

98-4 Contracting for Statewide Student Achievement Tests, Department of Public Instruction

Summary

Section 118.30, Wis. Stats., requires the Department of Public Instruction to adopt or approve standardized tests for statewide use to measure pupil attainment of knowledge and concepts in the fourth, eighth, and tenth grades. The tests, which are commonly known as Knowledge and Concepts Examinations and are part of the Wisconsin Student Assessment System, measure student achievement in reading, mathematics, science, social studies, and language arts, including writing. School districts are required to administer the tests to all enrolled fourth-, eighth-, and tenth-grade students, except those whose parents or guardians have requested exemptions. Accommodations, which can include exclusion from the tests, may be made for students enrolled in a special education program and for limited-English speaking students.

The Department's Office of Educational Accountability oversees the Knowledge and Concepts Examinations program. However, rather than provide testing services itself, the Department contracts with a vendor to develop, print, distribute, and score the tests and to develop and disseminate individual and summary score reports for individual students, schools, and school districts, as well as statewide results. Since 1992-93, the first school year in which the tests were administered, the Department has contracted with three different vendors to provide testing services. The third of these vendors—CTB/McGraw-Hill—first provided services in the 1996-97 school year at a total cost to the Department of \$1.425 million in general purpose revenue.

Although school districts generally gave high ratings to the contents of TerraNova, the testing instrument used, there was dissatisfaction with administration of the Knowledge and Concepts Examinations program in 1996-97. At the request of the Joint Legislative Audit Committee, we completed a review of administration of the program, including the provision and quality of testing materials and the content and comparability of score reports. We also reviewed the process through which TerraNova was procured, including whether it was appropriate. In completing our evaluation, we examined the Department's files related to the contracts for testing services; interviewed staff in the Department, representatives of CTB/McGraw-Hill, assessment coordinators in several school districts, and state procurement officials; and reviewed relevant state statutes and administrative code regarding the procurement of goods and services.

When the Department entered into the contract with CTB/McGraw-Hill, the company had not yet completed development of assorted testing materials and score reports the Department was purchasing. The vendor took longer than anticipated to complete the development of these materials, which limited the amount of time available to customize them in order to meet Wisconsin's needs. In addition, the Department was not initially clear about some of its expectations, such as demographic information to be collected for the first time in 1996-97 to meet federal requirements under Title I of the Improving America's Schools Act, and its expectation that separate pre-test guides be developed for each of the participating grades. Together, these factors contributed to problems encountered during the testing period, including errors in and delayed distribution of the test guides, booklets, and directions.

From a review of the problems encountered with administration of the tests in 1996-97, it is apparent that they led to a significant amount of inconvenience and frustration for school districts and classroom teachers. Nevertheless, many problems were ultimately of little consequence to the testing program. However, three problems contributed to delays in the test scoring and reporting process:

- because of the delays in receiving testing materials, Milwaukee Public Schools delayed its administration of the test, resulting in the late return of approximately 16,600 tests for scoring, which delayed compilation of statewide test results;

- two questions were missing from the fourth-grade test, which required CTB/McGraw-Hill to modify its software in order to exclude these questions from the scoring process; and
- pages became separated from test booklets, which required CTB/McGraw-Hill to match pages manually with individual student booklets before scoring the tests.

The delays in the reporting of scores were of greater consequence to the school districts than the problems encountered with the administration of the tests because it is through the score reports that the Department and school districts assess the quality of education in each school and school district, identify and correct curricular weaknesses, and provide guidance to individual students. For 1996-97, score reports consisted of scale scores, which reflect the number of correct answers of each student, and national percentile rankings in each subject for each student, as well as summaries of student performance for each school, each district, and the state.

Although the Department and school districts expected the score reports to be available in January and early February, as they had been in prior years, score reports for 1996-97 were distributed over a six-month period, from February 1997 through July 1997. In addition to problems encountered in the administration of the tests, two other factors contributed substantially to the delays in the scoring process. First, as was the case with its development of supplementary testing materials, CTB/McGraw-Hill did not design its standardized reports on schedule. For example, the vendor did not provide copies of the report formats until October 23, 1996, even though the vendor's schedule indicated the Department would receive copies for review by July 15, 1996. Discussions about the customization of score reports did not begin until November 1996, two months before reporting was initially expected to begin.

