Marilyn H. Oermann
Kathleen B. Gaberson

Evaluation and Testing in NURSING EDUCATION

FIFTH EDITION

Evaluation and Testing in Nursing Education

Marilyn H. Oermann, PhD, RN, ANEF, FAAN, is the Thelma M. Ingles Professor of Nursing and Director of Evaluation and Educational Research at Duke University School of Nursing, Durham, North Carolina. She is an author or coauthor of 18 books and many articles on evaluation, teaching in nursing, and writing for publication as a nurse educator. She is the editor of *Nurse Educator* and the *Journal of Nursing Care Quality* and past editor of the *Annual Review of Nursing Education.* Dr. Oermann lectures widely on teaching and evaluation in nursing.

Kathleen B. Gaberson, PhD, RN, CNOR, CNE, ANEF, is an owner of and nursing education consultant for OWK Consulting, Pittsburgh, Pennsylvania. She has over 35 years of teaching and administrative experience in graduate and undergraduate nursing programs. She is a coauthor of eight nursing education books and an author or coauthor of numerous articles on nursing education and perioperative nursing topics. Dr. Gaberson presents and consults extensively on nursing curriculum revision, assessment and evaluation, and teaching methods. The former research section editor of the *AORN Journal*, she currently serves on the *Journal* Editorial Board.

Evaluation and Testing in Nursing Education

Fifth Edition

MARILYN H. OERMANN, PhD, RN, ANEF, FAAN

KATHLEEN B. GABERSON, PhD, RN, CNOR, CNE, ANEF

SPRINGER PUBLISHING COMPANY

NEW YORK

The author and the publisher of this Work have made every effort to use sources believed to be reliable to provide information that is accurate and compatible with the standards generally accepted at the time of publication. Because medical science is continually advancing, our knowledge base continues to expand. Therefore, as new information becomes available, changes in procedures become necessary. We recommend that the reader always consult current research and specific institutional policies before performing any clinical procedure. The author and publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance on, the information contained in this book. The publisher has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Printed in the United States of America by Bradford & Bigelow.

Contents

PART IV: CLINICAL EVALUATION

**PART V: ISSUES RELATED TO TESTING AND
               EVALUATION IN NURSING EDUCATION**

APPENDICES

Preface

All teachers at some time or another need to assess learning. The teacher may write test items; prepare tests and analyze their results; develop rating scales and clinical evaluation methods; and plan other strategies for assessing learning in the classroom, clinical practice, online courses, simulation, and other settings. Often teachers are not prepared to carry out these tasks as part of their instructional role. This fifth edition of *Evaluation and Testing in Nursing Education* is a resource for teachers in nursing education programs and health care agencies; a textbook for graduate students preparing for their roles as nurse educators; a guide for nurses in clinical practice who teach others and are responsible for evaluating their learning and performance; and a resource for other health care professionals involved in assessment, measurement, testing, and evaluation. Although the examples of test items and other types of assessment methods provided in this book are nursing-oriented, they are easily adapted to assessment in other health fields.

The purposes of this book are to describe concepts of assessment, testing, and evaluation in nursing education and prepare teachers for carrying these out as part of their roles. The book presents qualities of effective assessment procedures; how to plan for testing, assemble and administer tests, and analyze test results; how to write all types of test items and develop assessment methods; and how to assess higher level cognitive skills and learning. There is a chapter on testing and evaluation in online courses and programs, which is particularly relevant considering the growth of online programs in nursing. The book describes the evaluation of written assignments in nursing, the development of rubrics, clinical evaluation, and methods for evaluating clinical performance. With the growth of simulation in nursing, we added a new chapter on using simulation for assessment and high-stakes evaluation. This edition also examines the social, ethical, and legal issues associated with testing and evaluation in nursing; the fundamentals of grading; and program evaluation. The content is useful for teachers in any setting who are involved in evaluating others, whether they are students, nurses, or other types of health care personnel.

Chapter 1 addresses the purposes of assessment, testing, measurement, and evaluation in nursing education. Differences between formative and summative evaluation and between norm-referenced and criterion-referenced measurements are explored. Because effective assessment requires a clear description of *what* and

how to assess, the chapter describes the use of outcomes for developing test items, provides examples of outcomes at different taxonomic levels, and describes how test items would be developed at each of these levels. Some teachers, however, do not use outcomes as the basis for testing but instead develop test items and other assessment methods from the content of the course. For this reason, Chapter 1 also includes an explanation of how to plan assessment using that process.

In Chapter 2, qualities of effective assessment procedures are discussed. The concept of assessment validity, the role of reliability, and their effects on the interpretive quality of assessment results are described. Tests and other assessment instruments yield scores that teachers use to make inferences about how much learners know or what they can do. Validity is the adequacy and appropriateness of those interpretations about learners' knowledge or ability based on those scores. Current ways of thinking about reliability and its relationship to validity are explained. Also discussed in Chapter 2 are important practical considerations that might affect the choice or development of tests and other instruments.

Chapter 3 describes the steps involved in planning for test construction, enabling the teacher to make good decisions about what and when to test, test length, difficulty of test items, item formats, and scoring procedures. An important focus of the chapter is how to develop a test blueprint and then use it for writing test items; examples are provided to clarify this process for the reader. Broad principles important in developing test items, regardless of the specific type, are described in the chapter.

There are different ways of classifying test items. One way is to group them according to how they are scored—objectively or subjectively. Another way is to group them by the type of response required of the test-taker—selected- or constructed-response—which is how we organized the chapters. Selected-response items require the test-taker to select the correct or best answer from options provided by the teacher. These items include true–false, matching, multiple-choice, and multiple-response. Constructed-response items ask the test-taker to supply an answer rather than choose from options already provided. These items include short answer (fill-in-the-blank) and essay (restricted and extended). Chapters 4 to 6 discuss these test items.

A true–false item consists of a statement that the student judges as true or false. In some forms, students also correct the response or supply a rationale as to why the statement is true or false. True–false items are most effective for recall of facts and specific information but may also be used to test the student's comprehension of the content. Chapter 4 describes how to construct true–false items and different variations, for example, correcting false statements or providing a rationale for the response, which allows the teacher to assess if the learner understands the content. Chapter 4 also explains how to develop matching exercises. These consist of two parallel columns in which students match terms, phrases, sentences, or numbers from one column to the other. Principles for writing each type of item are presented, accompanied by sample items.

In Chapter 5, the focus is on writing multiple-choice and multiple-response items. Multiple-choice items, with one correct answer, are used widely in nursing and other fields. This format of test item includes an incomplete statement or question, followed by a list of options that complete the statement or answer the

question. Multiple-response items are designed similarly, although more than one answer may be correct. Both of these formats of test items may be used for assessing learning at the remembering, understanding, applying, and analyzing levels, making them adaptable for a wide range of content and learning outcomes. There are three parts in a multiple-choice item, each with its own set of principles for development: (a) stem, (b) answer, and (c) distractors. In Chapter 5, we discuss how to write each of these parts and provide many examples. We also describe principles for writing multiple-response items, including the format used on the NCLEX®.

With true–false, matching, multiple-choice, and multiple-response items, the test-taker chooses the correct or best answer from the options provided by the teacher. In contrast, with constructed-response items, the test-taker supplies an answer rather than selecting from the options already provided. These items include short answer and essay questions. Short-answer items can be answered by a word, phrase, or number. One format presents a question that students answer in a few words or phrases. With the other format, completion or fill-in-the-blank, students are given an incomplete sentence that they complete by inserting a word or words in the blank space. On the NCLEX, candidates may be asked to perform a calculation and type in the number or to put a list of responses in proper order. In Chapter 6, we describe how to write different formats of short-answer items. We also explain how to develop and score essay items. With essay items, students construct responses based on their understanding of the content. Essay items provide an opportunity for students to select content to discuss, present ideas in their own words, and develop an original and creative response to a question. We provide an extensive discussion on scoring essay responses.

There is much debate in nursing education about students developing higher level thinking skills and clinical judgment. With higher level thinking, students apply concepts and other forms of knowledge to new situations; use that knowledge to solve patient and other types of problems; and arrive at rational and well-thought-out decisions about actions to take. The main principle in assessing higher level learning is to develop test items and other assessment methods that require students to apply knowledge and skills in a *new* situation; the teacher can then assess whether the students are able to use what they have learned in a different context. Chapter 7 presents strategies for assessing higher levels of learning in nursing. Context-dependent item sets or interpretive exercises are discussed as one format of testing appropriate for assessing higher level cognitive skills. Suggestions for developing these are presented in the chapter, including examples of different items. Other methods for assessing cognitive skills in nursing also are presented in this chapter: cases, case studies, unfolding cases, discussions using higher level questioning, debates, media clips, and short written assignments.

Chapter 8 focuses on developing test items that prepare students for licensure and certification examinations. The chapter begins with an explanation of the NCLEX test plans and their implications for nurse educators. Examples are provided of items written at different cognitive levels, thereby avoiding tests that focus only on recall and memorization of facts. The chapter also describes how to write questions about clinical practice or the nursing process and provides sample stems for use with those items. The types of items presented in the chapter are similar to those found on the NCLEX and many certification tests. When teachers

incorporate these items on tests in nursing courses, students acquire experience with this type of testing as they progress through the program, preparing them for taking licensure and certification examinations as graduates.

Through papers and other written assignments, students develop an understanding of the content they are writing about. Written assignments with feedback from the teacher also help students improve their writing ability, an important outcome in any nursing program from the beginning level through graduate study. Chapter 9 provides guidelines for assessing formal papers and other written assignments in nursing courses. The chapter includes criteria for assessing the quality of papers, an example of a scoring rubric, and suggestions for assessing and grading written assignments.

Chapter 10 explains how to assemble and administer a test. In addition to preparing a test blueprint and skillful construction of test items, the final appearance of the test and the way in which it is administered can affect the validity of its results. In Chapter 10, test design rules are described; suggestions for reproducing the test, maintaining test security, administering it, and preventing cheating are presented in this chapter as well.

Online education in nursing continues to expand at a rapid pace. Chapter 11 discusses assessment of learning in online courses, including testing and evaluating course assignments. The chapter begins with a discussion of online testing. To deter cheating and promote academic integrity, faculty members can use a variety of both low- and high-technology solutions. Providing timely and substantive feedback to students is critical in online courses, and we have included a sample rubric for an online discussion board assignment and evaluation. Clinical evaluation of students in online courses and programs presents challenges to faculty members and program administrators. The chapter includes discussion of methods for evaluating students' clinical performance in an online course. Other sections of this chapter examine assessment of online courses, student evaluation of teaching, and evaluating the quality of online nursing programs.

After administering the test, the teacher needs to score it, interpret the results, and then use the results to make varied decisions. Chapter 12 discusses the processes of obtaining scores and performing test and item analysis. It also suggests ways in which teachers can use posttest discussions to contribute to student learning and seek student feedback that can lead to test-item improvement. The chapter begins with a discussion of scoring tests, including weighting items and correcting for guessing, then proceeds to item analysis. How to calculate the difficulty index and discrimination index and analyze each distractor are described; performing an item analysis by hand is explained with an illustration for teachers who do not have computer software for this purpose. Teachers often debate the merits of adjusting test scores by eliminating items or adding points to compensate for real or perceived deficiencies in test construction or performance. We discuss this in the chapter and provide guidelines for faculty in making these decisions. A section of the chapter also presents suggestions and examples of developing a test-item bank. Many publishers also offer test-item banks that relate to the content contained in their textbooks; we discuss why faculty members need to be cautious about using these items for their own examinations.

