From **Gal, I. & Garfield, J. B. (editors).** **The Assessment Challenge in Statistics Education**. IOS Press, 1997 (on behalf of the ISI). Pages 1-13. ISBN 90 5199 333 1. Copyright holder: International Statistical Institute. Permission granted by ISI to the IASE to make the chapters from this book freely available on the Internet. The References section is a separate chapter and is available on the Internet at http://www.stat.auckland.ac.nz/~iase/publications/assessbkref.

1. Curricular Goals and Assessment Challenges in Statistics Education

Iddo Gal
Joan Garfield

Purpose

This chapter frames the main issues with which this volume deals. The chapter examines common goals for statistics education at both precollege (school) and college levels, describes the resulting challenges for assessment in statistics education, and outlines the main issues addressed by each of the chapters in this volume. Finally, needs for future research and development are discussed.

INTRODUCTION

Statistics has gained recognition as an important component of the precollege mathematics and science curriculum. Establishing a place in the elementary and secondary curriculum led to the production of new instructional materials for elementary and secondary schools (e.g., Landwehr & Watkins, 1987; Travers, Stout, Swift & Sextro, 1985; Friel, Russell & Mokros, 1990; Konold, 1990; COMAP, 1990; Ohio Math Project, 1992). At the college level, where statistics courses have traditionally been taught, changes in content and pedagogy are being recommended as part of a "statistics reform" effort (e.g., Cobb, 1992; Moore, in press).

Now that attention to the teaching of statistics has become visible at all educational levels, it has become apparent that assessment of student learning and understanding of statistics is not being adequately addressed in current projects and instructional efforts. Until recently (see Garfield, 1994; Konold, 1995; Gal & Ginsburg, 1994) very few publications addressed assessment issues in statistics education. This lack of information is alarming in light of extensive research showing that statistics and probability concepts are difficult to teach and often poorly understood (Garfield & Ahlgren, 1988; Shaughnessy, 1992).

Our goal in creating this book was to provide a useful resource to educators and researchers interested in helping students at all educational levels to develop statistical knowledge and reasoning skills. We advise readers focused on students at one level (e.g., secondary) to not skip over chapters describing students at other levels. We are convinced that students who are introduced to statistical ideas and procedures learn much the same material and concepts (e.g., creating graphical displays of data, describing the center and dispersion of data, inference from data, etc.) regardless of their grade level. Therefore, we believe that discussions of assessment issues couched in the reality of one age group will be of interest to those interested in instruction with other types of students.

This introductory chapter aims to establish a common base for the remaining chapters of this volume, and is organized in three parts. First, eight instructional goals in statistics education that are common to all or most levels of instruction are discussed and some tensions inherent in attempting to reach these goals are explored. Next, several challenges in assessment of students' progress towards these goals are noted and the unique issues addressed by each chapter are outlined. Finally, implications for needed practices, development, and research are outlined.

INSTRUCTIONAL GOALS

In a book discussing assessment of students' knowledge and understanding of any subject, it is desirable to start by presenting a coherent framework of the target concepts, knowledge, or understanding that are to be learned, and therefore, to be assessed. This section aims to identify common goals that apply across diverse contexts of instruction and educational levels and that are presupposed by many of the chapters in this book.

At the precollege level, statistics may be taught to elementary or secondary students, usually as part of the mathematics curriculum. It may be taught as a course, a single unit, as part of a classroom project, or embedded in a mathematics topic such as functions or graphing. The NCTM *Curriculum and Evaluation Standards* (1989), and Project 2061's *Benchmarks for Science Literacy* (AAAS, 1993), both outline broad learning goals for statistics and probability. These and related sources leave wide latitude for interpretation by individual teachers, curriculum developers, and teacher-training programs regarding specific curricular goals and preferred assessment methods.

