During the past few years, economic concerns have prompted a number of proposals aimed at reforming federal vocational education and employment training programs. Reports about the inadequate skills of high school graduates, the rapidly changing demands of the workplace, and the declining competitiveness of U.S. firms in the international marketplace have all fueled the idea that the organization and structure of employment preparation programs must change. The 104th Congress argued at length about this issue, but Democrats and Republicans were not able to reach a consensus about the shape of future federal vocational education programs.

However, despite the strong differences in their approaches to reform, all sides seem to agree on the need for trustworthy methods for assessing vocational students’ skills. A noteworthy example of this convergence can be seen in the continuing debate between those who recommend a greater focus on broad, industry skills at the secondary level (Boesel and McFarland, 1994) and those who place more emphasis on occupation-specific skills (Bishop, 1995). Both sides agree about the need for a system that assesses skills in reliable and valid ways. Almost all policy makers think it is essential to measure the degree to which participants have mastered the skills upon which training focuses.

Moreover, many vocational educators are advocating the wider use of alternative assessments, such as portfolios, exhibitions, and performance events, for measuring skills of either type. This interest in new measures derives in part from the changes occurring in vocational education. Educators and employers believe that the work
world is changing and vocational education must adapt if it is to serve students well. The changes in the workplace are complex and not completely understood, but most observers believe that future employees will need integrated academic and vocational knowledge, a broad understanding of occupational areas, the ability to interact creatively with their peers, and higher-order cognitive skills that allow them to be flexible, learn rapidly, and adapt to ever-changing circumstances. To the extent this belief is true, vocational training needs to place greater emphasis on integrated learning, critical-thinking skills, and connections between vocational and academic skills, rather than on the mastery of the narrow, occupation-specific skills that characterized vocational education in the past. This new vision may also require broader changes in vocational education, including rethinking the organization, goals, content, and delivery of services, as well as the manner in which students and programs are assessed.

The educational measurement community is engaged in an equally serious rethinking of the structure of assessment (Wolf, 1992; Mehrens, 1992; Wiggins, 1989). Traditional, selected-response methods (multiple choice, matching, true-false) are being criticized for a variety of reasons: they can lead to narrowing of curriculum, test preparation practices may inflate scores in high-stakes situations, there are consistent differences in average performance between racial/ethnic and gender groups, etc. (Koretz, Linn, Dunbar, and Shepard, 1991; Shepard and Dougherty, 1991; Koretz et al., 1993; Shepard, 1991; Smith and Rothenberg, 1991). Many educators advocate the use of alternative approaches, including open-response items, realistic simulations, extended performance events, exhibitions, judged competitions, portfolios, and other forms of elaborated student demonstration.

Educators and researchers are working to find ways to improve the technical quality and feasibility of such performance-based assessments. With regard to the quality dimension, researchers are concerned about the consistency of scoring and of student performance; the fairness of assessments that demand complex, contextually rich responses; and the interpretability of scores. From a practical point of view, educators worry about the complexity and cost of developing and scoring performance assessments, the additional time burdens they impose on students and teachers, and their acceptability to key
stakeholders, including the business community. On the positive side, the distinguishing feature of most alternative assessments is “authenticity,” i.e., students perform an activity or task as it would be done in practice rather than selecting from a fixed set of alternatives. On their face, these activities have greater validity than selected-response tests because success is clearly related to the criterion of interest, be it writing, problem solving, or performing job tasks. On the negative side, a student’s performance on complex tasks is not as consistent from one task to the next as it is with selected-response items, and the scores produced by alternative assessments are not as dependable or interpretable as those produced by traditional tests (Shavelson, Baxter, and Pine, 1992; Shavelson, Gao, and Baxter, 1993; Koretz, Stecher, Klein, and McCaffrey, 1994). These issues are unresolved at present, and there appear to be trade-offs among quality and feasibility considerations.

The uncertainties surrounding vocational education and educational assessment provide the context for our inquiry into vocational education assessment. Many educators believe that assessment can play an important role in systemic educational changes such as those being envisioned for vocational education (Wolf, 1992). The question we are exploring is: What forms of assessment might best meet the needs of vocational education, and how can educators make intelligent choices among the alternatives? Our evaluation of the alternatives paid particular attention to the purposes for which the assessment was to be used, the quality of the information provided, and the practicality or feasibility of the assessment approach.