Chapter 13 describes the process of clinical evaluation in nursing. It begins with a discussion of the outcomes of clinical practice in nursing programs and then presents essential concepts underlying clinical evaluation. In this chapter, we discuss fairness in evaluation, how to build feedback into the evaluation process, and how to determine *what* to evaluate in clinical courses.

Chapter 14 builds on concepts of clinical evaluation examined in the preceding chapter. Many evaluation methods are available for assessing competencies in clinical practice. We discuss observation and recording observations in notes about performance, checklists, and rating scales; written assignments useful for clinical evaluation such as journals, concept maps, case analyses, and short papers; electronic portfolio assessment and how to set up a portfolio system for clinical evaluation; and other methods such as conferences, group projects, and self-evaluation. The chapter includes a sample form for evaluating student participation in clinical conferences and a rubric for peer evaluation of participation in group projects. Because most nursing education programs use rating scales for clinical evaluation, we have included a few examples in Appendix A for readers to review.

Simulation is used widely for instruction in nursing, and it also can be used for assessment. A simulation can be developed for students to demonstrate procedures and technologies, analyze data, and make decisions. Students can care for the patient individually or as a team. Student performance in these simulations can be assessed to provide feedback or for verifying their competencies. Some simulations incorporate standardized patients, actors who portray the role of a patient with a specific diagnosis or condition. Another method for evaluating skills and clinical competencies of nursing students is Objective Structured Clinical Examination (OSCE). In an OSCE, students rotate through stations where they complete an activity or perform a skill, which then can be evaluated. Chapter 15, a new chapter in this edition, examines these methods for assessing clinical competencies of students.

Chapter 16 explores social, ethical, and legal issues associated with testing and evaluation. Social issues such as test bias, grade inflation, effects of testing on self-esteem, and test anxiety are discussed. Ethical issues include privacy and access to test results. By understanding and applying codes for the responsible and ethical use of tests, teachers can assure the proper use of assessment procedures and the valid interpretation of test results. We include several of these codes in the appendices. We also discuss selected legal issues associated with testing.

In Chapter 17, the discussion focuses on how to interpret the meaning of test scores. Basic statistical concepts are presented and used for criterion- and norm-referenced interpretations of teacher-made and standardized test results.

Grading is the use of symbols, such as the letters A through F or pass–fail, to report student achievement. Grading is for summative purposes, indicating how well the student met the outcomes of the course and clinical practicum. To represent valid judgments about student achievement, grades should be based on sound evaluation practices, reliable test results, and multiple assessment methods. Chapter 18 examines the uses of grades in nursing programs, types of grading systems, how to select a grading framework, and how to calculate grades with each of these frameworks. We also discuss grading clinical practice, using pass–fail and other systems for grading, and provide guidelines for the teacher to follow when

students are on the verge of failing a clinical practicum. We also discuss learning contracts and provide an example of one.

Program evaluation is the process of judging the worth or value of an educational program. With the demand for high-quality programs, there has been a greater emphasis on systematic and ongoing program evaluation. Thus, Chapter 19 presents an overview of program evaluation models and discusses evaluation of selected program components, including curriculum, outcomes, and teaching. We also discuss the development of a systematic plan for evaluation and include a sample format.

In addition to this book, we have provided an Instructor's Manual that includes a sample course syllabus, chapter-based PowerPoint presentations, and materials for an online course (with chapter summaries, student learning activities, discussion questions, and assessment strategies). To obtain your electronic copy of these materials, faculty should contact Springer Publishing Company at textbook@springerpub.com.

We wish to acknowledge Margaret Zuccarini, our editor at Springer, for her enthusiasm and continued support. We also thank Springer Publishing Company for its support of nursing education and for publishing our books for many years.

Marilyn H. Oermann
Kathleen B. Gaberson

Planning for Testing

It was Wednesday, and Paul Johnson was caught by surprise when he looked at his office calendar and realized that a test for the course he was teaching was only 1 week away, even though he was the person who had scheduled it! Thankful that he was not teaching this course for the first time, he searched his files for the test he had used last year. When he found it, his brief review showed that some of the content was outdated and that the test did not include items on the new content he had added this year. Because of a university policy that requires a minimum of 3 business days for the copy center to reproduce a test, Paul realized that he would have to finish the necessary revisions of the test and submit it for copying no later than Friday. He would be teaching in the clinical area on Thursday and teaching a class on Friday morning, and he was preparing to go out of town to attend a conference on Saturday.

He stayed up late on Wednesday night to revise the test, planning to proofread it on Thursday after he finished his clinical teaching responsibilities. But because of a family emergency, he was not able to proofread the test that night. Trusting that he had not made any serious clerical errors, he sent the test to the copy center before his class on Friday. When he returned to the office after his conference on Tuesday, he discovered that the photocopier in the copy center had been damaged by a lightning strike before his test had been copied, and had not been repaired or replaced. Paul picked up his test at the copy center that afternoon, but couldn't take it anywhere else to be copied that day because of a scheduled committee meeting. To complicate matters, the department secretary had called in sick that day, and Paul couldn't change his child-care arrangements to allow him to stay late at the office to finish copying the test. He came in very early on Wednesday morning to use the department photocopier, and finally finished the job just before the test was scheduled to begin.

With 5 minutes to spare, Paul rushed into the classroom and distributed the still-warm test booklets. As he was congratulating himself for meeting his deadline the first student raised a hand with a question: "On item three, is there a typo?" Then another student said, "I don't think that the correct answer for item six is there." A third student complained, "Item nine is missing; the numbers jump from 8 to 10" and a fourth student stated, "There are two ds for item 10." Paul knew that

45

it was going to be a long morning. But the worst was yet to come. As they were turning in their tests, students complained, "This test didn't cover the material that I thought it would cover," and "We spent a lot of class time analyzing case studies, but we were tested on memorization of facts." Needless to say, Paul did not look forward to the posttest discussion the following week.

Too often, teachers give little thought to the preparation of their tests until the last minute and then rush to get the job done. A test that is produced in this manner often contains items that are poorly chosen, ambiguous, and either too easy or too difficult, as well as grammatical, spelling, and other clerical errors. The solution lies in adequate planning for test construction before the item-writing phase begins, followed by careful critique of the completed test by other teachers. Exhibit 3.1 lists the steps of the test-construction process. This chapter describes the steps involved in planning for test construction; subsequent chapters will focus on the techniques of writing test items of various formats, assembling and administering the test, and analyzing the test results.

PURPOSE AND POPULATION

All decisions involved in planning a test are based on a teacher's knowledge of the purpose of the test and the relevant characteristics of the population of learners

EXHIBIT 3.1

Checklist for Test Construction

- ❐ Define the purpose of the test.
- ❐ Describe the population to be tested.
- ❐ Determine the optimum length of the test.
- ❐ Specify the desired difficulty and discrimination levels of the test items.
- ❐ Determine the scoring procedure or procedures to be used.
- ❐ Select item formats to be used.
- ❐ Construct a test blueprint or table of specifications.
- ❐ Write the test items.
- ❐ Have the test items critiqued.
- ❐ Determine the arrangement of items on the test.
- ❐ Write specific directions for each item format.
- ❐ Write general directions for the test and prepare a cover sheet.
- ❐ Print or type the test.
- ❐ Proofread the test.
- ❐ Reproduce the test.
- ❐ Prepare a scoring key.
- ❐ Prepare students for taking the test.

to be tested. The *purpose* for the test involves why it is to be given, what it is supposed to measure, and how the test scores will be used. For example, if a test is to be used to measure the extent to which students have met learning objectives to determine course grades, its primary purpose is summative. If the teacher expects the course grades to reflect real differences in the amount of knowledge among the students, the test must be sufficiently difficult to produce an acceptable range of scores. On the other hand, if a test is to be used primarily to provide feedback to staff nurses about their knowledge following a continuing education program, the purpose of the test is formative. If the results will not be used to make important personnel decisions, a large range of scores is not necessary, and the test items can be of moderate or low difficulty. In this chapter, the outcomes of learning are referred to as objectives, but as discussed in Chapter 1, many nurse educators refer to these as outcomes.

A teacher's knowledge of the population that will be tested will be useful in selecting the item formats to be used, determining the length of the test and the testing time required, and selecting the appropriate scoring procedures. The term *population* is not used here in its research sense, but rather to indicate the general group of learners who will be tested. The students' reading levels, English-language literacy, visual acuity, health, and previous testing experience are examples of factors that might influence these decisions. For example, if the population to be tested is a group of five patients who have completed preoperative instruction for coronary bypass graft surgery, the teacher would probably not administer a test of 100 multiple-choice and matching items with a machine-scored answer sheet. However, this type of test might be most appropriate as a final course examination for a class of 75 senior nursing students.

TEST LENGTH

The length of the test is an important factor that is related to its purpose, the abilities of the students, the item formats to be used, the amount of testing time available, and the desired reliability of the test scores. As discussed in Chapter 2, the reliability of test scores generally improves as the length of the assessment increases, so the teacher should attempt to include as many items as possible to adequately sample the content. However, if the purpose of the test is to measure knowledge of a small content domain with a limited number of objectives, fewer items will be needed to achieve an adequate sampling of the content.

It should be noted that assessment length refers to the number of test items or tasks, not to the amount of time it would take the student to complete the test. Items that require the student to analyze a complex data set, draw conclusions, and supply or choose a response take more test administration time; therefore, fewer items of those types can be included on a test to be completed in a fixed time period. When the number of complex assessment tasks to be included on a test is limited by test administration time, it is better to test more frequently than to create longer tests that test less important learning goals (Miller, Linn, & Gronlund, 2013; Waltz, Strickland, & Lenz, 2010).

Because test length probably is limited by the scheduled length of a testing period, it is wise to construct the test so that the majority of the students working at their normal pace will be able to attempt to answer all items. This type of test is called a *power* test. A *speeded* test is one that does not provide sufficient time for all students to respond to all items. Although most standardized tests are speeded, this type of test generally is not appropriate for teacher-made tests in which accuracy rather than speed of response is important (Brookhart & Nitko, 2015; Miller et al., 2013).

DIFFICULTY AND DISCRIMINATION LEVEL

The desired difficulty of a test and its ability to differentiate among various levels of performance are related considerations. Both factors are affected by the purpose of the test and the way in which the scores will be interpreted and used. The difficulty of individual test items affects the average test score; the mean score of a group of students is equal to the sum of the difficulty levels of the test items. The difficulty level of each test item depends on the complexity of the task, the ability of the students who answer it, and the quality of the teaching. It also may be related to the perceived complexity of the item; if students perceive the task as too difficult, they may skip it, resulting in a lower percentage of students who answer the item correctly (Brookhart & Nitko, 2015). See Chapter 12 for a more detailed discussion of item difficulty and discrimination. In general, items on an assessment should have a fairly narrow range of difficulty around the average difficulty level (Waltz et al., 2010), but this rule has different applications depending on how the test results will be interpreted.