As a postsecondary subject, many students encounter statistics in a stand-alone generic course (i.e., the infamous "introductory statistics" service course), but some are introduced to statistics couched within the needs and applications of a specific field, such as psychology, business, economics, or engineering. Statistics also continues to be taught as part of the core curriculum in most mathematics departments. These postsecondary versions of introductory statistics courses differ quite a bit in their goals for students, the level of mathematical background required, the coverage of probability theory, and the use of technology. However, general dissatisfaction with student outcomes across courses led to a task force report, commissioned by the Mathematical Association of America (Cobb, 1992). This report offered recommendations for the "reform" of most introductory statistics courses. The basic message was "more data and concepts, less theory, fewer recipes." Statistics instructors were also urged to include more active learning opportunities and to decrease the amount of lecturing.

We believe that there are some common goals to instruction in these seemingly diverse educational levels and contexts. If we think broadly about what it is that we want our students to learn and be able to do with their knowledge, an overarching goal of statistics education emerges. This goal is that, by the time students finish their encounters with statistics, they become informed citizens who are able to:

- Comprehend and deal with uncertainty, variability, and statistical information in the world around them, and participate effectively in an information-laden society.

- Contribute to or take part in the production, interpretation, and communication of data pertaining to problems they encounter in their professional life.

This is a broadly defined instructional vision, and one whose achievement may extend over several years or levels of schooling or over several statistics courses.

As part of achieving this broad vision, eight interrelated basic subgoals for statistics instruction are described below. These subgoals were gleaned from our own prior work and writings (e.g., Garfield & Ahlgren, 1988; Garfield, 1995a; Gal, 1993; Gal & Baron, 1996) and from relevant literature on mathematics, statistics, and science education (e.g., NCTM, 1989; Moore, 1990; AAAS, 1993; SCANS, 1990; ASA-NCTM/Burrill, 1994; Cobb, 1992). To be sure, these subgoals are described as "basic" since they relate to the primary phase of students' encounter with statistics in most levels of instruction. Additional subgoals (e.g., regarding statistical inference or probability) may be posed for specific student populations or contexts of instruction, and it is not assumed that statistics educators will find these subgoals equally important for all student populations. However, all chapters in this volume address issues in assessment related to at least one of these basic subgoals.

Goal 1: Understand the purpose and logic of statistical investigations

Students should understand why statistical investigations are conducted, and the "big ideas" that underlie approaches to data-based inquiries. These ideas include:

- The existence of variation
- The need to describe populations by collecting data
- The need to reduce raw data by noting trends and main features through summaries and displays of the data
- The need to study samples instead of populations and to infer from samples to populations
- The logic behind related sampling processes
- The notion of error in measurement and inference, and the need to find ways to estimate and control errors
- The need to identify causal processes or factors
- The logic behind methods (such as experiments) for determining causal processes

Goal 2: Understand the process of statistical investigations

Students should understand the nature of and processes involved in a statistical investigation and considerations affecting the design of a plan for data collection. They should recognize how, when, and why existing statistical tools can be used to help an investigative process. They should be familiar with the specific phases of a statistical inquiry. These phases include (not necessarily in a linear order):

- Formulating a question

- Planning the study (e.g., approach and overall design, sampling, choice of measurement tools)
- Collecting and organizing data
- Displaying, exploring, and analyzing data
- Interpreting findings in light of the research questions
- Discussing conclusions and implications from the findings, and identifying issues for further study

Goal 3: Master procedural skills

Students need to master the “component skills” that may be used in the process of a statistical investigation. This mastery includes being able to organize data, compute needed indices (e.g., median, average, confidence interval), or construct and display useful tables, graphs, plots, and charts, either by hand or assisted by technology (e.g., a calculator, graphing calculator, or computer).

Goal 4: Understand mathematical relationships

Students should develop an understanding, intuitive and/or formal, of the main mathematical ideas that underlie statistical displays, procedures, or concepts. They should understand the connection between summary statistics, graphical displays, and the raw data on which they are based. For example, they should be able to explain how the mean is influenced by extreme values in a data set and what happens to the mean and median when data values are changed.