Second, although the Department had envisioned a variety of adjustments to CTB/McGraw-Hill's standardized reports, it had not been sufficiently clear about its expectations in its contract with the vendor. In particular, the Department's request for proposals (RFP) for testing services, which was included as an attachment to the contract, included only a list of anticipated score reports labeled "example," although sample reports from the prior contractor were made available. The vendor's response, which was also included as an attachment to the contract, referred only to "score reports specified by the [Department]." No further clarification of expectations was included in the contract language. As a result, there were significant differences in the understanding of what changes needed to be made to the vendor's standard reports, which led to a negotiation process about the extent to which the reports would be customized to meet Wisconsin's needs that lasted through March 1997.

In addition to contributing substantially to delays in the availability of score reports, the Department's lack of clarity about its expectations also led to the production of reports that did not fulfill its or school districts' expectations. In particular, the reporting packages did not summarize schools' results by test question in order to allow a school district to assess its strengths and weaknesses on each item being tested, or identify the criteria by which student answers were graded to allow schools to adjust their writing instruction and ensure those components that are evaluated are covered adequately in class. The Department had expected CTB/McGraw-Hill to provide this information because it had been considered part of the assessment process, the prior vendor had produced these reports, and item analyses had been included in the sample list of reports in the RFP. After negotiating with the Department, CTB/McGraw-Hill subsequently provided similar, but less specific, reports for school district use.

Despite the difficulties encountered, the Department took no action to adjust the amount paid to CTB/McGraw-Hill in 1996-97 for the testing services it provided, nor did it invoke the penalties authorized under the terms of the contract. While the Department did request the removal and replacement of the CTB/McGraw-Hill manager with responsibilities for the project, staff in the Department state that contract reductions were not warranted because the vendor provided services beyond those specified in the contract. They also state that invocation of penalties would not have been prudent, given the Department's interest in maintaining an ongoing relationship with the vendor, which is reflected in the fact that the Department agreed to renew the contract with CTB/McGraw-Hill for 1997-98. Excluding \$175,000 for a one-time study, the contract increased by 20 percent from \$1.25 million to \$1.5 million. Department staff believe the contractual increase was necessary to compensate CTB/McGraw-Hill adequately and retain the firm as the Department's testing contractor.

The testing process has improved in the 1997-98 school year. Testing materials were available on time and free of

significant defects, and score reports have been available earlier than in 1996-97. Most school districts received test results for individual students and scores as scheduled in January 1998, but the Department announced a ten-day delay in the availability of score reports for districts, which was originally scheduled to begin February 13, 1998. According to the Department, shipping of these reports was completed February 27, 1998. In addition, statewide results were released in March, with the full statewide report expected to be available in April. These improvements result, in part, from the progress made throughout 1996-97 in further development of CTB/McGraw-Hill's TerraNova testing materials, scoring process, and customization of materials to meet the Department's expectations. In addition, the Department and CTB/McGraw-Hill have taken steps to improve contract management in 1997-98, including agreeing to a single project schedule.

Although some operational difficulties may be inevitable during the first year of a complex contract with a new vendor, the extent of the problems encountered during the 1996-97 school year raises questions about the Department's oversight and management of its contract for testing services. Therefore, we have recommended the Department improve its contract management by establishing performance-based contracts that include clear project goals and deadlines. Payment to the contractor should be contingent upon the timely fulfillment of these contractually established goals in order to ensure the services provided meet the Department's expectations.

While the Department's project management contributed to problems with both test administration and the distribution and content of score reports, problems related to comparability with test results from prior years could not have been avoided, because the complete comparability of test results requires continuity in the tests being used. That is, the selection of a new vendor—and thereby a new test—predetermined that test results would not be comparable with those of previous years. Because problems with continuity of the test results and other problems that were encountered in 1996-97 could have been avoided if the Department had not entered into a new contract for the provision of testing services, questions have been raised about why the change in vendors was made.

The Department must follow the same general procedures as most other state agencies when purchasing goods and services. Under these guidelines, which are determined by the State Bureau of Procurement and detailed in the State Procurement Manual, most purchases of \$10,000 or more—\$25,000 or more for some agencies—require formal procedures for selecting a vendor, including the rebidding of a contract for professional services every three years. The Department's contract for testing services falls under these guidelines and, therefore, needed to be rebid for services beginning in the 1996-97 school year. In reviewing the vendor-selection process, we found the Department complied with state purchasing requirements. The contract was awarded following the RFP process, which allowed the Department the flexibility to specify criteria in addition to price when selecting the vendor. The Department's selection criteria for 1996-97 did not, however, explicitly include test continuity.