If test results are to be used to determine the relative achievement of students (i.e., norm-referenced interpretation), the majority of items on the test should be moderately difficult. The recommended difficulty level for selection-type test items depends on the number of choices allowed. The percentage of students who answer each item correctly should be about midway between 100% and the chance of guessing correctly (e.g., 50% for true–false items, 25% correct for four-alternative multiple-choice items). For example, a moderately difficult true–false item should be answered correctly by 75% to 85% of students (Brookhart & Nitko, 2015; Waltz et al., 2010). When the majority of items on a test are too easy or too difficult, they will not discriminate well between students with varying levels of knowledge or ability.

However, if the teacher wants to make criterion-referenced judgments, more commonly used in nursing education and practice settings, the overall concern is whether a student's performance meets a set standard rather than on the actual score itself. If the purpose of the assessment is to screen out the least capable students (e.g., those failing a course), it should be relatively easy for most test-takers. However, comparing performance to a set standard does not limit assessment to testing of lower level knowledge and ability; considerations of assessment validity should guide the teacher to construct tests that adequately sample the knowledge or performance domain.

When criterion-referenced test results are reported as percentage scores, their variability (range of scores) may be similar to norm-referenced test results, but the interpretation of the range of scores would be more narrow. For example, on a

final exam in a nursing course the potential score range may be 0% to 100%, but the passing score is set at 80%. Even if there is wide variability of scores on the exam, the primary concern is whether the test correctly classifies each student as performing above or below the standard (e.g., 80%). In this case, the teacher should examine the difficulty level of test items and compare them between groups (students who met the standard and students who didn't). If item difficulty levels indicate a relatively easy or relatively difficult exam, criterion-referenced decisions will still be appropriate if the measure consistently classifies students according to the performance standard (Miller et al., 2013; Waltz et al., 2010).

It is important to keep in mind that the difficulty level of test items can only be estimated in advance, depending on the teacher's experience in testing this content and knowledge of the abilities of the students to be tested. When the test has been administered and scored, the actual difficulty index for each item can be compared with the expected difficulty, and items can be revised if the actual difficulty level is much lower or much higher than anticipated (Waltz et al., 2010). Procedures for determining how the test items actually perform are discussed in Chapter 12.

ITEM FORMATS

Some students may be particularly adept at answering essay items; others may prefer multiple-choice items. However, tests should be designed to provide information about students' knowledge or abilities, not about their skill in taking certain types of tests. A test with a variety of item formats provides students with multiple ways to demonstrate their competence (Brookhart & Nitko, 2015). All item formats have their advantages and limitations, which are discussed in later chapters.

Selection Criteria for Item Formats

Teachers should select item formats for their tests based on a variety of factors, such as the learning outcomes to be evaluated, the specific skill to be measured, and the ability level of the students. Some objectives are better measured with certain item formats. For example, if the instructional objective specifies that the student will be able to "discuss the comparative advantages and disadvantages of breast- and bottle-feeding," a multiple-choice item would be inappropriate because it would not allow the teacher to evaluate the student's ability to organize and express ideas on this topic. An essay item would be a better choice for this purpose. Essay items provide opportunities for students to formulate their own responses, drawing on prior learning, and to express their ideas in writing; these often are desired outcomes of nursing education programs.

The teacher's time constraints for constructing the test may affect the choice of item format. In general, essay items take less time to write than multiple-choice items, but they are more difficult and time-consuming to score. A teacher who has little time to prepare a test and therefore chooses an essay format, assuming that this choice is also appropriate for the objectives to be tested, must plan for considerable time after the test is given to score it.

In nursing programs, faculty members often develop multiple-choice items as the predominant, if not exclusive, item format because for a number of years, licensure and certification examinations contained only multiple-choice items. Although this type of test item provides essential practice for students in preparation for taking such high-stakes examinations, it negates the principle of selecting the most appropriate type of test item for the outcome and content to be evaluated. In addition, it limits variety in testing and creativity in evaluating student learning. Although practice with multiple-choice items questions is critical, other types of test items and evaluation strategies also are appropriate for measuring student learning in nursing. In fact, although the majority of NCLEX® examination items currently are four-option multiple-choice items, the item pools now contain other formats such as completion and multiple response (National Council of State Boards of Nursing, 2013). It is clear from this example that nurse educators should not limit their selection of item formats based on the myth that learners must be tested exclusively with the item format most frequently used on a licensure or certification test.

On the other hand, each change of item format on a test requires a change of task for students. Therefore, the number of different item formats to include on a test also depends on the length of the test and the level of the learner. It is generally recommended that teachers use no more than three item formats on a test. Shorter assessments, such as a 10-item quiz, may be limited to a single item format.

Objectively and Subjectively Scored Items

Another powerful and persistent myth is that some item formats evaluate students more objectively than do other formats. Although it is common to describe true–false, matching, and multiple-choice items as "objective," objectivity refers to the way items are scored, not to the type of item or their content (Miller et al., 2013). Objectivity means that once the scoring key is prepared, it is possible for multiple teachers on the same occasion or the same teacher on multiple occasions to arrive at the same score. Subjectively scored items, like essay items (and short-answer items, to a lesser extent), require the judgment of the scorer to determine the degree of correctness and therefore are subject to more variability in scoring.

Selected-Response and Constructed-Response Items

Another way of classifying test items is to identify them by the type of response required of the test-taker (Miller et al., 2013; Waltz et al., 2010). *Selected-response* (or "choice") items require the test-taker to select the correct or best answer from among options provided by the teacher. In this category, the item formats are true–false, matching exercises, and multiple-choice. *Constructed-response* (or "supply") formats require the learner to supply an answer, and may be classified further as limited response (or short response) and extended response. These are the short answer and essay formats. Exhibit 3.2 depicts this schema for classifying test-item formats and the variations of each type.

EXHIBIT 3.2

Classification of Test Items by Type of Response

**SELECTED-RESPONSE
ITEM FORMATS**
("CHOICE" ITEMS)

**CONSTRUCTED-RESPONSE
ITEM FORMATS**
("SUPPLY" ITEMS)

True–false	Short-answer
Matching exercises	Completion or fill-in-the-blank
Multiple-choice	Restricted-response essay
Multiple-response	Extended-response essay

SCORING PROCEDURES

Decisions about what scoring procedure or procedures to use are somewhat dependent on the choice of item formats. Student responses to short-answer, numerical-calculation, and essay items, for instance, usually must be hand-scored, whether they are recorded directly on the test itself, on a separate answer sheet, or in a booklet. Answers to objective test items such as multiple-choice, true–false, and matching also may be recorded on the test itself or on a separate answer sheet. Scannable answer sheets greatly increase the speed of objective scoring procedures and have the additional advantage of allowing computer-generated item analysis reports to be produced. The teacher should decide if the time and resources available for scoring a test suggest that hand scoring or electronic scoring would be preferable. In any case, this decision alone should not influence the choice of test-item format.

TEST BLUEPRINT

Most people would not think of building a house without blueprints. In fact, the word *house* denotes diverse attributes to different individuals. For this reason, a potential homeowner would not purchase a lot, call a builder, and say only, "Build a house for me on my lot." The builder might think that a proper house consists of a two-story brick colonial with four bedrooms, three baths, and a formal dining room, whereas the homeowner had a three-bedroom ranch with two baths, an eat-in kitchen, and a great room with a fireplace in mind. Similarly, the word *test* might mean different things to different teachers; students and their teacher might have widely varied expectations about what the test will contain. The best way to avoid misunderstanding regarding the nature of a test and to ensure that the teacher will be able to make valid judgments about the test scores is to develop a test blueprint, also known as a test plan or a table of specifications, before "building" the test itself.

The elements of a test blueprint include (a) a list of the major topics or instructional objectives that the test will cover, (b) the level of complexity of the task to be assessed, and (c) the emphasis each topic will have, indicated by number or percentage of items or points. Exhibit 3.3 is an example of a test blueprint for a unit test on nursing care during normal pregnancy that illustrates each of these elements.

The row headings along the left margin of the example are the content areas that will be tested. In this case, the content is indicated by a general outline of topics. Teachers may find that a more detailed outline of content or a list of the relevant objectives is more useful for a given purpose and population. Some teachers combine a content outline and the related objectives; in this case, an additional column of objectives would be inserted before or after the content list.

The column headings across the top of the example are taken from the taxonomy of cognitive domain (Anderson & Krathwohl, 2001). Because the test blueprint is a tool to be used by the teacher, it can be modified in any way that makes sense to the user. Accordingly, the teacher who prepared this blueprint

EXHIBIT 3.3

Example of a Test Blueprint for a Unit Test on Normal Pregnancy (75 Points)

CONTENT	LEVEL OF COGNITIVE SKILL[a]				
	R	U	Ap	An	Total #[b]
I. Conception and fetal development		2	3	3	8
II. Maternal physiological changes in pregnancy	2	3	1	2	8
III. Maternal psychological changes in pregnancy		2	2	3	7
IV. Social, cultural, and economic factors affecting pregnancy outcome		3	2	3	8
V. Signs and symptoms of pregnancy	2	2	2		6
VI. Antepartal nursing care		8	10	12	30
VII. Preparation for childbirth		4	1	3	8
TOTAL #[b]	4	24	21	26	75

[a]According to Anderson and Krathwohl (2001) taxonomy of the cognitive domain. Selected levels are included in this test blueprint and are represented by the following key:
R = Remembering
U = Understanding
Ap = Applying
An = Analyzing
[b]Number of points. Test blueprints also may include the number or the percentage of items.

chose to use only selected levels of the taxonomy. Other teachers might include all levels or different levels of the taxonomy, or use a different taxonomy.

The body of the test blueprint is a grid formed by the intersections of content topics and cognitive levels. Each of the cells of the grid has the potential to represent one or more test items that might be developed. The numbers in the cells of the sample test blueprint represent the number of points on the test that will relate to it; some teachers prefer to indicate numbers of items or the percentage of points or items represented by each cell. The percentage is a better indicator of the amount of emphasis to be given to each content area (Miller et al., 2013), but the number of items or points may be more helpful to the teacher in writing actual test items. It is not necessary to write test items for each cell; the teacher's judgment concerning the appropriate emphasis and balance of content governs the decision about which cells should be filled and how many items should be written for each.

Rigorous classification of items into these cells also is unnecessary and, in fact, impossible; the way in which the content is actually taught may affect whether the related test items will be written at the applying or understanding level, for example. For this reason, the actual test may deviate slightly from the specifications for certain cells, but the overall balance of emphasis between the test and the actual instruction should be very similar (Brookhart & Nitko, 2015; Miller et al., 2013).

Once developed, the test blueprint serves several important functions. First, it is a useful tool for guiding the work of the item writer so that sufficient items are developed at the appropriate level to test important content areas and objectives. Without a test blueprint, teachers often use ease of construction as a major consideration in writing test items, resulting in tests with a limited and biased sample of learning tasks that may omit outcomes of greater importance that are more difficult to measure (Miller et al., 2013). Using test blueprints also helps teachers to be accountable for the educational outcomes they produce. The test blueprint can be used as evidence for judging the validity of the resulting test scores. The completed test and blueprint may be reviewed by content experts who can judge whether the test items adequately represent the specified content domain, as described in the procedures for collecting content-related evidence in Chapter 2.

