scores of students with or without disabilities. However, a read-aloud accommodation, when used with accommodations recommended by teachers, did positively and significantly affect test scores for both groups of students.

There was much individual variability in the accommodation effects. The accommodations raised the scores for 50% of all students with disabilities and 38% of all students without disabilities. Although the scores of both groups were higher with the read-aloud than without any accommodation, the read-aloud effect was not significant when the scores of groups receiving the read-aloud accommodation were compared with the scores of groups receiving only the teacher-recommended accommodations.

Elliott notes that educators who have cooperated with him and his students have rarely requested a summary of research on the effective use of testing accommodations. Perhaps they recognize there is little research on this issue, he says, "If research is going to guide practice, researchers and test publishers interested in seeing all students participate meaningfully in assessments will need to help front-line educators more," he says. They need to understand which testing accommodations are most likely to be valid and how they can go about making decisions about the validity of testing accommodations for individual students prior to testing.

For more information, visit the testing accommodations research Web site at http://www.wcer.wisc.edu/testacc/.

Elliott

Kratochwill

Better teacher evaluation practices

Statements of two teachers about expectations and input were typical across the districts. One said, "[The evaluations] are more two-way, I think. In the past, you just went in and sat down and they did the [form]. Now, I'm more of a participator in the evaluation process."

Over the last century, states and school districts have attempted to structure teacher evaluation practices to promote teacher accountability and improvement in practice. However, traditional teacher evaluation systems and repeated attempts at reform appear to have done little to enhance either accountability or practice.

Teacher evaluation systems are sometimes criticized for lacking commitment at the district or school level, using criteria reflecting narrow conceptions of effective teaching, providing inadequate feedback, and operating with perceived subjectivity.

Standards-based teacher evaluation represents one approach to addressing these problems. Standards aim to represent a common conception of instruction. Proponents of new standards-based teacher evaluation suggest that feedback and objectivity can be strengthened when an evaluation system uses comprehensive standards and rubrics, along with multiple data sources and authentic samples of teaching work, to generate discussion of instruction between teachers and evaluators. Accountability for teaching to new standards and instructional improvement may result from such systems.

UW–Madison assistant researcher Steven Kimball studied the design and operation of three district evaluation systems that use new sets of standards for teaching—specifically, the standards proposed by Charlotte Danielson in *Enhancing Professional Practice: A Framework for Teaching* (Alexandria, VA: Association for Supervision and Curriculum Development, 1996). Danielson designed her teaching framework generically so that it could be applied to various subject areas and to teachers with varying levels of experience.

The three school districts Kimball studied had made considerable investments in time and personnel resources to design, field-test, and implement new teacher evaluation systems. The systems aimed to enhance instructional accountability and teachers' professional growth. All three sites chose Danielson's framework for teaching because it

- is comprehensive and applies empirical and theoretical research,
- has multiple levels of performance defined by specific rubrics,
- represents a conception of teaching that can be agreed upon by teachers and evaluators,
- emphasizes teacher reflection and self-directed growth, and
- saves the extensive time and effort that would be needed to research and develop new unique standards and evaluation criteria.

Most of the teachers interviewed by Kimball considered their new evaluation systems substantial improvements over the prior systems. In comparison with the prior systems, the new systems established more structure, provided more opportunities for teacher discussion, and drew on more data sources for evaluation decisions. Teachers and others underscored the need for ongoing evaluator training to promote quality and consistency in evaluations.

Timing, credibility, and utility

On the surface, most teachers interviewed by Kimball appeared to have a positive view of the performance feedback they received. But when their comments were examined more closely, it became clear that they had expressed a variety of perceptions about the feedback timing, credibility, and utility, which influenced their experiences. In particular, the perceived credibility of evaluators' feedback varied, according to the teachers interviewed. Evaluator credibility was influenced by the amount of time evaluators had invested in the evaluation process and the extent to which teachers and evaluators had similar content expertise. Time constraints may have inhibited opportunities for feedback and evaluation dialogue in some cases.