If test continuity is not considered in the selection criteria specified in the RFP, the need to rebid the contract every three years may undermine the Department's ability to ensure continuity within the Knowledge and Concepts Examinations program. To address concerns about the frequent change in vendors, the Department has requested and received approval from the Governor for an exemption from the next scheduled procurement process in 1998-99. This exemption will permit successive one-year extensions of the Department's contract with CTB/McGraw-Hill through 2001-02. It should be noted, however, that while it will enable the Department to secure the benefits of program continuity, it creates other risks. For example, it may limit the Department's ability to ensure it is getting the best services possible at the most competitive price.

In addition, even if the Department continues its contract with CTB/McGraw-Hill for several years, the testing program's continuity will not be assured. Three ongoing initiatives may result in significant changes to the program. First, under 1997 Wisconsin Act 27, school districts must administer a high school graduation test beginning in 2000-01 and require passage of the test, or of an alternative assessment, for all high school graduates beginning in 2002-03. Tenth-grade students will no longer be required to be tested as part of the Knowledge and Concepts Examinations program after 2000-01. However, the high school graduation test, which will be designed to qualify individual students for high school graduation, may not also be able to provide assessment data for schools, districts, and the state that are consistent with data provided by the Knowledge and Concepts Examinations.

Second, score reports for 1997-98 include, for the first time, evaluations of students' performance in comparison to

proficiency standards adopted by the Department. These standards define what fourth-, eighth-, and tenth-grade students ought to know and translate these content standards into minimum expected proficiency scores in each of four levels: advanced, proficient, basic, and minimal performance. This method of reporting differs substantially from the method used in the past, which compared students in Wisconsin only with one another and with students nationwide. Two different concerns about the Department's new proficiency categories have been expressed:

- The Department adopted its proficiency standards for the Knowledge and Concepts Examinations in November 1997, before the Governor's Advisory Taskforce on Education and Learning completed its statewide model academic standards. This has led some to question whether the Department's standards are aligned appropriately with the statewide standards. That question is significant because under 1997 Wisconsin Act 27, each district must adopt academic standards by August 1, 1998. If the standards adopted by school districts differ substantially from the Department's proficiency standards, then school districts will teach students based on one set of standards while testing them on another.

This would become even more problematic if the current proposal to end "social promotion"—which is included in 1997 Senate Bill 436 and its companion, 1997 Assembly Bill 768—is enacted. Under this proposal, before students may advance to grades five and nine they would be required to score at the "basic" level of performance on the Knowledge and Concepts Examinations for the fourth and eighth grades or to complete satisfactorily an alternate examination adopted by a school board. In order to enable students to perform at the basic level, school district academic standards must be aligned with those the tests are designed to measure.

- Some school district administrators and others are concerned that the change to proficiency standards will reflect poorly on Wisconsin's schools, because comparisons to proficiency standards can provide a less-favorable assessment than comparisons to test results of other students across the country. Initial results indicate that in some districts, few students are performing at a level the Department deems proficient.

In response to these concerns, 1997 Assembly Bill 642 would authorize a school board to exempt itself from the Knowledge and Concepts Examinations and, therefore, avoid receiving test results. The bill would also allow a school board to adopt a resolution exempting itself from the requirements regarding the administration of a high school graduation test. However, enactment of the bill could jeopardize Wisconsin's receipt of aid payments under Title I of the Improving America's Schools Act, which requires states to make use of high-quality assessments to determine adequate yearly progress in participating schools in at least three levels of proficiency and to identify how well children are learning. These payments are expected to total \$128 million in fiscal year 1997-98.

Third, the Department has decided to move the Knowledge and Concepts Examinations to the spring semester beginning in the 1998-99 school year. The Department changed the test schedule in response to criticism that tests administered at the beginning of the year cannot measure learning that takes place during the year. However, at the school, district, and statewide levels, the change in the test schedule could complicate comparisons of some scores for the first two years of TerraNova testing with scores for subsequent years. It may also limit the usefulness of the scores for individual students, because test results may not be provided until late April or early May.

Although some changes to the pupil assessment program can improve assessment and accountability of Wisconsin's educational system, frequent structural changes can diminish the program's ability to identify long-term trends and contribute to administrative challenges. The Legislature and the Department will need to weigh the benefits of any changes to the Knowledge and Concepts Examinations program that might be considered in the future against the advantages of a stable and consistent program.