Another important use of the test blueprint is to inform students about the nature of the test and how they should prepare for it. Although the content covered in class and assigned readings should give students a general idea of the content areas to be tested, students often lack a clear sense of the cognitive levels at which they will be tested on this material. Although it might be argued that the objectives might give students a clue as to the level at which they will be tested, teachers often forget that students are not as sophisticated in interpreting objectives as teachers are. Also, some teachers are good at writing objectives that specify a reasonable expectation of performance, but their test items may in fact test higher or lower performance levels. Students need to know the level at which they will be tested because that knowledge will affect how they prepare for the test, not necessarily how much they prepare. They should prepare differently for items that test their ability to apply information than for items that test their ability to synthesize information.

Some teachers worry that if the test blueprint is shared with students, they will not study the content areas that would contribute less to their overall test scores, preferring to concentrate their time and energy on the more important areas of emphasis. If this indeed is the outcome, is it necessarily harmful? Lacking any guidance from the teacher, students may unwisely spend equal amounts of time reviewing all content areas. In fact, professional experience reveals that some knowledge is more important for use in practice than other knowledge. Even if they are good critical thinkers, students may be unable to discriminate more important content from that which is less important because they lack the practice experience to make this distinction. Withholding information about the content emphasis of the test from students might be perceived as an attempt to threaten or punish them for perceived shortcomings such as failure to attend class, failure to read what was assigned, or failure to discern the teacher's priorities. Such a use of testing would be considered unethical.

The best time to share the test blueprint with students is at the beginning of the course or unit of study. If students are unfamiliar with the use of a test blueprint, the teacher may need to explain the concept as well as discuss how it might be useful to the students in planning their preparation for the test. Of course, if the teacher subsequently makes modifications in the blueprint after writing the test items, those changes also should be shared with the students (Brookhart & Nitko, 2015).

WRITING THE TEST ITEMS

After developing the test blueprint, the teacher should begin to write the test items that correspond to each cell. Regardless of the selected item formats, the teacher should consider some general factors that contribute to the quality of the test items.

General Rules for Writing Test Items

1. *Every item should measure something important.* If a test blueprint is designed and used as described in the previous section, each test item will measure an important objective or content area. Without using a blueprint, teachers often write test items that test trivial or obscure knowledge. Sometimes the teacher's intent is to determine whether the students have read assigned materials; however, if the content is not important information, it wastes the teacher's time to write the item and wastes the students' time to read it and respond to it. Similarly, it is not necessary to write "filler" items to meet a targeted number; a test with 98 well-written items that measure important objectives will work as well as or better than one with 98 good items and two meaningless ones. Although the reliability of test results is related to the length of the assessment, this rule presumes that items added to a test to increase the number of tasks would be of the same quality as those that are already a part of the test. Adding items that are so easy that every student will answer the questions correctly, or so difficult that every student will answer them

incorrectly, will not improve the reliability estimate (Miller et al., 2013). In fact, students who know the content well might regard a test item that measures trivial knowledge with annoyance or even suspicion, believing that it is meant to trick them into answering incorrectly. There is no reason other than ease of mentally calculating a percentage score for setting an absolute target number of points on a test at 100.

2. *Every item should have a correct answer*. The correct answer should be one that would be agreed on by experts (Miller et al., 2013). This may seem obvious, but the rule is violated frequently because of the teacher's failure to make a distinction between fact and belief. In some cases, the correct or best answer to a test item might be a matter of opinion, and unless a particular authority is cited in the item, students might justifiably argue a different response than the one the teacher expected. For example, one answer to the question, "When does life begin?" might be "When the kids leave home and the dog dies." If the intent of the question was to measure understanding of when a fetus becomes viable, this is not the correct answer, although if the latter was the teacher's intent, the question was poorly worded. There are a variety of opinions and beliefs about the concept of viability; a better way to word this question is, "According to the standards of the American College of Obstetricians and Gynecologists, at what gestational age does a fetus become viable?" If a test item asks the student to state an opinion about an issue and to support that position with evidence, that is a different matter. That type of item should not be scored as correct or incorrect, but with variable credit based on the completeness of the response, rationale given for the position taken, or the soundness of the student's reasoning (Brookhart & Nitko, 2015).

3. *Use simple, clear, concise, precise, grammatically correct language*. Students who read the test item need to know exactly what task is required of them. Wording a test item clearly is often difficult because of the inherent abstractness and imprecision of language, and it is a challenge to use simple words and sentence structure when writing about highly technical and complex material. The teacher should include enough detail in the test item to communicate the intent of the item but without extraneous words or complex syntax that only serve to increase the reading time. Additionally, grammatical errors may provide unintentional clues to the correct response for the testwise but unprepared student and, at best, annoy the well-prepared student.

This rule is particularly important when testing students for whom English is a second language or non-native speakers (NNSs). Bosher and Bowles (2008) found that in a majority of cases, linguistic modification of test items improved NNSs' comprehension of nursing exam items. The process of linguistic modification or simplification maintains key content area vocabulary but reduces the semantic and syntactic complexity of written English. Linguistic structures such as passive voice constructions, long question phrases, conditional and subordinate clauses, negation, and grammatical errors are particularly difficult for NNSs to understand, and they require more time to read and process (Bosher, 2009; Bosher & Bowles, 2008). Although arguments might be made that no accommodation is made for NNSs on the NCLEX, consideration of measurement validity

must take into account that any test that employs language is at least partially a measure of language skills (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Miller et al., 2013).

The following item stems, adapted from an example given by Bosher and Bowles (2008), illustrate the effect of linguistic simplification:

Original stem: A patient with chronic pain treated over a period of months with an oral form of morphine tells you that she is concerned because she has had to gradually increase the amount of medication she takes to achieve pain control. Your response should include:

Linguistically simplified stem: A patient has chronic pain. She is treated over a period of months with an oral form of morphine. She tells the nurse that she is concerned because she has gradually needed more medication to achieve the same level of pain control. How should the nurse respond? (Bosher & Bowles, 2008, p. 168)

Note that the same content is emphasized, but that the revised example contains four short simple sentences and ends with a question to be answered rather than a completion format. Given growing concerns that even native English speakers are entering postsecondary programs with poorer reading skills, such linguistic modification should benefit all students.

4. *Avoid using jargon, slang, or unnecessary abbreviations.* Health care professionals frequently use jargon, abbreviations, and acronyms in their practice environment; in some ways, it allows them to communicate more quickly, if not more effectively, with others who understand the same language. Informal language in a test item, however, may fail to communicate the intent of the item accurately. Because most students are somewhat anxious when taking tests, they may fail to interpret an abbreviation correctly for the context in which it is used. For example, does MI mean myocardial infarction, mitral insufficiency, or Michigan? Of course, if the intent of the test item is to measure students' ability to define commonly used abbreviations, it would be appropriate to use the abbreviation in the item and ask for the definition, or give the definition and ask the student to supply the abbreviation. Slang almost always conveys the impression that the item-writer does not take the job seriously. As noted previously, slang, jargon, abbreviations, and acronyms contribute to linguistic complexity, especially for NNSs. Additionally, growing alarm about health care errors attributed to poor communication, including the overuse of abbreviations, suggests that nurse educators should set positive examples for their students by using only abbreviations generally approved for use in clinical settings.

5. *Try to use positive wording.* It is difficult to explain this rule without using negative wording, but in general, avoid including words like *no*, *not*, and *except* in the test item. As noted previously, negation contributes to linguistic complexity that interferes with the test performance of NNSs. The use of negative wording is especially confusing in true–false items. If using a negative form is unavoidable, underline the negative word or phrase, or use bold text and all uppercase letters to draw students' attention to it. It is best to avoid asking students to identify the incorrect response, as in the following example:

Which of the following is **NOT** an indication that a skin lesion is a Stage IV pressure ulcer?

a. Blistering[1]
b. Sinus tracts
c. Tissue necrosis
d. Undermining

The structure of this item reinforces the wrong answer and may lead to confusion when a student attempts to recall the correct information at a later time. A better way to word the item is:

Which of the following is an indication that a skin lesion is a Stage II pressure ulcer?

a. Blistering[1]
b. Sinus tracts
c. Tissue necrosis
d. Undermining

6. *No item should contain irrelevant clues to the correct answer.* This is a common error among inexperienced test-item writers. Students who are good test-takers can usually identify such an item and use its flaws to improve their chances of guessing the correct answer when they do not know it. Irrelevant clues include a multiple-choice stem that is grammatically inconsistent with one or more of the options, a word in the stem that is repeated in the correct option, using qualifiers such as "always" or "never" in incorrect responses, placing the correct response in a consistent position among a set of options, or consistently making true statements longer than false statements (Brookhart & Nitko, 2015; Miller et al., 2013). Such items contribute little to the validity of test results because they may not measure what students actually know, but how well they are able to guess the correct answers.

7. *No item should depend on another item for meaning or for the correct answer.* In other words, if a student answers one item incorrectly, he or she will likely answer the related item incorrectly. An example of such a relationship between two completion items follows:

    1. Which insulin should be used for emergency treatment of ketoacidosis?

       _____

    2. What is the onset of action for the insulin in Item 1?

       _____

In this example, Item 2 is dependent on Item 1 for its meaning. Students who supply the wrong answer to Item 1 are unlikely to supply a correct

---

[1]Correct answer.

answer to Item 2. Items should be worded in such a way as to make them independent of each other. However, a series of test items can be developed to relate to a context such as a case study, database, diagram, graph, or other interpretive material. Items that are linked to this material are called interpretive or context-dependent items, and they do not violate this general rule for writing test items because they are linked to a common stimulus, not to each other.

8. *Eliminate extraneous information unless the purpose of the item is to determine whether students can distinguish between relevant and irrelevant data.* Avoid the use of patient names in clinical scenarios; this information adds unnecessarily to reading time, it may distract from the purpose of the item, and it may introduce cultural bias (see Chapter 16). However, some items are designed to measure whether a student can evaluate the relevance of clinical data and use only pertinent information in arriving at the answer. In this case, extraneous data (but not patient names) may be included.

9. *Arrange for a critique of the items.* The best source of this critique is a colleague who teaches the same content area or at least someone who is skilled in the technical aspects of item writing. If no one is available to critique the test items, the teacher who developed them should set them aside for a few days. This will allow the teacher to review the items with a fresh perspective to identify lack of clarity or faulty technical construction.

10. *Prepare more items than the test blueprint specifies.* This will allow for replacement items for those discarded in the review process. The fortunate teacher who does not need to use many replacement items can use the remainder to begin an item bank for future tests.

PREPARING STUDENTS TO TAKE A TEST

A teacher-made test usually measures students' maximum performance rather than their typical performance. For this reason, teachers should create conditions under which students will be able to demonstrate their best possible performance. These conditions include adequate preparation of students to take the test (Brookhart & Nitko, 2015; Miller et al., 2013). Although this is the last point on the test-construction checklist (Exhibit 3.1), the teacher should begin preparing students to take the test at the time the test is scheduled. Adequate preparation includes information, skills, and attitudes that will facilitate students' maximum performance on the test.

