CHAPTER 1

Exploring Classroom Assessment

In this book, we describe a journey of discovery. Assessments of student achievement command in excess of a billion dollars of educational resources in the US each year and are becoming very prominent forces in the setting of educational policy at international, national, state and local levels. Yet we still have so much to learn about whether or how these assessments really influence students' learning. To fill some of the gaps in our assessment wisdom, several years ago, we set out on a journey to explore the nature, quality and influences of one as yet unexplored part of this vast and growing assessment world: those assessments developed and used by teachers in their classroom on a day-to-day basis. This book details what we found. It can be read as if it were a journal in which we describe our travels: the people we met along the way, the assessment practices they use, their feelings about those assessments, how they use their assessments and their assessment-related needs.

We departed on this journey because these are times of profound change in education—times of "school improvement," and the pursuit of educational excellence. Clearly our innovations cannot succeed, we surmised, if assessments of educational outcomes are inadequate. But what assessments need to be of high quality? Is it just the highly visible, politically important standardized tests that must be sound? They certainly seem to command sufficient research and development resources to ensure high quality. But, we asked ourselves, what about the rest of the assessments that shape student learning: teachers' classroom assessments? Do they need to be of high quality too? Are they sound now? They seem to be so

critical to student well-being. Not only are they one of our indicators of educational *outcomes,* but these classroom assessments also are part of the very instructional *treatments* that produce the desired outcomes.

However, our initial forays into the research literature on the nature and quality of classroom assessment (described in detail in Chap. 2) revealed very few meaningful studies, and those that had been completed foreshadowed the possibility of serious quality control problems.

And so, armed with our intense curiosity and a supply of research tools, we launched our adventure into a process of understanding. We began with research methods that relied on interviews and questionnaires. These provided us with a general outline of teachers' assessment practices. But we wanted and needed a much higher resolution picture. So we turned to indepth studies of classroom assessment environments using ethnographic research methods (after we labored to learn about these methods). As these more powerful methods allowed us to move from broad strokes to finer lines, our focus began to improve and details began to emerge.

In time, we came to understand that our adventure and the insights it was producing were allowing us to create an ever sharper and more colorful portrait of what appeared to be a sleeping giant—the truly huge and complex world of classroom assessment. As the picture gained clarity, we were forever transformed both as researchers and professionals in the field of educational measurement.

We were startled by the complexity of what we found and awed by the scope of teachers' need for help. We were encouraged by their desire to improve their assessments. We were both tantalized and frightened by the anticipation of what might result if the giant awoke and demanded to be served. But, from the outset, we were committed to making that happen. This volume represents our attempt to sound the alarm.

THE CRITICAL ROLE OF ASSESSMENT
IN TIMES OF CHANGE

As we enter the decade of the 1990s, reforms are sweeping education as never before. Most of the basic underlying assumptions and key ingredients of the teaching and learning process are being reexamined (Shulman 1986). In effect, we are reconsidering everything from the knowledge and skills that comprise the teaching profession to the very way we organize schools and use instructional time. At the same time, we are conducting an equally intense and every bit as important reexamination of the role and training of school administrators, as we strive to come to terms with the meaning of the concept of instructional leadership.

The one constant during this reevaluation of education is the fact that schools and educators are to be held accountable for students' attainment of educational outcomes. School reforms are to be tested in terms of their impact on students. This fact has stimulated several very exciting developments.

For example, virtually every professional association of teachers in every subject matter area spent considerable time and effort at the national level in the 1980s defining the achievement targets they want students to hit. The clarity and quality of the resulting goal statements is far superior to those developed in the past. Clearer targets are easier for students to hit and easier to assess well.

In addition, concern for accountability has given rise to innovative thinking about alternative forms of assessment. New forms of paper and pencil and performance assessment are emerging. This will provide for the proper translation of a broader array of valued outcomes into sound assessment.

The optimist says assessment will drive instruction in the future and new and better assessments are being developed to do the job. But the cautious optimist says this will only happen if educators at all levels understand the difference between sound and unsound assessment and can integrate sound assessments into the instruction process in effective ways. As our research results will reveal, there is reason to believe that neither teachers nor

administrators—those who make schools happen—can meet these standards.

The challenge we faced in initiating this program of research was to find ways to help them. None of the currently popular school reforms, whether they be "outcome-based education," the "restructuring of schools," "instructional theory into practice (ITIP)," special programs for "at-risk youth," or any other attempt at innovation, can be (or has been) properly evaluated unless and until they have been evaluated in terms of their ability to help more students hit more clearly articulated and carefully assessed achievement targets.

We can only build a foundation of solid understanding of critical assessment issues with and for practitioners if we come to understand the real-world assessment demands they face. We build our emerging structure of school reform and improvement on a foundation of sand if we choose to remain uninformed about the nature, role and quality of classroom assessment.

A ROAD MAP

Our journey into the domain of classroom assessment was long and complex. For this reason, we provide a road map here at the outset to assist the reader. The chapters that follow describe a series of studies in the order in which they were carried out. Each study was influenced by those that preceded it and, in turn, influenced those that followed. The road map provided below outlines each study in terms of the question(s) of interest, why the questions were important, the research methodology employed and a very brief overview of results. The story that unfolds for the reader evolves very much as it did for us as we progressed along the way.