Despite the increased workload associated with the new evaluation system, teachers largely saw the system standards, procedures, and outcomes as fair. Some teachers did question inconsistencies in evaluation data gathering and evaluator decisions. Increased administrator burdens, combined with insufficient training in the new systems, may have contributed to this problem.

The quality of many teachers' performance was confirmed, and they felt encouraged to continue what they were doing. Others changed some of their instructional strategies based on the feedback they received. Still others did not consider the feedback particularly useful but nonetheless altered aspects of their instruction after reflecting on their performance in relation to the evaluation standards. It was rare, however, for teachers to report substantial changes in their instructional practice as a result of their evaluation experiences, and the large majority of teachers did not see the evalua-

Teachers would benefit from having evaluators familiar with their content area.

Kimball

tion process as an incentive to seek out professional development opportunities.

Untenured teachers were more likely than more experienced teachers to report improvements in their instruction based on feedback received during their evaluations. Kimball offers some explanations for this finding: First, untenured teachers were often novices with little classroom teaching experience. Many new teachers struggle with the most basic (yet important) aspects of teaching: lesson organization, record keeping, and classroom interactions. Principals and other evaluators were often able to draw on their prior teaching experience and their work as instructional leaders to provide constructive feedback on these practices that novice teachers could put to immediate use.

Second, evaluators were required to spend more time with untenured teachers, observing more of their classroom sessions and conducting more pre- and postobservation conferences. The additional time evaluators devoted to assessing untenured teachers gave these teachers more opportunity for direct and ongoing feedback.

Based on his study, Kimball concluded that each of the three districts could benefit from addressing three fundamental questions related to the implementation of standards-based teacher evaluation reforms:

- Can other evaluators, besides the school-based administrators, provide effective evaluations?
- How does one strike a balance between making valid decisions about performance, on the one hand, and minimizing evaluation burdens, on the other?
- What is the appropriate amount and nature of evaluator training?

In the districts studied, administrators who traditionally conducted teacher evaluations (primarily principals) were chosen to implement the new evaluation systems. As a result, some teachers were evaluated by administrators with little or no experience in the teachers' content area or grade level. This problem was most common at the high school level, given secondary teachers' greater specialization in content fields.

Evaluators' own content backgrounds affect their ability to make informed inferences about the quality of teachers' instruction on specific content. Evaluators' backgrounds also may limit their ability to provide deep feedback. One solution, leading to deeper analysis and more credible feedback on content-related instruction, would be for evaluators to develop in-depth knowledge about teaching and learning in one subject.

For more information, contact Kimball at skimball@education.wisc.edu, (608) 265–6201.

Comprehensive Centers offer silent service

Cotherman

Fifteen Comprehensive Regional Assistance Centers around the country provide technical assistance, professional development services, educational materials, and distance learning to educators in their regions. Most Comprehensive Centers' customers gave the centers high ratings for the accessibility, quality, and utility of their services.[1] The Midwest regional center, the Comprehensive Center–Region VI (CC-VI) serves six states and is based at WCER. Center Director Audrey Cotherman calls the assistance provided by CC-VI a "silent service," involving practitioners thoroughly and deeply and enabling them to find their own solutions.

Congress created the Comprehensive Regional Assistance Centers in 1995 to provide quality services to schools, districts, state education departments, community members, and Bureau of Indian Affairs schools.

In the last 6 months, WCER's CC-VI staff have served 22,500 educators in North Dakota, Minnesota, Wisconsin, South Dakota, Iowa, and Michigan. CC-VI also coordinates its services with state departments of education to determine statewide needs. Principals and other administrators are involved by CC-VI in developing ways to change school cultures and to support high performance by teachers and students. CC-VI activities are funded by the U.S. Department of Education's Office of Elementary and Secondary Education.

Teachers' classroom practice benefits from research-based training.