Information Needs

Students need information about the test to plan for effective preparation. They need sufficient time to prepare for a test, and the date and time of a test should be announced well in advance. Although many teachers believe that unannounced or "pop" tests motivate students to study more, there is no evidence to support this position. In fact, surprise (unscheduled) tests can be considered punitive or threatening and, as such, represent an unethical use of testing (Brookhart &

Nitko, 2015). Adult learners with multiple responsibilities may need to make adjustments to their work and family responsibilities to have adequate study time, and generous notice of a planned test date will allow them to set their priorities.

In addition, students need to know about the conditions under which they are to be tested, such as how much time will be allotted, whether they will have access to resources such as textbooks, how many items will be included, the types of item formats that will be used, and if they need special tools or supplies to take the test, such as calculators, pencils, or black-ink pens (Miller et al., 2013). They also should know what items and resources they will not be able to use during the test. For example, the teacher may direct students not to bring cell phones, personal digital assistants, chiming watches, watches with calculators, backpacks, briefcases, or any books or papers to the testing site. Some teachers do not allow students to wear caps or hats with brims to discourage cheating. In fact, such requirements may be good practice for prelicensure students who must observe similar restrictions for the NCLEX.

Of course, students also should know what content will be covered on the test, how many items will be devoted to each content area, the cognitive level at which they will be expected to perform, and the types of items to expect. As previously discussed, giving students a copy of the test blueprint and discussing it with them is an effective way for teachers to convey this information. Students should also have sufficient opportunity to practice the type of performance that will be tested. For example, if students will be expected to solve medication dose calculation problems without the use of a calculator, they should practice this type of calculation in class exercises or out-of-class assignments. Students also need to know if spelling, grammar, punctuation, or organization will be considered in scoring open-ended items so that they can prepare accordingly. Finally, teachers should tell students how their test results will be used, including the weight assigned to the test score in grading (Brookhart & Nitko, 2015; Miller et al., 2013).

Another way that teachers can assist students in studying for a test is to have students prepare and use a "cheat sheet." Although this term can be expected to have negative connotations for most teachers, cheat sheets commonly are used in nursing practice in the form of memory aids or triggers such as procedure checklists, pocket guides, and reminder sheets. When legitimized for use in studying and test-taking, cheat sheets capitalize on the belief that although dishonest behavior must be discouraged, the skills associated with cheating can be powerful learning tools.

When students intend to cheat on a test, they usually try to guess potential test items and prepare cheat sheets with the correct answers to those anticipated items. Using this skill for a more honest purpose, the teacher can encourage all of the students to anticipate potential test items. In a test-preparation context, the teacher requires the students to develop a written cheat sheet that summarizes, prioritizes, condenses, and organizes content that they think is important and wish to remember during the test. The teacher may set parameters such as the length of the cheat sheet—for example, one side of one sheet of 8½ × 11-inch paper. The students bring their cheat sheets on the day of the test and may use them during the test; they submit their cheat sheets along with their test papers. Students who do not submit cheat sheets may be penalized by deducting points from their test scores or may not be permitted to take the test at all.

Some students may not even consult their cheat sheets during the test, but they still derive benefit from the preparation that goes into developing them. The teacher also may review the cheat sheets with students whose test scores are low to identify weaknesses in thinking that may have contributed to their errors. When used for this purpose, the cheat sheet becomes a powerful diagnostic and feedback tool.

Test-Taking Skills

Because of an increasingly diverse population of learners in every educational setting, including growing numbers of students for whom English is a second language and whose testing experiences may be different from the teacher's expectations, teachers should determine if their students have adequate test-taking skills for the type of test to be given. If the students lack adequate test-taking skills, their test scores may be lower than their actual abilities. Skill in taking tests sometimes is called *testwiseness*. To be more precise, testwiseness is the ability to use test-taking skills, clues from poorly written test items, and test-taking experience to achieve a test score that is higher than the student's true knowledge would predict. Common errors made by item writers do allow some students to substitute testwiseness for knowledge. But, in general, all students should develop adequate test-taking skills so that they are not at a disadvantage when their scores are compared with those of more testwise individuals. Adequate test-taking skills include the following abilities (Brookhart & Nitko, 2015):

1. Reading and listening to directions and following them accurately
2. Reading test items carefully
3. Recording answers to test items accurately and neatly
4. Avoiding physical and mental fatigue by paced study and adequate rest before the test rather than late-night cram sessions supplemented by stimulants
5. Using test time wisely and working at a pace that allows for careful reflection but also permits responding to all items that the student is likely to answer correctly
6. Bypassing difficult items and returning to them later
7. Making informed guesses rather than omitting answers
8. Outlining and organizing responses to essay items before beginning to write
9. Checking answers to test items for clerical errors and changing answers if a better response is indicated

Many teachers advise students not to change their answers to test items, believing that the first response usually is the correct answer and that changing responses will not increase a student's score. Research findings, however, do not support this position. Studies of answer-changing and its effect on test performance have revealed that most students do change their answers to about 4% of test items and that approximately two thirds of answer changes become correct responses. As item difficulty increases, however, this payoff diminishes; consequently, more knowledgeable students benefit more than less knowledgeable students from changing answers (Brookhart & Nitko, 2015).

Students should be encouraged to change their first response to any item when they have a good reason for making the change. For example, a student who has a clearer understanding of an item after rereading it, who later recalls additional information needed to answer the item, or who receives a clue to the correct answer from another item should not hesitate to change the first answer. Improvement in test scores should not be expected, however, when students change answers without a clear rationale for making the change.

Test Anxiety

Finally, teachers should prepare students to approach a test with helpful attitudes. Although anxiety is a common response to situations in which performance is evaluated, high levels of anxiety are likely to interfere with maximum performance (Miller et al., 2013).

Whether some students can be characterized as test-anxious is a matter of frequent debate. Test anxiety can be viewed in several ways. Students who are motivated to do well often experience increased emotional tension in response to a test. Their perceptions of the testing situation affect their thoughts during test preparation and test-taking. Students who perceive a test as a challenge usually have thoughts that are task-directed. They can focus on completing the task and easily manage any tension that is associated with it. Some students perceive tests as threats because they have poor test-taking skills, inadequate knowledge, or both. These students often have task-irrelevant thoughts about testing. They focus on what could happen if they fail a test, and their feelings of helplessness cause them to desire to escape the situation (Brookhart & Nitko, 2015).

Test anxiety can be characterized as a trait with three components: physical, emotional, and cognitive. Test-anxiety research suggests an interaction among these components: negative thoughts and perceptions about testing can create negative feelings, which interfere with performance (Poorman, Mastorovich, Molcan, & Liberto, 2011). The physical component, or autonomic reactivity, involves unpleasant feelings and reactions such as perspiration, increased heart rate, headaches, and gastrointestinal symptoms, although not all test-anxious individuals have physical reactions. The emotional component involves mood and feelings (e.g., nervousness, uneasiness, fear, dread, panic) associated with testing situations.

The cognitive component refers to thoughts or concerns related to performance and its consequences, occurring before or during a test. Essentially, the cognitive component involves worry about possible negative outcomes: "catastrophic fantasies" about what might happen if the student fails, and "competitive worry" that other students are doing better (Poorman et al., 2011). Cognitive indications of test anxiety include impaired ability to concentrate and easy distractibility during the test, difficulty recalling information ("going blank"), misreading or misunderstanding directions or test items, and feeling pressured to be perfect. Additionally, individuals with true test anxiety often have a history of poor performance on tests and other evaluative situations, particularly high-stakes tests. For example, these individuals may repeatedly fail a driver's license examination or achieve good scores on quizzes or unit tests but fail final examinations (Poorman et al., 2011).

The combination of negative feelings and thoughts often results in behaviors that interfere with students' ability to prepare adequately for a test. One of the most dangerous behaviors is avoidance—procrastinating rather than beginning preparation early, and engaging in activities that seem to be related to preparing for the test but really are just distractions. For example, students often report that they studied for many hours and still failed a test, but a record of their activities would reveal that much of that time was spent highlighting material in the textbook or "preparing to study"—organizing their notes, doing household chores with the intention of minimizing interruptions, and so on. Negative thinking creates anxiety, which students try to avoid by avoiding the studying that they believe is causing the discomfort (Poorman et al., 2011).

Students whose test anxiety interferes with their performance often benefit from treatment that addresses the feeling or emotional component of anxiety and the negative thinking or worry aspect as well as training to improve their general test-taking skills. For example, the test-anxious student may learn techniques for stopping negative thoughts during study periods and testing situations, and behavioral techniques such as progressive relaxation and visual imagery (Poorman et al., 2011). A more comprehensive discussion of the diagnosis and treatment of test anxiety is beyond the scope of this textbook. However, teachers may be able to identify students whose performance suggests that test anxiety may be a factor, and to refer those students for treatment.

Students need to view tests and other assessment procedures as opportunities to demonstrate what they know and what they can do. To foster this attitude, the teacher should express confidence in the students' abilities to prepare for and perform well on an upcoming test. It may be helpful for the teacher to ask the students what would help them to feel more relaxed and less anxious before and during a test. Conducting a review session, giving practice items similar to those that will be used on the test, and not talking or interrupting students during a test are examples of strategies that are likely to reduce students' anxiety to manageable levels (Brookhart & Nitko, 2015; Miller et al., 2013).

SUMMARY

Teachers who leave little time for adequate preparation often produce tests that contain poorly chosen and poorly written test items. Sufficient planning for test construction before the item-writing phase begins, followed by a careful critique of the completed test by other teachers, is likely to produce a test that will yield more valid results.

All decisions involved in planning a test should be based on a teacher's knowledge of the purpose of the test and relevant characteristics of the population of learners to be tested. The purpose for the test involves why it is to be given, what it is supposed to measure, and how the test scores will be used. A teacher's knowledge of the population that will be tested will be useful in selecting the item formats to be used, determining the length of the test and the testing time required, and selecting the appropriate scoring procedures. The students' English-language

literacy, visual acuity, and previous testing experience are examples of factors that might influence these decisions.

The length of the test is an important factor that is related to its purpose, the abilities of the students, the item formats that will be used, the amount of testing time available, and the desired reliability of the test scores. The desired difficulty of the test and its ability to differentiate among various levels of performance are affected by the purpose of the test and the way in which the scores will be interpreted and used.

A test with a variety of item formats usually provides students with more opportunity to demonstrate their competence than a test with only one item format. Test items may be classified as selected-response or constructed-response types, depending on the task required of the learner. All item formats have advantages and limitations. Teachers should select item formats based on a variety of factors, such as the objectives, specific skill to be measured, and the ability level of the students. Many objectives are better measured with certain item formats.

Decisions about what scoring procedure or procedures to use are somewhat dependent on the choice of item formats. Student responses to some item formats must be hand-scored, whether they are recorded directly on the test itself or on a separate answer sheet or in a booklet. The teacher should decide whether the time and resources available for scoring a test suggest that hand-scoring or machine-scoring would be preferable.

The best way to ensure measurement validity of a teacher-constructed test is to develop a test blueprint, also known as a test plan or a table of specifications, before building the test itself. The elements of a test blueprint include (a) a list of the major topics or instructional objectives that the test will cover, (b) the level of complexity of the task to be assessed, and (c) the emphasis each topic will have, indicated by number or percentage of items or points. The test blueprint serves several important functions. It is a useful tool for guiding the work of the item writer so that sufficient items are developed at the appropriate level to test important content areas and objectives. The blueprint also should be used to inform students about the nature of the test and how they should prepare for it.