Goal 5: Understand probability and chance

Moore (in press) recommends that students need only an informal grasp of probability in order to follow the reasoning of statistical inference. This understanding would develop from experiences with chance behavior starting with devices (e.g., coins and dice) and leading to computer simulations. In this way students should gain an understanding of a few key ideas (Garfield, 1995b). These ideas include:

- Concepts and words related to chance, uncertainty, and probability appear in our everyday lives, particularly in the media.
- It is important to understand probabilistic processes in order to better understand (the likelihood of) events in the world around us, as well as information in the media.
- Probability is a measure of uncertainty.
- Developing a model and using it to simulate events is a helpful way to generate data to estimate probabilities.
- Sometimes our intuition is incorrect and can lead us to the wrong conclusion regarding probability and chance events.

Goal 6: Develop interpretive skills and statistical literacy

In carrying out a statistical investigation, students need to be able to interpret results and be aware of possible biases or limitations on the generalizations that can be drawn from data. We realize that most students are more likely to be consumers of data than researchers, and will seldom have to collect or analyze data as adults. Instead, they will need to be able to make sense of published results from studies and surveys reported in the media or in a workplace context. Therefore, students need to learn what is involved in interpreting results from a statistical investigation and to pose critical and reflective questions about arguments that refer to summary statistics or to data reported in the media or in project reports from their classroom peers (e.g., How reliable are the measurements used? How representative was the sample? Are the claims being made sensible in light of the data and sample?).

Goal 7: Develop ability to communicate statistically

Strong writing and speaking skills are needed if students are to effectively communicate about statistical investigations and probabilistic phenomena or processes. Good reading comprehension and communication skills are required so that students can effectively discuss or critique statistical or probabilistic arguments they encounter which claim to be based on some data (e.g., “8 out of 10 doctors use...”, “there is a 20 percent chance that...”). Students should be able to use statistical and probabilistic terminology properly, convey results in a convincing way, and be able to construct proper arguments based on data or observations. They should also be able to argue thoughtfully about the validity of other people’s interpretations of data or graphical displays, and raise questions about acceptability of generalizations made on the basis of a single study or a small sample.

Goal 8: Develop useful statistical dispositions

Students should develop an appreciation for the role of chance and randomness in the world and for statistical methods and planned experiments as useful scientific tools and as powerful means for making personal, social, and business-related decisions in the face of uncertainty. They should realize that the process of statistical inquiry can often lead to better conclusions than relying on anecdotal data or on their own subjective experiences or intuitions, but that this is not guaranteed. Students should also learn to adopt a questioning stance when they are faced with an argument that purports to be based on data (e.g., “all people are...”) or a report of results or conclusions from a statistical investigation, survey, or empirical research.

CHALLENGES IN STATISTICS EDUCATION

The unique challenges that statistics teachers face stem from the existence of multiple subgoals as listed above, which require teachers to address a wide range of conceptually distinct issues during instruction. Educators are further challenged by the need to make sure that students understand the real-world problems that motivate statistical work and investigations, and by the need to help students become familiar with the many nuances, considerations, and decisions involved in generating, describing, analyzing, and interpreting data and in reporting findings.

Doing statistics versus being informed consumers of statistics

The eight subgoals listed above are made up of two overlapping but separate clusters. The first cluster includes the first six subgoals that deal mainly with “doing” statistics (understanding of purposes and uses, and of the logic of various procedures). A second, overlapping cluster includes the last three subgoals which are concerned with sense-making and communicative skills as well as with reflection and questioning. As Gal and Brayer-Ebby (in preparation) point out, these different clusters of subgoals “pull” educators in somewhat different directions as each has a different set of instructional implications, and teachers and textbooks are often more concerned with selected aspects of the first cluster than with the second cluster.

Statistics versus mathematics

The eight subgoals listed above emphasize a shift from traditional views of teaching statistics as a mathematical topic (with an emphasis on computations, formulas, and procedures) to the current view that distinguishes between mathematics and statistics as separate disciplines. As Moore (1992) argues, statistics is a mathematical science but is not a branch of mathematics, and has clearly emerged as a discipline in its own right, with characteristic modes of thinking that are more fundamental than either specific methods or mathematical theory.