Center mission

Before the formation of the 15 comprehensive centers, there were 46 assistance centers for different programs. Congress consolidated these centers into the 15 comprehensive centers so that practitioners could more easily find assistance. The centers were designed to provide a comprehensive approach to school improvement by focusing on whole school reform, and to view the special needs of special populations within the context of high performance for all students.

CC-VI trainers work with educators at their schools over a sustained time to improve the teaching and learning of reading or math; to train educators on how to collect and use data for decision making; and to provide assistance in forming professional learning communities.

CC-VI trainers also publish and distribute research studies; share effective ways to involve parents as teaching partners; guide schools in working toward safe and drug-free environments; and share research on best practices for teaching non-English-speaking students.

Generally, CC-VI's training programs focus on seven research-based areas: reading, math, English as a second language, the role of the principal as instructional leader, student assessment, parental involvement, and smaller learning communities.

CC-VI staff respect and use the knowledge practitioners already have. They design structures for educators to share their successes; they acquaint educators with results of education research; show teachers and principals how to implement the research in the classroom; and provide sustained rather than one-time or sporadic training, materials, and guidance.

Summer institute

Throughout the year, CC-VI staff spend most of their time on the road, visiting educators in their home districts. Each summer, CC-VI offers numerous institutes. This summer, educators participated in the Fifth Annual CGI National Institute on the UW-Madison campus.

CGI (Cognitively Guided Instruction) is a problem-solving mathematics program, developed at WCER, for students in kindergarten through Grade 3. CGI has proven effective for boys and girls of diverse socioeconomic, racial, ethnic, and language backgrounds. Originally developed and tested in Madison and Madison area schools, this

continued on page 8

Educational equity for deaf and hearing students

Students with disabilities, even those in mainstreamed classrooms, have historically received lower quality instruction and have often been excluded from the required curriculum.

Many classrooms now do include students (including deaf students) who traditionally would have been placed in special classes. At the same time as deaf students are increasingly integrated into diverse educational settings, national and state agencies are calling for higher standards of student achievement. Inclusion policies, such as the Individuals with Disabilities Education Act (IDEA), require educators to evaluate deaf students' participation in curricular content and assessment. However, it is not clear whether inclusive placement results in equitable exposure to curricular content—that is, equitable opportunity to learn (OTL).

UW–Madison graduate student and WCER research assistant Stephanie Cawthon recently completed a study of the access hearing-impaired students have to the standards-based reading curriculum. Her study adapted methods pioneered by WCER Director Andrew C. Porter and his colleagues to study the "enacted curriculum" (i.e., what is taught and how). The measure of the enacted curriculum developed by Porter and his colleagues evaluates the relationship—or alignment—between content standards and classroom practices in core content areas. It provides information on content coverage in both time (number of classes) and depth (level of cognitive work asked of students).

Cawthon investigated in particular how the Wisconsin state standards for reading are covered in classrooms with deaf and hearing students. Cawthon's study was unique in a couple of ways: First, by addressing standards for reading in the early elementary grades, it expanded the study of alignment to a new subject area. Second, it focused on a special education population and educational equity across classroom settings.

Cawthon asked teachers to complete a reading curriculum survey describing their instruction during the spring 2001 semester. Teachers were asked to indicate how much time they allocated to specific reading curriculum topics (i.e., plot, character, spelling, application of information from world events) and what types of learning goals they had for their students (e.g., to memorize, understand, apply, and analyze material).

Teachers in deaf-only, mixed, and hearing-only classrooms reported equitable exposure to standards-based curriculum. This is a promising result in the context of inclusive placement and standards-based reform. The IDEA emphasizes the need for a continuum of services based on the needs of deaf students. Cawthon's results indicate there is less need for concern over possible differences in OTL standards in different educational settings for deaf students.

Standards-based analyses may not capture all differences in instructional experiences.