After developing the test blueprint, the teacher writes the test items that correspond to it. Regardless of the selected item formats, the teacher should follow some general rules that contribute to the development of high-quality test items. Those rules were discussed in the chapter.

Because teacher-made tests typically measure students' maximum performance rather than their typical performance, teachers should create conditions under which students will be able to demonstrate their best possible performance. These conditions include adequate preparation of the students to take the test. Adequate preparation includes information, skills, and attitudes that will facilitate students' maximum performance on the test.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York, NY: Longman.

Bosher, S. D. (2009). Removing language as a barrier to success on multiple-choice exams. In S. D. Bosher & M. D. Pharris (Eds.), *Transforming nursing education: The culturally inclusive environment* (pp. 259–284). New York, NY: Springer Publishing.

Bosher, S., & Bowles, M. (2008). The effects of linguistic modification on ESL students' comprehension of nursing course test items. *Nursing Education Perspectives, 29*, 165–172.

Brookhart, S. M., & Nitko, A. J. (2015). *Educational assessment of students* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). Upper Saddle River, NJ: Prentice Hall.

National Council of State Boards of Nursing. (2013). *2013 NCLEX-RN® detailed test plan: Item writer/ item reviewer/ nurse educator version.* Retrieved from https://www.ncsbn.org/2016_RN_DetTestPlan_ Educator.pdf

Poorman, S. G., Mastorovich, M. L., Molcan, K. L., & Liberto, T. L. (2011). *Good thinking: Test taking and study skills for nursing students* (3rd ed.). Pittsburgh, PA: STAT Nursing Consultants.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). *Measurement in nursing and health research* (4th ed.). New York, NY: Springer Publishing.

Test Construction and Preparation of Students for Licensure and Certification Examinations

One of the outcomes of prelicensure nursing programs is for graduates to pass an examination that measures their knowledge and competencies to engage in safe and effective nursing practice. At the entry level for professional nursing, graduates take the National Council Licensure Examination for Registered Nurses (NCLEX-RN®) or, if graduating from a practical or vocational nursing program, they take the National Council Licensure Examination for Practical Nurses (NCLEX-PN®). Certification validates knowledge and competencies for professional practice in a specialized area of nursing. As part of this process, nurses may take certification examinations that assess their knowledge and skills in a nursing specialty such as critical care or pediatric nursing. Other certification examinations measure knowledge and competencies for advanced practice, for teaching, and for administrative roles. As students progress through a nursing program, they should have experience with tests that are similar to and prepare them for taking licensure and certification examinations when they graduate.

Because the focus of the NCLEX and most certification examinations is on nursing practice, the other advantage to incorporating items of these types in teacher-made tests is that it provides a way of assessing whether students can apply their theoretical learning to clinical situations. Teachers can develop items that present new and complex clinical situations for students to critically analyze. Items can focus on collecting and analyzing data, setting priorities, selecting interventions, and evaluating outcomes as related to the content taught in the course. This type of testing is a means of assessing higher and more complex levels of learning and provides essential practice before students encounter similar questions on licensure and certification examinations.

This chapter begins with an explanation of the NCLEX test plans and implications for nurse educators. Examples are provided of items written at different cognitive levels, thereby avoiding tests that focus only on recall and memorization of facts. The chapter also describes how to write questions within the framework of clinical practice or based on the nursing process and provides sample stems for use with those items. The types of items presented in the chapter are similar to

those found on the NCLEX and many certification tests. By incorporating items of these types on tests in nursing courses, teachers help students acquire experience with this type of testing as they progress through the program, preparing them for taking licensure and certification examinations as graduates. The reader should keep in mind that Chapter 7 presented other ways of assessing higher level learning such as context-dependent testing, case scenarios, and other strategies for assessing critical thinking and clinical reasoning.

NCLEX TEST PLANS

In the United States and its territories, graduates of nursing programs cannot practice as registered nurses (RNs) or as practical nurses (PNs) or vocational nurses (VNs) until they have passed a licensure examination. These examinations are developed by the National Council of State Boards of Nursing, Inc. (NCSBN) based on extensive analyses of the practice requirements of RNs and licensed practical nurses (LPNs) or vocational nurses (LVNs). Because the NCLEX examinations are high stakes, items are piloted and tested extensively to ensure they are valid, reliable, and legally defensible, including an analysis of potential biases such as those related to ethnicity and gender (Woo & Dragan, 2012, p. 29). The licensure examination results then are used by the state boards of nursing as one of the requirements for practice in that state or territory.

NCLEX-RN TEST PLAN

In developing the NCLEX-RN, the NCSBN conducts an analysis of the current practice of newly licensed RNs across clinical areas and settings. This is a continuous process allowing the licensure examination to stay current with the knowledge and competencies needed by entry level nurses. To ensure that the NCLEX-RN measures the essential competencies for safe and effective practice by a newly licensed RN, the NCSBN reviews the test plan or blueprint every 3 years (NCSBN, 2015a, 2015b). For the most recent revision of the test plan, more than 3,500 newly licensed RNs prioritized how frequently they performed 139 nursing care activities and rated the overall importance of each activity. The nursing care activities are then analyzed based on the frequency with which they are performed, impact on patient safety, and settings where the care is provided (NCSBN, 2015a). In addition to the practice analysis, the NCSBN conducts a survey to assess the knowledge, skills, and abilities needed by newly licensed RNs to provide safe and effective care (NCSBN, 2015a). Findings from the RN practice analysis and the knowledge survey are used to develop the NCLEX-RN test plan and items on the test. A test plan is developed from this analysis, guiding the selection of content to be tested and the percentage of items for each of the categories of the test. The NCLEX-RN is based on this test plan.

Client Needs

Test items on the NCLEX-RN are categorized by client needs: (a) safe and effective care environment, (b) health promotion and maintenance, (c) psychosocial

integrity, and (d) physiological integrity. Two of the categories, safe and effective care environment and physiological integrity, also have subgroups. The client needs represent the content tested on the examination.

Safe and Effective Care Environment

In the safe and effective care environment category, two subcategories of content are tested on the NCLEX-RN: (a) management of care and (b) safety and infection control. In the management of care subcategory, the test items focus on providing and directing nursing care that enhances care delivery to protect clients and health care providers (NCSBN, 2015a). Examples of content tested in this category include advance directives; advocacy; assignment, delegation, and supervision; case management; collaboration with the interdisciplinary team; concepts of management; confidentiality/information security; continuity of care; establishing priorities; ethical practice; informed consent; legal rights and responsibilities; and performance improvement (quality improvement) among others (NCSBN, 2015a). In the NCLEX-RN, 17% to 23% of the items assess management of care.

In the safety and infection control subcategory, test items focus on prevention of accidents, emergency response planning, ergonomic principles, handling hazardous and infectious materials, reporting of incidents and irregular occurrences, safe use of equipment, standard precautions, and use of restraints, among others (NCSBN, 2015a). Between 9% and 15% of the items on the NCLEX-RN relate to safety and infection control.

Health Promotion and Maintenance

The second category of client needs is health promotion and maintenance. Between 6% and 12% of the items on the NCLEX-RN relate to health promotion and maintenance. There are no subcategories of needs. Examples of content tested in this category are the aging process, ante/intra/postpartum and newborn care, developmental stages and transitions, health promotion and disease prevention, lifestyle choices, physical assessment techniques, and others.

Psychosocial Integrity

The third category of client needs, Psychosocial Integrity, also has no subgroups. This category focuses on nursing care that promotes the emotional, mental, and social well-being of clients experiencing stressful events, and the care of patients with acute and chronic mental illness (NCSBN, 2015a). Examples of content tested include abuse, behavioral interventions, chemical and other dependencies, cultural awareness and influences on health, end-of-life care, grief and loss, mental health, sensory and perceptual alterations, and therapeutic communication and environment (NCSBN, 2015a). Six percent to 12% of the items on the NCLEX-RN ask questions about psychosocial integrity.

Physiological Integrity

The final client needs category, physiological integrity, is a significant content area tested on the NCLEX-RN. Items in this category focus on nursing care that promotes physical health and comfort, reduces risk potential, and manages health alterations. Four subcategories of content are examined by these items on the NCLEX-RN:

1. *Basic care and comfort*: In this area, items focus on comfort measures and assistance with activities of daily living. Related content includes assistive devices, elimination, mobility and immobility, nonpharmacological comfort interventions, nutrition and oral hydration, personal hygiene, and rest and sleep. Six percent to 12% of the items are on basic care and comfort.

2. *Pharmacological and parenteral therapies*: Items focus on adverse effects, contraindications, side effects, and interactions; blood and blood products; calculating dosages; central venous access devices; medication administration; parenteral/intravenous therapy; pharmacological pain management; and total parenteral nutrition. More test items are included on the NCLEX-RN in this subcategory than the others in the physiological integrity category. Between 12% and 18% of the items relate to pharmacological and parenteral therapies.

3. *Reduction of risk potential*: The content in this subcategory relates to measures for reducing the risk of developing complications or health problems. For example, items relate to diagnostic tests; laboratory values; potential for complications from diagnostic tests, treatments, and procedures, and from surgical procedures; and system-specific assessments, among others. In the test plan, 9% to 15% of the items relate to these content areas.

4. *Physiological adaptation*: The last subcategory includes nursing care of patients with acute, chronic, or life-threatening physical health problems. Between 11% and 17% of the test items relate to physiological adaptation. Sample content areas are alterations in body systems, fluid and electrolyte imbalances, hemodynamics, management of illness and medical emergencies, pathophysiology, and unexpected responses to therapies (NCSBN, 2015a).

Integrated Processes

Five processes that are fundamental to nursing practice are integrated throughout each of the categories of the test plan: (a) nursing process, (b) caring, (c) communication and documentation, (d) teaching and learning, and (e) culture and spirituality (NCSBN, 2015a). Thus, there can be test items on teaching patients and the nurse's ethical and legal responsibilities in patient education as part of the management of care subcategory, teaching nursing assistants about the use of restraints in the safety and infection control subcategory, health education for different age groups in the health promotion and maintenance category, and teaching about diagnostic tests in the reduction

of risk potential subcategory. The other processes are integrated similarly throughout the test plan. Many of the items on the NCLEX examinations are developed based on clinical scenarios. Those scenarios can involve any age group of patients in acute care hospitals, long-term care, community health, or other types of settings.

Cognitive Levels

The taxonomy for the cognitive domain is used for developing and coding items on the NCLEX-RN (NCSBN, 2015a). This taxonomy was presented in Chapter 1. The majority of items are at the application (applying) and higher cognitive levels (NCSBN, 2015a). This has implications for testing in prelicensure nursing education programs. Faculty members should avoid preparing only items that require remembering and understanding facts and specific information on their tests. Although some low-level questions are essential to assess students' knowledge, test items also need to ask students to *use* their knowledge and think critically to arrive at an answer. Test blueprints can be developed to list not only the content and number of items in each content area but also the level of cognitive complexity at which items should be written. An example of a blueprint of this type was provided in Exhibit 3.3 in Chapter 3.

