The purpose of the test is to determine the level of mastery in the material. The purpose of the grade is to show the level of that mastery. There are several themes in the literature about tests that are tied to competency of the material and mastery of the material. Competency-based and Mastery-based education has preset expectations on what a student should know or be able to do in order to pass a module with the desired level of competency or mastery of the material. The following guidelines are a variation in the idea of mastery-based and competency-based learning.

1. **Having preset grade ranges are better than curving grades:** Students perceive curves as inherently unfair because it does not grade them on the mastery of the material, rather on how well they did relative to their peers. Students do not like the uncertainty of curves since they generally do not know where their overall course average falls relative to that of other students. Students prefer to have fair tests with preset grade ranges for the class (such as 90-100 = A, 80-89 = B, etc.). That way, they know where they stand in the course. Furthermore, curves create a sense of competition that may limit collaboration and group studying, particularly in smaller classes.

   A student who has proven competency and completed all or nearly all assignments should receive at least a C in the course. Students who receive a B or A have proven a higher level of mastery of the material. To test competence of the material, a reasonable practice is to have
   
   a. 75-80% of the test made up of problems that are very similar in level and composition to problems given previously in the class (homework, worksheets, class examples). These problems should be on the major topics in the class. A student who can work those problems has proven competency.

   b. 20%-25% of the test can be made up of
      
      i. more difficult problems needing a higher level of understanding, such as changing a typical problem to see if they can handle new assumptions or conditions.
      
      ii. Shorter, lower-credit problems that test minor topics. You can give several short problems on these topics, each having low weight commensurate with their relative importance in the course.

2. **Fewer well-chosen questions are better than time-pressured tests with too many questions:** The test should be a true measure of student performance in answering questions on the test. Students often complain that they knew the material for a test but that the test was too long. There is a theory in cognitive psychology that supports the notion of giving shorter tests to get a better measure of student understanding of the material. The Yerkes-Dobson curve shows a relationship between performance and stress (The original curve related performance to mental arousal or stimulus, but many more recent studies examine stress and pressure instead.)
A test that is too long for the time allotted means that students work under high levels of stress and are performing on the right, downward side of that curve. In this situation, chemicals are released in the brain that inhibit attention and decision-making, and degrades working memory.

How to make a test that is of appropriate length: Time yourself taking the test, working every problem to the level of detail as you would expect from the students. A rule of thumb is that the test should take you no more than 1/3 of the time allotted to the class. Why? You know exactly how to work the problems, so your time is a measure of the mechanics of working through a problem without accounting for the need to think about the problem. The extra 2/3 of the time allotted give students the time to think about the problem and to answer the questions closer to their peak performance level.

I have been giving tests this way for many years, competency-based and of appropriate length. The resulting averages are typically 75%-80%. I may adjust a particular test by adding points if I realized that the test was overly difficult or focused incorrectly (too much emphasis on topics that received little attention in the class). I do not curve the final grades for the class and use the standard scale of 90-100 = A, 80-89=B, etc. Students always know where they stand in the class. They may be more willing to help each other since the grades are based on an absolute rather than relative scale.

Note: The term “mastery-based learning” is often used for the practice of allowing students chances to retake a test (or other assignment) if they did not display the desired level of mastery the first time. In these cases, they are not allowed to move on to the next course module until they demonstrate that level of mastery on the previous module.

Additional comments from Kate Williams, PhD, Assistant Director of CTL:

“Here is my favorite resource on test creation. It proposes a series of principles for writing different types of questions and then provides an activity asking students to evaluate the quality of questions based on the design principles they propose.

http://ctl.byu.edu/sites/default/files/principles-of-test-creation.pdf

If you’d like something with more background/content, try this resource. It is directed toward alternate teacher training for K12, but the principles are applicable.

http://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf
The previous resource mentions Blooms, which I think is a useful framework for thinking about what you want to your test to accomplish and at what level the questions should be written. There are some additional resources like this to further investigate this idea if needed.

For grading, Vanderbilt's center for teaching and learning has a good overview: https://cft.vanderbilt.edu/guides-sub-pages/grading-student-work/. I also like the advice from Yale's center: http://ctl.yale.edu/teaching/teaching-how/chapter-5-grading-and-evaluation/grades-and-grading. And, finally, Brown has a good resource on creating and using rubrics: https://www.brown.edu/about/administration/sheridan-center/teaching-learning/assessing-student-learning/grading-criteria-rubrics”