We began our investigations in the early 1980s with a comprehensive analysis of prior research on school testing with the intent of gleaning important insights about assessment in the classroom. In Chapter 2, we detail the results of our initial exploration of the literature. Two bodies of work were tapped, research on the use of tests in schools and research on the teaching and learning process. In the latter case, we concentrated on the emerging re-

search on teacher decision making in the classroom and their use of assessment data in that context.

Our first conclusion was that the literature was shallow. Almost every study of school testing focused on the use of standardized test results in schools. Classroom assessment drew very little attention. However, two key points seemed to emerge. Teachers assess a great deal and student well-being in the classroom hinges on the quality of teachers' assessments. We also found evidence of quality control problems with classroom assessments and clear indications that the educational measurement community is unaware of some fundamental differences between assessment in the context of large-scale testing programs and assessment in the classroom.

This led us to our first exploration of classroom assessment processes, described in Chapter 3. We relied on two forms of teacher self-report data collection methods to help us begin to understand the assessment experience from the teachers' perspective. First, we conducted a national survey of teachers' uses of and concerns about paper and pencil assessment methods, performance assessments and published, standardized tests. Second, we enlisted several teachers to serve as research partners. They kept daily journals in which they described important assessment events in their classrooms.

The results provided our first insights into the student characteristics teachers assess, why they assess, and how they do so. We also were able to detect differences in assessment approaches across teachers teaching at different grade levels and in different subject matter areas. However, paradoxically, we found that individual teachers are quite stable in their assessment methods, even when they assess in different subjects. In addition, our first investigation provided our first clear evidence that teachers are concerned about the quality of their own assessments.

This start told us what teachers *say* they do by way of assessment. We wanted to know more about what they *actually* do. So we moved beyond self-report data to the direct observation of classroom assessment process. In Chapter 4, we describe our first attempt to see assessment happening for ourselves. Three members of the NWREL research staff—all of whom had sound assessment

backgrounds, but none of whom had prior public school teaching experience—became participant observers for ten weeks in three separate and quite different sixth-grade classrooms. Our mission was to describe achievements assessed, purposes for assessment, methods, results, and use of assessment results.

We obtained our first detailed portraits of immense complexity of classroom assessment environments. We surmised that the reason prior assessment researchers had not delved into this arena must have been the fear of trying to come to terms with and make sense of this immense complexity. The similarities and differences we found in our three observational experiences provided many long and enjoyable hours of story telling at weekly research team meetings. We were awed and fascinated.

In an initial attempt to come to terms with the complexity we found, we endeavored to create a written framework to use in profiling the important aspects of what we came to refer to as "classroom assessment environments." Our profiling scheme is presented in detail in Chapter 5. It breaks classroom assessment environments down into fine detail, covering eight major dimensions and dividing those into over 400 specific features. We were determined to come to terms with the complexity. We were certain our framework could be used to analyze the assessment environment in any classroom.

In Chapter 6 we describe our initial attempts to do just that. We took the framework to school and profiled the assessment environments in eight high school classrooms, two each in four subject matter areas: mathematics, science, social studies and language arts. We sought to describe each environment thoroughly and test the sensitivity of the framework to differences across classrooms.

Limitations in space did not permit us to include all of the profiles in this volume. However, in Chapter 6, we present one complete profile and compare it to the profile of another class on another subject taught by the same teacher. The results illustrate the extreme complexity of this aspect of classroom life, but also reveal in sharp detail how that complexity can be described so as to capture both their richness and variation.

And then to fill-in some detail about the remaining six class-

room profiles, we devote Chapter 7 to describing the contrasts we found among the remaining assessment environments. This chapter reveals the sensitivity of the profiling process to differences across classrooms and portrays more clearly than ever before the richness of school testing when considered from the classroom perpective. In effect, in this presentation, we turn the spotlight on aspects of assessment in schools not previously examined and are able to reveal both strengths and weaknesses in the assessments of these teachers.

In Chapter 8 we present our first attempts to take parts of the framework and study each with more powerful microscopes. One such attempt relied on classroom observations and assessment document analysis to describe teachers' assessments of student thinking across ten grades and in four different subject matter areas. The other used similiar information-gathering strategies to uncover precise details about the extent to which the report card grading practices of a small number of high school teachers conform to suggested principles of sound grading practice. We use these two studies to illustrate the kind of detail needed to understand both the assessment competencies teachers need and the extent to which those competencies often are missing.

In the final two chapters, we explore the implications of the classroom assessment research we have completed to date. Chapter 9 addresses implication for the professional preparation of teachers. We spell out in detail both what it is that teachers and their supervisors need to know about assessment in order to manange classroom assessment environments effectively, and how little they are taught about this topic in undergraduate, graduate, preservice and inservice training programs. Specific action programs are proposed to change this critical aspect of teacher preparation.

And in Chapter 10 we explore the implications of our work for assessment policy and future research on classroom assessment. Policy issues addressed include standards for teacher and administrator licensing and certification, standards for the evaluation of teacher performance, the development of building and district staffing plans to help teachers meet the challenges of classroom assessment, and the allocation of limited assessment resources to assure the quality of both large-scale standardized and classroom

assessments. In addition, we seek to stimulate additional research on classroom assessment by identifying and discussing important questions about classroom assessments, their quality and use that remain unanswered.

With this map in mind, then, explore with us the meaning and importance of high-quality classroom assessment.