The following points explicate some of the key differences between the two disciplines:

1. In statistics, the context motivates procedures and is the source of meaning and basis for interpretation of results of such activities. Moore (1990) points out that data should be viewed as numbers with a context.

2. The indeterminacy, “messiness,” or context-boundedness of statistics is markedly different from the more precise, finite nature characterizing traditional learning in other mathematical domains.

3. Mathematical concepts and procedures are used as part of the attempt to manage or “solve” statistical problems, and some technical facility with them may be expected in certain courses and educational levels. However, the need for accurate application of computations or execution of procedures is rapidly being replaced by the need for selective, thoughtful, and accurate use of technological devices and increasingly more sophisticated software programs.

4. The fundamental nature of many (but not all) statistical problems is that they do not have a single mathematical solution. Rather, realistic statistical problems usually start with a question and culminate with the presentation of an opinion that may have different degrees of reasonableness.

5. A primary goal of statistics education is to enable students to be able to render reasoned descriptions, judgments, inferences and opinions about data, or argue about the interpretation of data, using various mathematical tools only to the degree needed. Judgments and inferences expected of students (e.g., predictions about a population based on sample data students collected in a survey they planned and conducted) very often cannot be characterized as “right” or “wrong,” instead having to be evaluated in terms of quality of reasoning, adequacy of methods employed, and nature of data and evidence used, and may often depend on (or be biased by) students’ world knowledge, which might be limited.

The need for alternative approaches to assessment

In light of the complex goals for students and the emerging differences between mathematics and statistics, an adequate assessment of student outcomes is not possible using only multiple choice or short answer questions. These types of questions are all too often divorced from context and focus on accuracy of statistical computations, correct application of formulas, or correctness of graphs and charts, thus assessing only one or two of the subgoals listed earlier. Questions and task formats that culminate in simple "right or wrong" answers do not reflect the nature of many statistical problems. Such tasks provide only limited information about students' statistical reasoning processes, their ability to construct or interpret statistical arguments, their understanding of the logic behind the use of certain procedures (e.g., sampling, averaging), or their ability to clearly and correctly use statistical or mathematical terminology when discussing their work or reasoning.

A range of different assessment methods is needed to provide broad information about the quality of students' thinking, communication, and reasoning processes (NCTM, 1995). Methods are needed that are appropriate not only for assessing specific skills (such as those under subgoals 3 and 4, which in part can be assessed by traditional item formats), but that also reveal students' understanding of the "big ideas" in statistics and their ability to choose and apply statistical tools appropriately when making sense of realistic data. A student's ability to correctly calculate the average of given data either manually or with the help of an electronic aid, for example, says little about her understanding of when the average is a reasonable way to summarize information, or what other statistical tools may be better suited for the task of describing the data (Gal, 1995).

The challenge faced by all educators involved in statistics education is to identify assessment methods that are able to elicit and reveal student learning corresponding to each of the eight subgoals outlined in the previous section. These methods need to gauge the degree of *integration* between students' skills, knowledge and dispositions and their ability to manage meaningful, realistic questions, problems, or situations, both as generators as well as interpreters of data, findings, or statistical messages. This challenge provides the rationale behind the creation of this book.

OVERVIEW OF THIS BOOK

Each chapter in this book discusses assessment issues pertaining to one or more of the eight subgoals, within the context of a specific educational level. However, the issues presented are usually relevant for other levels as well. Whenever possible, authors present examples for assessment tasks suitable for a range of instructional levels and analyze student responses to illustrate interpretive issues. The chapters have been grouped into four parts, each of which is summarized below.

Part I: Curricular Goals and Assessment Frameworks

The chapters in this part outline key curricular goals and desired outcomes in statistics education, describe recent changes and reforms in thinking about the desired processes of instruction in the mathematical sciences, and frame the assessment issues which all educators involved in statistics education have to address.

In Chapter 2, Begg reviews key issues in mathematics education and in education in general that underpin work on assessment in statistics education. Begg discusses the broadening of instruction in mathematics and statistics to include the key processes of problem-solving, reasoning, communicating, making connections, and using tools, introduces the emerging constructivist views on learning, and reiterates purposes and general principles for assessment of relevance to statistics education.