PURPOSE AND PROCEDURES

Our project, which ran from 1994 to 1996, had a twofold purpose: (1) to provide information about promising alternative assessments that may meet the needs of vocational educators, and (2) to develop materials to help vocational educators make better decisions regarding the use of alternative assessments. This report addresses the first goal, i.e., it evaluates alternative assessments in the context of the current needs of vocational educators. To that end, we gathered information about selected assessment systems, summarized it in a set of detailed case descriptions, reviewed it critically from the perspective of vocational education, and identified some of the important
factors that affect the choice of assessments in the vocational context. The second goal was to convert our knowledge into training and evaluation materials useful to vocational educators and to other groups considering assessment reforms in similar contexts. These are to be published as a practitioner guide by the National Center for Research in Vocational Education in 1997.

We began our investigation broadly, reviewing the literature and contacting experts in the field to look for promising examples of operational assessment systems applicable to vocational education. We developed a set of frameworks for organizing our thinking about the needs of vocational educators and for classifying types of assessment, uses of assessment, and dimensions of assessment quality. These ideas are reflected in our discussions about the needs of vocational educators in Chapter Two and the range of alternative assessments in Chapter Three. We then identified a tentative list of exemplary alternative assessments—both within vocational education and in related education and training sectors. For each project, we collected initial descriptive data from the printed record and from telephone interviews. We then compiled these data into a working casebook.

A panel of expert advisors familiar with vocational education and assessment was formed to guide our work from both technical and practitioner perspectives and to select the assessment reform efforts that would be reviewed in depth. The panel met in March of 1995 and offered advice about our selection of cases and plans for organizing information. To achieve our goal of providing vocational educators with relevant information and helpful procedures for selecting (or developing) alternative assessments, the panel members felt it was important to include a diverse set of assessments in our sample. They encouraged us to select assessment cases that differed in terms of their purposes and the uses of the results, the types of knowledge and skills being assessed, the types of assessment strategies being used, and the organization and structure of the assessment system.

With these factors in mind, we selected six cases for in-depth investigation:

- Career-Technical Assessment Program (C-TAP)
- Kentucky Instructional Results Information System (KIRIS)
• Laborers-AGC (Associated General Contractors) environmental training and certification programs

• National Board for Professional Teaching Standards (NBPTS) certification program

• Oklahoma Department of Vocational-Technical Education competency-based testing program

• Vocational/Industrial Clubs of America (VICA) national competition

We developed a common set of questions to guide our examination of the six cases. The questions focused on description, implementation, administration, consequences, feasibility, quality, and applicability to vocational education. We gathered information to address these questions from a variety of sources, including descriptive materials provided by the assessment activities, the research literature, telephone interviews, and one- to two-day site visits. During the site visits (which involved four of the six cases), we interviewed staff and observed various activities.

After the data were collected, we constructed a thorough description of each assessment, including the features we deemed to be most relevant to vocational education. Each member of the research team assumed primary responsibility for one of the assessment activities. This person coordinated the data collection and was responsible for writing up the case summary according to a common format. One person assumed an editorial role and rewrote the case summaries to provide greater consistency of presentation and voice. We were not formally evaluating each of the efforts and did not attempt to reach a conclusive judgment about each one. When questions could not be answered via available sources, we left them unresolved; when there was disagreement among sources or other contradictory information, we reported the differences. Finally, we conducted an impressionistic review of the case reports, looking for insights that would be relevant to vocational educators.

The next section briefly describes the cases studied. Subsequent chapters use these cases to illustrate the range of concerns and choices that confront vocational educators. Specifically, Chapter Two discusses the educational uses of assessment and the specific
needs of vocational educators, Chapter Three describes various types of measures that vocational educators might use, and Chapter Four presents the factors that might affect the choice of assessment, including the quality of the information provided and the feasibility of various options. In each instance, we present information from the cases to illustrate the issues being discussed.

Our sample was both too small and too diverse to permit strong generalizations about the type of assessment to use in a particular situation. Instead, it provided illustrations of a variety of trade-offs that confront the developers of educational assessments, trade-offs that are relevant to vocational educators as well. In Chapter Five, we discuss these trade-offs, presenting illustrations drawn from the cases and relating these cases to the vocational education context. In the concluding chapter, we consider two prominent assessment challenges facing vocational educators—improving programs and certifying occupational mastery—and draw some implications from our study for selecting or developing assessments to support those functions.

A note on terminology: The collection of assessment activities we reviewed was quite diverse, creating minor problems in description. The sample ranges from developmental efforts (C-TAP) to fully operational testing programs (Laborers-AGC); from job-specific measures (VICA) to broader, occupational assessments (NBPTS); and from single tests (Oklahoma) to assessment systems (KIRIS). Because of this diversity, it is difficult to find simple terminology to refer to all these assessment efforts. They are not all “tests” in the traditional use of the word, nor are they all “testing programs.” We use the terms assessment, assessment activity, and accountability system to refer to our cases in general. When discussing a specific case, we often use a narrower, more focused term, such as test or measure, as appropriate.