NCLEX-PN TEST PLAN

The test plan for the NCLEX-PN is developed and organized similarly to the RN examination. For the 2014 test plan, PNs and VNs who were newly licensed were asked how frequently they performed 160 nursing activities and the importance of those activities (NCSBN, 2013). The activities were then used as the framework for the development of the test plan for the PN examination.

The test plan is structured around client needs and integrated processes fundamental to the practice of practical and vocational nursing. The same four client needs categories are used for the NCLEX-PN with differences in some of the subcategories, related content, and percentage of items in each category and subcategory. In the safe and effective care environment category on the NCLEX-PN, 16% to 22% of the items are on coordinated care, and 10% to 16% are on safety and infection control. Between 7% and 13% of the items assess the second category of client needs: health promotion and maintenance. Eight percent to 14% of the items on the NCLEX-PN address psychosocial integrity. The last category of client needs, physiological integrity, has four subcategories similar to the RN test plan: basic care and comfort (7%–13% of the items); pharmacological therapies (11%–17% of the items); reduction of risk potential (10%–16% of the items); and physiological adaptation (7%–13% of the items). Four processes are integrated throughout the test: (a) the clinical problem-solving process (nursing process), (b) caring, (c) communication and documentation, and (d) teaching and learning.

Items are developed at all cognitive levels with the majority written at the application or higher levels of cognitive abilities, similar to the NCLEX-RN test plan (NCSBN, 2013).

TYPES OF ITEMS ON THE NCLEX EXAMINATIONS

The NCLEX examinations contain multiple-choice items and alternate item formats. Earlier chapters described how to construct each type of item used on the NCLEX: multiple-choice (Chapter 5); the alternate formats of multiple-response (Chapter 5), fill-in-the-blank, calculation, and ordered response (Chapter 6); and hot-spot and chart or exhibit (Chapter 7). Any of these item formats on the NCLEX including multiple-choice might include a table, a chart, an image, or sound and video as part of the item. These items with multimedia have the capacity to assess higher levels of thinking and do so more authentically than a text-based item.

The NCLEX-RN Detailed Test Plan provides valuable information about the practice activities used for developing the items and content areas assessed in each of the categories and subcategories on the examination. As described earlier, the NCSBN analyzes the current practices of newly licensed RNs and PNs/VNs across clinical specialties and settings and the knowledge needed for safe and effective practice. This analysis identifies nursing activities that are used frequently by entry-level nurses and are important to ensure patient safety. Development of the NCLEX using these practice activities provides evidence of validity as a measure of entry-level nursing practice.

The NCLEX-RN Detailed Test Plan includes a list of the activity statements and related content for each category and subcategory. This information is of value in developing items for tests in a nursing program. For example, in the Safety and Infection Control subcategory, the activity statements describe the practices that RNs use to protect clients and health care personnel from health and environmental hazards. An example of one of these activity statements is: "Apply principles of infection control (e.g., hand hygiene, surgical asepsis, isolation, sterile technique, universal/standard precautions)" (p. 15). A sample test item also is provided with each category and subcategory. The sample item in the Safety and Infection Control subcategory assesses student understanding about infection control precautions for a patient with streptococcal pneumonia. In the NCLEX-PN test plan, the categories and subcategories are described with related content areas.

ADMINISTRATION OF NCLEX EXAMINATIONS

The NCLEX are administered to candidates by computerized adaptive testing (CAT). The CAT model is such that each candidate's test is assembled interactively as the person is answering the questions. Each item on the NCLEX

has a predetermined difficulty level. As each item is answered, the computer reestimates the candidate's ability based on whether the answer is correct or incorrect. The computer then searches the item bank for an item with the same degree of difficulty (NCSBN, 2015a). This is an efficient means of testing, avoiding questions that do not contribute to determining a candidate's level of nursing competence.

The standard for passing the NCLEX is criterion-referenced. The standard is set by the NCSBN based on an established protocol and is used as the basis for determining if the candidate has passed or failed the examination. The NCLEX-RN can range from 75 to 265 items, with 15 of those being pretest items that are not scored. After candidates answer the minimum number of items, the testing stops when the candidate's ability is above or below the standard for passing, with 95% certainty (NCSBN, 2015a). Because the NCLEX is an adaptive test, candidates complete different numbers of items, and therefore the test takes varying amounts of time. If a candidate's ability has not been determined by the time the maximum number of items has been presented or when the time limit has been reached, the examination then stops.

All RN candidates must answer a minimum number of 75 items. The maximum number they can answer is 265 within a time limit of 6 hours (NCSBN, 2016). On the NCLEX-PN, PN and VN candidates must answer a minimum of 85 items. The maximum number of items they can answer is 205, during the 5-hour testing period allowed (NCSBN, 2016).

PREPARATION OF ITEMS AT VARIED COGNITIVE LEVELS

When courses have higher level outcomes, tests in those courses need to measure learning at the applying and analyzing levels rather than at remembering and understanding. This principle was discussed in earlier chapters. Items at higher levels of cognitive complexity are more difficult and time-consuming to develop, but they provide a way of evaluating ability to apply knowledge to new situations and to engage in analytical thinking. The majority of items on the NCLEX are written at higher levels of cognitive ability, requiring application of knowledge and analytical thinking.

Students are at a disadvantage if they encounter only test items that ask them to recall facts as they progress through a nursing program. Low-level items assess how well students memorize specific information, not if they can use that knowledge to analyze clinical situations and arrive at the best decisions possible for those situations. Students need experience answering questions at the application and analysis levels before they take the NCLEX. More important, if course outcomes are at higher levels of cognitive complexity, then tests and other methods need to assess learning at those levels. In graduate nursing programs, test items should be developed at higher cognitive levels to assess students' ability to problem solve and think critically and to prepare them for certification examinations they might take as graduates.

When developing a new test, a blueprint is important in planning the number of items at each cognitive level for the content areas to be assessed. By using a blueprint, teachers can avoid writing too many items that require only recall of information. For existing tests that were not developed using a blueprint, teachers can code items using Bloom's taxonomy or the updated taxonomy of the cognitive domain and then decide if more higher level items should be added.

Remembering (Knowledge)

In developing items at varying cognitive levels, it is important to remember the learning outcome intended at each of these levels. Questions at the remembering or knowledge level deal with facts, principles, and other specific information that is memorized and then recalled to answer the item. An example of a multiple-choice item at the remembering level follows:

> Your patient is taking pseudoephedrine for his stuffy nose. Which of the following side effects is common among patients using this medication?
>
> a. Diarrhea
> b. Dyspnea
> c. Hallucinations
> d. Restlessness[1]

Understanding (Comprehension)

At this level, items assess understanding of concepts and ability to explain them. These questions are written at a higher level than remembering facts, but they do not assess use of information in a new context. An example of an item at the understanding level is:

> An adult female patient is a new admission with the diagnosis of acute renal failure. Her total urine output for the previous 24 hours was 90 mL. A urinary output of this amount is known as _____.

Applying (Application)

At the applying level, students apply concepts and other types of knowledge as a basis for responding to the item. At this level, test questions measure *use* of knowledge in new or unique situations. One method for developing items at this level is to prepare stems that have information that students did

---

[1]Correct answer.

not encounter in their learning about the content. The stem might present patient data, problems, or interventions different from the ones discussed in class or in the readings. If examples related to nursing care of adults, items might test the ability to use those concepts when the patient is an adolescent or has multiple coexisting conditions. An example of an item at the applying level is as follows:

A mother tells you that she is worried about her 4-year-old daughter's development because her daughter seems to be "behind." You complete a developmental assessment. Which of the following behaviors suggests the need for further developmental testing?

a. Cannot follow five commands in a row
b. Has difficulty holding a crayon between thumb and forefinger[1]
c. Is unable to balance on each foot for 6 seconds
d. Keeps making mistakes when asked about the day of the week

Analyzing (Analysis)

Questions at the analyzing level are the most difficult to construct. They require analysis of a clinical or other situation to identify critical elements and relationships among them. Items should provide a new situation for students to analyze, not one encountered previously for which the student might recall the analysis. Many of these items require learners to solve a problem and make a decision about priorities or the best approach to take among the options. Or, items might ask students to identify the most immediate course of action to meet patient needs or manage the clinical situation.

The difference between applying and analyzing items is not always readily apparent. Items at the analyzing level, though, should ask students to identify relevant data, critical elements in the scenario, and their interrelationships. In analysis level items, students should distinguish between significant and nonsignificant information and select the best approach or priority among those cited in the alternatives. An example of an item written at the analysis level is as follows:

You receive a report on the following patients at the beginning of your evening shift at 3 p.m. Which patient should you assess first?

a. An 82-year-old with pneumonia who seems confused at times[1]
b. A 76-year-old patient with cancer with 300 mL remaining of an intravenous infusion
c. A 40-year-old who had an emergency appendectomy 8 hours ago
d. An 18-year-old with chest tubes for treatment of a pneumothorax following an accident

PREPARATION OF ITEMS WITHIN FRAMEWORK OF CLINICAL PRACTICE OR NURSING PROCESS

One of the processes integrated into the NCLEX test plans is the nursing process. This is also a framework taught in many nursing programs. If not presented as the nursing process, most clinical courses address, in some form, assessment, data analysis, problems or diagnoses, interventions, and evaluation. These areas provide another useful framework for developing test questions. Items can examine assessment of patients with varied needs and health problems, analysis of data, identification of problems, selection of evidence-based interventions and treatments, and evaluation of the outcomes of care.

Current practices suggest that many test items focus on scientific rationale, principles underlying patient care, and selection of interventions. Fewer items are developed on collecting and analyzing data, determining patient problems, setting priorities and realistic goals of care, and evaluating the effectiveness of interventions and outcomes. Developing items based on clinical scenarios in the stems provides an opportunity to examine these outcomes of learning. These items also facilitate testing at a higher cognitive level because they are written in relation to specific scenarios in the stems, requiring students to apply their knowledge to the clinical situation.

The process for developing items within the framework of clinical practice or the nursing process begins with identifying the total number of items to be written. This includes specifying the number of items on each phase of the nursing process or more generally on assessment, problem identification, and so forth. On some tests, greater weight may be given to certain areas, for example, assessment, if these were emphasized in the instruction. As part of this planning, the teacher also maps out the clinical scenario to be developed in the stem as related to the course content. For instance, the teacher may develop a scenario on a young adult with sickle cell disease being seen in the emergency department and plan for two assessment items on pain, an item on a high-risk medication, and one item on evaluating the quality of the nurse's communication with the patient. A similar process may be used with other content areas for which this type of testing is intended. Items may stand alone, or a series of items may be developed related to one clinical scenario. In the latter format the teacher has an option of adding data to the situation and creating an unfolding case, which was discussed in Chapter 7.

EXHIBIT 8.1

Examples of Stems for Clinical Practice and Nursing Process Items
Assessment
The nurse should collect which of the following data?
Which of the following information should be collected as a priority in the assessment?
Which data should be collected first?
Which questions should the nurse ask (the patient, the family, others) in the assessment?
Your patient develops (symptoms). What data should the nurse collect now?

(continued)

EXHIBIT 8.1

Examples of Stems for Clinical Practice and Nursing Process Items (*continued*)

What additional data are needed to establish the patient's problems?

Which resources should be used to collect the data?

Which of the following information is a priority to report in a SBAR (situation–background–assessment–recommendation) communication to the (physician, nurse, other provider)?