Colvin and Vos provide in Chapter 3 an introduction to the authentic assessment movement, which aims to measure student performance within tasks relevant to situations or problems outside of the school setting. An authentic assessment model for statistics education in primary and secondary schools is described which is also relevant for college level statistics instruction. The questions and stages involved in establishing an authentic assessment system are outlined and some dilemmas inherent in this process are noted. Examples for authentic assessment tasks and for student performance on such tasks are included to illustrate the ideas presented.

Chapter 4, Gal, Ginsburg, and Schau focus on assessment challenges that pertain to the curricular goal of developing productive dispositions and attitudes as part of statistics education. These authors argue that non-cognitive or meta-cognitive factors, such as (negative) attitudes or beliefs towards statistics can impede learning of statistics, or hinder the extent to which students will develop useful statistical intuitions that can be applied outside the classroom. The chapter examines current approaches for assessing students' attitudes and beliefs, and suggests ways to extend and integrate such assessment into ongoing instruction.

Part II: Assessing Conceptual Understanding of Statistical Ideas

Learning the many procedures, techniques, and analytic processes in statistics assumes that students understand the underlying concepts and "big ideas" (e.g., variation, visual representation of data, center, spread). However, understanding such concepts and ideas requires that students undergo a process of conceptual development, and that teachers acknowledge the presence of such a process and become familiar with its stages. The chapters in this part discuss the challenges involved in identifying learners' conceptual understanding and demonstrate assessment approaches that can be useful in this regard.

In Chapter 5, Friel, Bright, Frierson, and Kader outline a framework for thinking about what teachers and students know and should able to do with respect to learning statistics in K-8 schools. They focus on assessment of teachers' and students' understanding of concepts and ideas related to graphical representations, in light of the limited knowledge about the complexities of learning such concepts. The authors grapple with questions about the nature of "good tasks" that may be used to assess graph knowledge, and demonstrate assessment tasks that can advance our understanding of the complexities of assessing students' graph knowledge.

Lesh, Amit, and Schorr describe in Chapter 6 an approach to statistical education which focuses on models, modeling, and "model-eliciting" activities. Characteristics of realistic problems which prompt students to construct conceptual models for statistical reasoning are described. An illustrative project-size "performance assessment" activity is presented in which students work in teams, using appropriate technology-based tools, and in which students are able to simultaneously learn and document (self-assess) what they are learning. Examples for students' ways of thinking are provided, and criteria for designing model-eliciting activities that highlight students' conceptual understanding are offered.

Students' lack of conceptual understanding is the topic of Chapter 7. Kelly, Sloane, and Whittaker contend that students approach statistical terminology as a foreign language, disassociated from students' existing mathematical knowledge. This causes many students to learn statistics in a rote fashion rather than understand the underlying logic, or to adopt a series of statistical routines that are poorly understood. This chapter presents college-level classroom examples illustrating how students rush to apply techniques when given statistical tasks, and examines the implications of students' lack of conceptual understanding for needed assessment practices.

In Chapter 8, Schau and Mattern posit that understanding of conceptual connections (i.e., knowing and understanding the relationships between different statistical concepts and techniques) is necessary for students to be successful in statistical problem solving. They argue that an explicit goal in teaching statistics is to assist students in gaining such understanding, and discuss ways for assessing college students' understanding of the interrelationships among concepts, with a particular focus on the use of concept maps.

Part III: Innovative Models for Classroom Assessment

The chapters in this part illustrate new or improved ways to assess aspects of statistical knowledge in specific contexts of instruction. Several chapters focus on assessment of knowledge elements or skills pertaining to one of the eight common instructional goals listed earlier, while others discuss assessment methods (e.g., using multiple-choice exams or portfolios, assessing projects) which encompass several goals. Regardless of their focus, chapters in this section present diverse and detailed examples for tasks, items, extended problems, or projects that can be used to assess students in a range of instructional environments and levels. Where relevant, chapters also address general theoretical or background issues which guide both the design as well as the interpretation of student responses, and highlight the instructional benefits that can be obtained from using specific techniques or approaches to assessment.