**BRIEF DESCRIPTION OF CASES**

This section briefly summarizes the six assessment activities to familiarize the reader with the range of our sample and the variety of approaches represented. More thorough descriptions of the six assessment activities are contained in Appendices A through F.
Career-Technical Assessment Program

The Career-Technical Assessment Program (C-TAP) is a standards-based assessment system designed to support instruction of important career skills and assess the preparedness of California students for entry-level jobs and postsecondary educational training. The program was developed by WestEd under the direction of the California Department of Education (CDE) and the Sacramento County Office of Education (SCOE). C-TAP assessments are being implemented in five career areas: agriculture, business, health careers, home economics, and industrial and technology education. Originally, in 1990, C-TAP was planned as a set of specific occupational tests for over twenty-nine occupations, and C-TAP’s primary purpose was to be a standardized statewide student certification system. However, C-TAP’s purpose and content have changed over time. The focus switched from specific occupations and job skills to clusters of related occupations and broader skills, and expanded to include cumulative assessment components. The C-TAP assessment system includes three components, each of which addresses academic skills, general workplace skills, and job-specific skills: (1) the portfolio, (2) the project, and (3) the written scenario. Scoring for the portfolios, projects, and written scenarios is done by teachers, using scoring rubrics designed by WestEd. C-TAP was recently expanded to include multiple choice and short written-response questions. These on-demand components are designed to measure the breadth of students’ career-technical knowledge. They will be administered and scored by CDE. C-TAP is currently being used primarily as a teaching/learning tool and as an assessment contributing to grades in vocational education programs. It is not yet widely used in any standardized fashion; rather, teachers tend to adapt the materials for use in their own classrooms. The portfolio is generally deemed to be the strongest component of C-TAP by teachers using the system. Widespread adoption of portfolios by individual teachers appears to be feasible, if teachers can find time to learn how to use them and agree that substantive course material may have to be dropped, at least initially, to free up time for their use.
Kentucky Instructional Results Information System

The Kentucky Instructional Results Information System (KIRIS) is a statewide assessment system for elementary and secondary schools that is part of a major reform of public education in Kentucky. Designed both as an accountability tool and as a lever to promote changes in curriculum and instruction, KIRIS uses multiple measures of achievement, including open-ended written questions, group performance events, and portfolios of students' best work to produce school-level scores. Also factored into a school’s accountability score are two noncognitive measures, attendance and retention. Significant rewards are attached to success (schools can earn thousands of dollars for high performance) and significant penalties to failure (the threat of external intervention for continued failure to improve performance). Partially as a result, KIRIS has affected teacher behavior and brought about changes in schools that are consistent with the larger state reform effort. But stakeholders have raised questions about the quality of the scores and the fairness of the awards. Independent evaluations have identified technical shortcomings that threaten the validity of the awards and of school comparisons (Hambleton et al., 1995). Kentucky educators are working to respond to these concerns and to improve the system.

Laborers-AGC Environmental Training and Certification Programs

The Laborers International Union of North America and the Associated General Contractors of America (AGC) cooperatively fund and manage a program of courses and assessments whose purpose is to train and certify environmental cleanup workers (and construction laborers). The courses, which are taught at affiliated local training schools, must comply with federal government regulations that focus on avoiding potential threats to health and safety. The assessment system includes both performance events with real equipment (which take place multiple times during the course) and criterion-referenced multiple-choice tests (which occur at the end). The assessments are used to certify each individual’s competence, as well as to monitor program success and report program completion information. In the last few years, Laborers-AGC has started to evaluate and strengthen the technical quality of its environmental assess-
ments, though employer evaluation of certified employees is already quite positive. Because the fund is a shared venture between labor and management, employers have immediate input if their needs are not being met. The Laborers-AGC model carries high operational costs because of the depth and breadth of its hands-on activities and the need for extensive space (e.g., to create mock hazard sites), expensive equipment, and supplies actually used on the job.