Analysis

These data support the (diagnosis, problem) of_____.

Which (diagnosis, problem) is most appropriate for this patient?

The priority nursing diagnosis is _____.

The priority problem of this (patient, family, community) is _____.

A patient with (a diagnosis of, symptoms of) is at risk for developing which of the following complications?

Planning

Which outcomes are most important for a patient with a (problem of)?

What are the priority outcomes for a patient receiving (treatment)?

Which nursing measures should be included in the plan of care for a patient with (problem, surgery, treatment, diagnostic test)?

Which of the following nursing interventions would be most effective for a patient with (diagnosis of, problem of, symptoms of)?

The nurse is teaching a patient who is (years old). Which teaching strategy would be most appropriate?

Which intervention is most likely to be effective in managing (symptoms of)?

Implementation

Which of the following actions should be implemented immediately?

Nursing interventions for this patient include _____.

Following this (procedure, surgery, treatment, test), which nursing measures should be implemented?

Which of these nursing interventions is a priority for a patient with (problem)?

What evidence supports (nursing intervention)?

A patient with (a diagnosis of) complains of (symptoms). What should the nurse do first?

Which explanation should the nurse use when teaching a patient (with a diagnosis of, prior to procedure, surgery, treatment, test)?

Which of the following instructions should be given to the (patient, family, caregiver, nurse) at discharge?

Which of the following situations (incidents) should be reported immediately?

Evaluation

Which of these responses indicates that the (intervention, medication, treatment) is effective?

A patient is taking (medication) for (diagnosis, problem). Which of these data indicate a side effect of the medication?

<div align="center">**EXHIBIT 8.1**</div>

Examples of Stems for Clinical Practice and Nursing Process Items (*continued*)

Which response by the patient indicates improvement?

Which of the following observations indicates that the (patient, caregiver) knows how to (perform the procedure, give the treatment, follow the protocol)?

Which statement by the (patient, caregiver) indicates the need for further teaching?

Examples of stems that can be used to develop these items are provided in Exhibit 8.1. Teachers can select a stem and add content from their own course, providing an easy way of writing items within the framework of clinical practice or the nursing process. Sample items are as follows:

Assessment

An 8-year-old boy is brought to the emergency department by his mother after falling off his bike and hitting his head. Which of the following data is most important to collect in the initial assessment?

- a. Blood pressure
- b. Level of consciousness
- c. Pupillary response
- d. Respiratory status[1]

Analysis

The nurse practitioner is admitting a patient who complains of fatigue and myalgia, and has a rash across the bridge of the nose and cheeks. The practitioner finds a few ulcers in the patient's mouth. Prior laboratory tests included a positive C-reactive protein. These findings support a likely diagnosis of:

- a. Fibromyalgia
- b. Rheumatoid arthritis
- c. Scleroderma
- d. Systemic lupus erythematosus[1]

Planning

Your patient is being discharged after a sickle cell crisis. Which of the following measures should be included in your teaching plan for this patient? Select all that apply.

- ❏ 1. Avoid warm temperatures inside and outdoors.
- ❏ 2. Do not use nonsteroidal anti-inflammatory drugs for pain.
- ☑ 3. Drink at least eight glasses of water a day.
- ☑ 4. Eat plenty of grains, fruits, and green leafy vegetables.
- ☑ 5. Get a vaccination for pneumonia.
- ❏ 6. Keep cold packs handy for joint pain.

This is a sample from EVALUATION AND TESTING IN NURSING EDUCATION, FIFTH EDITION
Visit This Book's Web Page / Buy Now / Request an Exam/Review Copy

Chapter Eight Test Construction for Licensure Examinations **139**

Implementation

Your patient is in active labor with contractions every 3 minutes lasting about 1 minute. She appears to have a seizure. Which of the following interventions is the top priority?

a. Assess her breathing pattern[1]
b. Attach an external fetal monitor
c. Call the physician
d. Prepare for a cesarean delivery

Evaluation

A male adult patient was discharged following a below-the-knee amputation. You are making the first home health visit after his discharge. Which of the following statements by the patient indicates that he needs further instruction?

a. "I know to take my temperature if I get chills again like in the hospital."
b. "I won't exert myself around the house until I see the doctor."
c. "The nurse said to take more insulin when I start to eat more."[1]
d. "The social worker mentioned a support group. Maybe I should call about it."

PREPARATION OF STUDENTS FOR THE NCLEX EXAMINATIONS

A number of studies have been done over the years to identify predictors of success on the NCLEX-RN. Some factors related to performance on the NCLEX-RN are: SAT and ACT scores (Grossbach & Kuncel, 2011); scores on exit or prelicensure readiness examinations (Brodersen & Mills, 2014; McGahee, Gramling, & Reid, 2010; Penprase & Harris, 2013; Schooley & Kuhn, 2013; Sosa & Sethares, 2015); standardized assessment scores for specific content areas, for example, fundamentals, pharmacology, and medical–surgical nursing (Emory, 2013; McCarthy, Harris, & Tracz, 2014; Yeom, 2013); grades in nursing courses and graduation grade point average (Grossbach & Kuncel, 2011; Landry, Davis, Alameida, Prive, & Renwanz-Boyle, 2010; McGahee et al., 2010; Romeo, 2013); being a transfer student (Simon, McGinniss, & Krauss, 2013); and grades in science courses (Elder, Jacobs, & Fast, 2015; McGahee et al., 2010). Academic achievement, in terms of nursing course grades and overall grade point average, has been found across studies as predictive of student performance on the NCLEX-RN. In a meta-analysis of 31 independent samples with 7,159 participants, admissions test scores (SAT and ACT) and grades earned in nursing programs, especially grades in the second year, were the best predictors of performance on the NCLEX (Grossbach & Kuncel, 2011).

A second area of the literature on the NCLEX-RN focuses on methods of preparing students to pass the examination. Many schools use standardized examinations designed to predict student performance on the NCLEX-RN and determine students' readiness for taking the examination. By analyzing the results of standardized tests for NCLEX readiness, faculty members and students can work together

to design individual plans for remediation so that students will be more likely to experience first-time success on the licensure examination.

Some of these approaches include curriculum review to identify areas needing improvement, using readiness tests combined with remedial learning activities to better prepare students for the NCLEX, student self-assessment of content areas needing improvement, instruction for content mastery, test-taking tips, managing test anxiety, cooperative study groups, courses that guide formal NCLEX-RN preparation, and careful planning for the day of testing. Hanna, Roberts, and Hurley (2016) discussed the benefits of collaborative testing as an NCLEX enrichment activity. Experience with test items that are similar to the NCLEX prepares students for the types of items they will encounter on the licensing examination. In addition to these item formats, students also need experience in taking practice tests.

SUMMARY

The chapter summarized the NCLEX test plans and their implications for nurse educators. One of the principles emphasized was the need to prepare items at different cognitive levels as indicated by the outcomes of the course. Items at the remembering or knowledge level assess how well students memorized facts and specific information; they do not, however, provide an indication of whether students can use that information in practice or can engage in higher level thinking. To assess those higher level outcomes, items must be written at the applying or analyzing levels or evaluated by methods other than tests. It is worthwhile for faculty members to develop a test blueprint that specifies the number of items to be developed at each cognitive level for content areas in the course. By using a blueprint, teachers can avoid writing too many lower level items on an examination.

As students progress through a nursing program, they develop knowledge and skills to assess patients, analyze data, identify patient needs and problems, set priorities for care, select appropriate and evidence-based interventions, and evaluate the outcomes of care. Testing within the framework of clinical practice or the nursing process provides an opportunity to assess those learning outcomes. Items may be written about data to collect in the particular clinical scenario, possible problems, approaches to use, priorities of care, decisions to be made, varying judgments possible in a scenario, and other questions that examine students' thinking and clinical judgment as related to the situation described in the stem of the item. This format of testing also provides experience for students in answering the types of items encountered on licensure and certification examinations.

REFERENCES

Brodersen, L. D., & Mills, A. C. (2014). A comparison of two nursing program exit exams that predict first-time NCLEX-RN outcome. *Computers, Informatics, Nursing, 32*, 404–412.

Elder, B. L., Jacobs, P., & Fast, Y. J. (2015). Identification and support of at-risk students using a case management model. *Journal of Professional Nursing, 31*, 247–253.

Emory, J. (2013). Standardized mastery content assessments for predicting NCLEX-RN outcomes. *Nurse Educator, 38*, 66–70.

Grossbach, A., & Kuncel, N. R. (2011). The predictive validity of nursing admission measures for performance on the National Council Licensure Examination: A meta-analysis. *Journal of Professional Nursing, 27*, 124–128.

Hanna, K., Roberts, T., & Hurley, S. (2016). Collaborative testing as NCLEX enrichment. *Nurse Educator, 41,* 171–174.

Landry, L., Davis, H., Alameida, M., Prive, A., & Renwanz-Boyle, A. (2010). Predictors of NCLEX-RN success across 3 prelicensure program types. *Nurse Educator*, *35*, 259–263.

McCarthy, M. A., Harris, D., & Tracz, S. M. (2014). Academic and nursing aptitude and the NCLEX-RN in baccalaureate programs. *Journal of Nursing Education, 53*, 151–159.

McGahee, T. W., Gramling, L., & Reid, T. F. (2010). NCLEX-RN® success: Are there predictors? *Southern Online Journal of Nursing Research, 10*, 208–221.

National Council of State Boards of Nursing. (2013). *NCLEX-PN® detailed test plan for the National Council Licensure Examination for Practical Nurses.* Chicago, IL: Author.

National Council of State Boards of Nursing. (2015a). *NCLEX-RN® detailed test plan for the National Council Licensure Examination for Registered Nurses.* Chicago, IL: Author.

National Council of State Boards of Nursing. (2015b). *2014 RN Practice analysis: Linking the NCLEX-RN® to practice.* Chicago, IL: Author.

National Council of State Boards of Nursing. (2016). *2016 NCLEX® Examination candidate bulletin.* Chicago, IL: Author.

Penprase, B. B., & Harris, M. A. (2013). Accelerated second-degree nursing students: Predictors of graduation and NCLEX-RN first-time pass rates. *Nurse Educator, 38*, 26–29.

Romeo, E. M. (2013). The predictive ability of critical thinking, nursing GPA, and SAT scores on first-time NCLEX-RN performance. *Nursing Education Perspectives, 34*, 248–253.

Schooley, A., & Kuhn, J. R. (2013). Early indicators of NCLEX-RN performance. *Journal of Nursing Education, 52*, 539–542.

Simon, E. B., McGinniss, S. P., & Krauss, B. J. (2013). Predictor variables for NCLEX-RN readiness exam performance. *Nursing Education Perspectives, 34*, 18–24.

Sosa, M. E., & Sethares, K. A. (2015). An integrative review of the use and outcomes of HESI testing in baccalaureate nursing programs. *Nursing Education Perspectives, 36*, 237–243.

Woo, A., & Dragan, M. (2012). Ensuring validity of NCLEX® with differential item functioning analysis. *Journal of Nursing Regulation, 2*(4), 29–31.

Yeom, Y. J. (2013). An investigation of predictors of NCLEX-RN outcomes among nursing content standardized tests. *Nurse Education Today, 33*, 1523–1528.