Watson discusses in Chapter 9 issues in assessing students' ability to interpret probabilistic and statistical concepts which appear in the media. The chapter suggests a hierarchy of learning goals for judging outcomes of instruction in this area, provides examples of viable assessment tasks and analysis of student responses based on items from the media, and discusses the implementation of media-based items in ongoing classroom assessment.

In Chapter 10, Curcio and Arzt focus on the problems involved in assessment of statistical knowledge when students work in groups. The authors describe the design and application of an instrument to assess the graph comprehension of middle school students working in a small-group setting. Based on a theoretical problem-solving framework, the assessment instrument is explained and examples for students work are provided to highlight the categories sampled by the instrument. Issues involved in assessing the contribution of group processes to the problem-solving process are also examined.

Starkings offers in Chapter 11 practical advice for teachers who want to use projects in their courses. Project work is a method of allowing students to use what they have learned in statistics lessons in a practical context. This chapter describes the process of designing and completing a project and introduces a method of assessment in stages that gives students an indication of their progress on a project and induces them to continue with the work. Examples of projects are given, two different models for assessment of project work are described, and teachers' experiences with these assessment models are described. The project models and examples described have been extensively used with students of 14 to 18 years of age, but can be adapted for younger or older students as well.

The assessment of project work is also the topic of Chapter 12, by Holmes, though this chapter focuses on assessments done by external examiners, as opposed to assessments conducted by a classroom teacher. In secondary schools in England and Wales, at the end of two years of study of mathematics or statistics, many students undergo a formal assessment of their projects by external examination boards. As a departure from the traditional written examination papers used in the past, this assessment attempts to uncover students' deep understanding of statistics. The assessment methods used to evaluate projects and issues involved in using external examiners are described.

The use of a portfolio assessment in a statistics class is discussed in Chapter 13 by Keeler. The portfolio is a purposeful collection of student work that exhibits the student's efforts, progress and achievements over time. It displays the products of instruction in a way which challenges teachers and students to focus on meaningful outcomes and provides a context for guidance and critique. The use of portfolios is explained and evaluated here in the context of a graduate level statistics course, yet the processes involved in developing portfolios and using them to support learning can easily be generalized to different educational levels. The author addresses how an instructor determines the critical knowledge to be assessed in this format, the learning principles used to guide the assessment process, and the considerations involved in deciding how to implement portfolio assessment and how to rate and score student work.

In Chapter 14, Lajoie describes a way in which computer technology can be used to assess as well as extend statistical learning. The chapter describes a project for eighth grade students in which a computerized library of exemplars (based on audio and video recordings) was created and used to illustrate samples of average and above average performance on statistical projects. A case is built for how technology can be used to provide detailed information, beyond that which is normally provided by traditional paper and pencil tests, on how students progress over time in their statistical problem solving and reasoning.

The first six chapters in this section describe ways that can help capture multiple aspects of students' ability to apply or extend their emerging statistical knowledge and problem-solving skills. However, some of these approaches require that teachers break out of traditional forms of instruction as well. The remaining two chapters take a fresh look at the assessment challenges facing teachers when they have to continue to rely on more traditional forms of paper and pencil tests. Such assessments may be called for to evaluate mastery of computational procedures or of certain statistical techniques, or may be inevitable due to classroom realities related to class size or conditions of instruction.

In Chapter 15, Jolliffe aims to help teachers develop their own assessment instruments for formative purposes through exemplars of written classrooms tasks. Unsatisfactory tasks are used to illustrate the pitfalls, and alternative versions are given as examples of good practice. Some comments on grading are included. The emphasis is on written assessment in the classroom, mainly of pupils aged about 14-19, but much is relevant also to introductory statistics courses at college and university level. Consideration is given to ways of assessing factual knowledge, the ability to use computers, understanding of concepts and application of techniques, and communication skills. The pros and cons of multiple choice and open-ended questions are discussed as are the challenges of oral assessment and assessment of group work.