Cawthon

However, notwithstanding Cawthon's finding of equitable OTL, it is still likely that teachers have different reading instruction strategies for different students. In other words, a standards-based analysis may not capture what are real differences in the instructional experiences of different participant groups. One way to graphically depict how instruction time is used is by constructing content maps. These figures are topographical maps that show relative "highs" and "lows" of instructional time, similar to the "mountains" and "valleys" over rough terrain. They demonstrate differences in how teachers implement standards-based curricula, and show levels of alignment with standards. Diversity in instruction from one teacher to the next is not only to be expected, but hoped for, given the range of communicative and educational needs of deaf and hearing students. Thus, differences in *how* teachers implement standards is a rich area for further research.

Cawthon's study has implications for the emerging field of alignment analysis in standards-based reform. Educators assume that higher levels of alignment to standards will result in higher levels of academic achievement on assessments tied to those standards. Yet achieving better results

continued on page 8

WCER HIGHLIGHTS STAFF

Director Andrew Porter
Associate Director Stephen N. Elliott
Editor Paul Baker
Editorial Consultant Cathy Loeb
Production University Communications

WCER Highlights is published by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. WCER is funded through a variety of federal, state, and private sources, including the U.S. Department of Education, the National Science Foundation, and UW–Madison. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the funding agencies. Fourth-class, bulk-rate postage is paid at UW–Madison, Madison, WI. Send changes of address to WCER, 1025 West Johnson Street, Madison, WI 53706 or call (608) 263–4200. Include the address label from this issue.

No copyright is claimed on the contents of *WCER Highlights*. In reproducing articles, please use the following credit: "Reprinted with permission from *WCER Highlights*, published by the Wisconsin Center for Education Research, UW–Madison School of Education." If you reprint, please send a copy to *Highlights*.

WCER Highlights is available on the Web at *http://www.wcer.wisc.edu.*

ISSN 1073–1822
Vol. 14, No.1
Summer 2002

Comprehensive centers

continued from page 6

program has been replicated in Austin and San Antonio, Texas; Los Angeles, California; Dearborn and Lansing, Michigan; Milwaukee, Wisconsin; Prince George's County, Maryland; Bismarck and Fargo, North Dakota; Washington, D.C.; and numerous other sites. This year's institute was specially designed for teams composed of a trainer and two or more primary grade teachers. Over five days, participants learned how critical mathematical ideas develop in children and planned how to refocus instruction to build on children's natural mathematical abilities and to integrate the learning of skills and problem solving.

For more information, visit the CCVI Web site at www.wcer.wisc.edu/CCVI; e-mail CCVI@mail.wcer.wisc.edu; or call 608–263–4220.

[1]To view a copy of the report "Comprehensive Regional Assistance Centers Program: Final Report on the Evaluation" (2000) point your browser to: http://www.ed.gov/offices/OUS/PES/ExecutiveSummary11.1.00.PSAcorrecteddoc.doc

Deaf and hearing

continued from page 7

involves a substantial number of other variables—for example, school resources, class size, teacher characteristics, student characteristics, assessment validity, and professional development. The strength of the relationship between degree of alignment and student achievement thus is modified by these additional factors and needs to be verified by current and future research.

Alignment between components of accountability measures is one proposed method of documenting the success of standards-based reform.

Cawthon says that this area of research must also identify the educational significance of alignment between standards, curriculum, and assessments. In the hearing-only, mixed, and deaf-only classrooms investigated in this study, teachers report their reading instruction to be moderately aligned to the standards. Perhaps moderate alignment to standards produces "good enough" results in terms of student achievement. Studies of simultaneous alignment with standards and related assessments will be needed to clarify the educational significance of these findings.

For more information, contact Cawthon at stephanie_cawthon@hotmail.com.

Wisconsin Center for Education Research
School of Education
University of Wisconsin–Madison
1025 West Johnson Street
Madison, WI 53706

Address service requested

Nonprofit Organization
U.S. POSTAGE
PAID
Madison, Wisconsin