**National Board for Professional Teaching Standards Certification Program**

The National Board for Professional Teaching Standards (NBPTS) offers voluntary national certification to recognize highly accomplished K–12 teachers. Although NBPTS is the only case we studied that does not focus on student assessment, we included it because of the interesting alternative assessment strategies it employs and the lessons that can be learned about assessment for certification. The board aims “to establish high and rigorous standards for what teachers should know and be able to do, to certify teachers who meet those standards, and to advance other education reforms for the purpose of improving student learning in American schools” (NBPTS, 1989, p. iii). The standards and tasks by which candidates are judged were developed mainly by teachers. To obtain the NBPTS certificate, teachers must prepare an extensive portfolio demonstrating their preparation, classroom work, teaching strategies, and professional activities, and must participate in a day of performance activities at a regional assessment center. Standards committees (mostly teachers) used a multistage process to develop subject-matter standards, and assessment development laboratories created the initial assessments. Assessments are still being developed/tested for many of the categories (combining one of four grade levels with one of fourteen subjects);¹ in 1995–1996, NBPTS certification was available in two fields. Extensive reviews of validity, reliability, and other quality-related factors have, on the whole, produced positive results (Bond et al., 1994). Areas that need improvement include reduction of the costs of test development, administration, and

¹Draft standards have been developed for vocational education teachers, but no work has yet been done to develop these assessments.
scoring; and clarification of directions for some activities. Candidates who complete the process find it extremely rewarding despite the substantial burdens. The board intends for the system to drive preservice and inservice training and even to influence state licensing standards.

Oklahoma Department of Vocational-Technical Education Competency-Based Testing Program

The Oklahoma competency-based testing program encompasses a range of multiple-choice and performance-based assessments for both secondary and postsecondary students. The Oklahoma Department of Vocational-Technical Education (Oklahoma Vo-Tech) developed and oversees these tests, which are used to certify students for employment, to improve instruction and student learning through competency-based curriculum and assessment, and to report program improvement and accountability data at the state level. Students are required to pass two local performance assessments, attain all locally identified competencies, and then pass a written multiple-choice test. The responsibility for establishing competencies, certifying mastery, and conducting performance assessments rests with individual programs, with the associated variation from site to site. The multiple-choice component of the program is administered centrally and standardized across sites. Criterion-referenced multiple-choice tests have been developed for 250 occupational titles categorized into fifty-five program areas. The tests measure occupation-specific knowledge and skills. State staff feel confident about the tests’ content validity based on the strong employer input into the assessment system, but no formal validation research has been done. Schools and occupational programs use the state curriculum guides to different degrees, so instruction and testing are not always closely tied. Oklahoma has a long-standing tradition of centralization and political support for vocational education, which has translated into substantial state funding. Without this acceptance of centralized authority and this level of support, other states may be hard-pressed to follow Oklahoma’s example.
Vocational/Industrial Clubs of America National Competition

Vocational/Industrial Clubs of America (VICA) is a national organization for secondary and postsecondary students in some sixty vocational/technical fields. VICA conducts the Skills USA Championships, a national competition that focuses on performing occupationally specific skills in realistic contexts. Many of the skill areas include a written exam as well. The national competition is the culmination of local, regional, and state contests; winners proceed to the next level. The main purpose of the contests is to document students’ skill mastery, encourage excellence, and improve the workforce. VICA also aims to improve curriculum and instruction. Performance in the contests is judged by experienced industry practitioners using specific task-related criteria. The organization places high priority on fairness and consistency in judging; however, no research has been done on the validity, reliability, or equity of the test content or scoring methods. VICA aims for its contests to be closely tied to instruction in the relevant field, though the closeness of the tie varies across competition fields and across instructors. Industry practitioners develop the performance tasks and the written tests under VICA’s guidance. This extensive industry involvement increases the relevance of the assessments to the workplace. Students and teachers gain a reality-based and up-to-date picture of the performance and skills expected in their industry from the involvement of practitioners. The written tests are primarily multiple choice, but there are a few open-ended items as well. The VICA model would be relatively easy to replicate in schools. The most substantial obstacle would be recruiting experienced and knowledgeable industry people to design and judge the competitions.

ORGANIZATION

The rest of this report is organized as follows. Chapter Two examines the primary purposes served by assessments in education and the two specific conditions that are generating demands for alternative methods of assessment among vocational educators: the changing student population and the rapidly evolving skill mix that must be reflected in vocational programs. Chapter Three describes the range of assessment methods, from common multiple-choice tests to new
constructed-response alternatives, including performance tasks, senior projects, and portfolios. Chapter Four discusses the quality and feasibility of alternative assessments. Chapter Five identifies other issues, such as standardization, consequences, and voluntariness, that are relevant to choosing appropriate assessment strategies, and summarizes the advantages associated with particular choices. Chapter Six presents examples of the kinds of assessment decisions confronting vocational educators and shows how the results of this study can contribute to those decisions. Six appendices are also provided, each describing one of the case studies in detail.