In the last chapter in this section, (Chapter 16), Wild, Triggs, and Pfannkuch discuss effective uses of multiple-choice items in large introductory statistics college-level classes. These authors acknowledge that many of the methods and ideas presented in this volume about “alternative” assessments, while promising and important, tend to be very demanding in terms of teacher-time and other resources, which are scarce in many academic teaching contexts. It is essential to find ways to employ inexpensive traditional methods more creatively to assess and reinforce attainment of key goals in large classes. Wild et al. discuss the pros and cons of multiple choice items and describe strategies for writing items that address multiple teaching goals, including both the ability to plan and carry out investigations and choose appropriate analyses, as well as interpret what these analyses tell about the world. Conclusions are formed about what can and cannot be accomplished with forced-choice testing.

Part IV: Assessing Understanding of Probability

Over the years, numerous studies have examined understanding of probability, chance, and processes with multiple outcomes using children as well as adults. These studies reveal that many students and adults find such topics difficult to learn and understand in both formal and everyday contexts and that learning and understanding may be influenced by ideas and intuitions developed in early years. Both research and classroom observations often suggest that there may be a gap between the formal computations and probability rules taught and (presumably) learned in many classrooms (e.g., regarding the outcomes of certain chance events, such as flipping coins several times), and the informal understandings and beliefs that students have before instruction and may still possess after instruction.

In this volume, three chapters illustrate selected topics and dilemmas and describe possible approaches to some assessment challenges in different populations and instructional contexts. To be sure, what is considered an assessment challenge is contingent on a teacher’s instructional goals, on the extent to which the teacher finds it important to link the teaching of probability with the teaching of data analysis and any other area of statistics or mathematics (such as combinatorics), and on his or her resulting teaching methods. While some teachers may emphasize the learning of the logic behind and rules for probability computations, others may be more interested in making sure students acquire “correct” intuitions about chance processes, or improve their subjective estimates of probability. Overall, the goals for student learning and the assessment of students’ understanding of probability are complex and involve many different issues whose full coverage requires a separate volume.

In Chapter 17, Metz examines the nature and assessment of two core ideas that underlie students' understanding of probability in a real-world context: randomness and chance variation. Metz argues that these “big ideas” should serve as instructional goals and that they involve both conceptual construction and beliefs about the place of chance in the occurrence of events in the world. The author considers how young students’ ideas or beliefs are acquired within or influenced by the culture of a mathematics classroom, and examines assessment of students’ understanding and application of chance in a classroom context, with attention to cognitive constructions, beliefs, and classroom culture.

The role of combinatorics in teaching and learning probability and variables that influence students’ work and cause errors when solving combinatorial problems are examined in Chapter 18 by Batanero, Godino, and Navarro-Pelayo. Based on a theoretical perspective about desired knowledge of combinatorics in the context of learning statistics, the authors present different tasks for assessing combinatorial reasoning of secondary school students and examine issues in the interpretation of student responses.

Finally, to shed light on yet another and different context for teaching and assessing probabilistic knowledge and reasoning, Cohen and Chechile examine in Chapter 19 issues in assessing students' understanding of probability in a software-assisted environment. These authors describe methods for assessing students' understanding and interpretations of probability and sampling distributions while using instructional software in a college statistics course. Ideas about identifying concepts embodied in such instructional software and issues in designing questions to test concepts are described. The authors illustrate students' interpretations of computer displays of probability distributions and analyze students' errors to highlight the unique issues involved in assessing knowledge developed through computer applications.

IMPLICATIONS FOR FUTURE RESEARCH AND DEVELOPMENT NEEDS

Many of the chapters in this volume struggle with two overarching challenges: how to create tasks which can tap students' status with regard to one or several of the instructional subgoals described earlier, and what kinds of interpretations can validly be applied to students' responses or behavior (e.g., students' knowledge, reasoning processes, communicative skills, and dispositions). While some assessment tasks may be relatively context-free, the majority of chapters assume that assessment (and of course learning) activities should be embedded in contexts that are meaningful to students. Thus, at all levels of instruction teachers need to deal with a third challenge, that of the establishment of meaningful contexts for teaching and assessment.

Creating meaningful contexts is crucial yet not trivial, as problems have to be arranged so that their conditions and parameters are understood and shared by both learners and their teachers. For younger students as well as college-level students, what seems meaningful or realistic to a teacher may not be so to the student. Furthermore, problems and tasks used in assessment should be designed to reveal how students approach, model, and reason statistically about a given situation, as opposed to how they apply routine procedures.

Although the chapters in this book provide a thoughtful discussion of many current assessment issues and offer many practical suggestions for statistics educators and researchers, they only begin to address the full range of existing challenges in assessment in statistics education. As more attention is focused on statistics education and as more is known about the processes involved in acquisition of statistical knowledge, skills, and dispositions by students at different levels, we hope to see additional areas for assessment explored in more depth, such as:

Assessment of students in computer-assisted environments

This involves two separate subtopics: searching for effective ways to assess what students can do and how they reason when they use computers or other technological aids, and the nature of and limitations on inferences that can be drawn from assessments when students learn with computers but are tested without computers, as is common in many classes.

Assessment of "statistical literacy"

This would involve testing the application or transfer of student learning to interpretive or functional tasks such as those encountered in media or outside the classroom. The challenge in assessment of statistical literacy is that it should involve examining not only what students can do and how they think when asked to, for example, reflect on a report in the media, but also their tendency or disposition to do so without being cued.

Assessment of students' understanding of "big ideas"

Throughout their education students encounter important statistical ideas such as variation, error, bias, sampling, or representativeness. These ideas underlie much of students' overall understanding of the uses and limitations of statistics, yet may not have direct mathematical or functional representation. Their meaning may depend on the context in which they are invoked. Tasks are needed that can assess students' understanding of and sensitivity to the prevalence and importance of such "big ideas" in different contexts.

Assessment of students' intuitions and reasoning involving probability concepts and processes

This area has seen much research activity and there are clear indications that many students have misconceptions or intuitive beliefs that are not being changed during instruction. There is a need to transfer and adapt promising assessment methods and instruments used by researchers (mostly involving in-depth clinical interviews with selected students) to formats that are reasonably acceptable and accessible to teachers and that can be used for "routine" classroom use. Some preliminary work has been done to develop paper and pencil instruments to assess statistical reasoning but these are difficult to validate using traditional measurement approaches.

Assessment of outcomes of group work

A common teaching format in statistics, especially at the precollege level, is the use of groups, especially for project work. However, the assessment and grading of the outcomes of students' work when it is done in groups has been described as the most frequent stumbling block for novice statistics teachers who participate in workshops emphasizing active learning and cooperative group activities. Promising approaches for assessing group work are being developed and implemented in mathematics education and should be examined for their relevance to statistics education.

A broader challenge, and one pertaining to all areas of instruction in statistics, has to do with the need to develop models and methods for comparative assessments of student learning. We need reliable, valid, and practical assessment instruments for measuring and evaluating the relative utility of instructional approaches or curricula used in teaching statistics, or for monitoring the overall level of statistical knowledge or statistical attitudes and dispositions in a certain student population. As long as statistics items used in large-scale or standardized assessments remain focused on computations (as opposed to statistical reasoning) and provide little context, the relative effectiveness of statistics courses or units will remain difficult to ascertain. Furthermore, there will be little chance to change practices of teachers who tailor their instruction to the content of standardized tests. The ideas presented in several of the chapters in this volume can be used to advance the area of comparative assessment in statistics but further investment in this area is warranted.

Overall, we believe that the challenges listed above can be adequately addressed only through collaborative efforts involving teachers, statisticians, measurement experts, psychologists, mathematics educators, and technology specialists. The merits of such a collaborative approach are demonstrated by the rich array of perspectives offered in this volume by an interdisciplinary set of contributing authors. It is essential that we continue to jointly pursue improvements in current methods of assessment in order to ascertain that all students can function effectively as citizens and workers in an information-laden, statistically-oriented society.