Answers to: Knowledge Check Questions
Reliability and Validity

1. Weighing yourself on a scale 3 times and getting the following readings: 150 lbs., 157 lbs., and 153 lbs.
   This example directly illustrates low reliability: the scale is yielding inconsistent output (a 7 pound range) simply by getting on and off the scale three times. Measures with low reliability always have low validity as well. Although the construct (i.e., concept) of “weight” has validity, this scale could not provide a valid measure of weight because it doesn’t even yield consistent measurements in the first place.

2. Five hundred school-children take a personality test measuring shyness at the start of the school year. Test scores successfully help teachers predict children’s popularity during the school year and identify children who are most likely to be socially isolated.
   Criterion validity focuses on how well a measure predicts some outcome, and in particular, predicts future behavior. In this example, the shyness test has high criterion validity because it predicts popularity/social isolation. (Ultimately, if supported by other findings, this high criterion validity will help to establish the construct validity of this shyness test as well.)

3. Students who score in the top 10% on the ACTs (a reasoning ability test) tend to score in about the same percentile on the SATs (a different reasoning ability test).
   Construct validity represents the degree to which a measure actually assesses what it claims to assess. One type of support for construct validity is convergent validity: Scores on a measure should correlate strongly with scores on other measures of the same construct. If the ACT and SAT both claim to measure the same construct (reasoning ability), and if one of these tests has already established good construct validity, then the high ACT-SAT correlation helps to establish the convergent validity (and hence the construct validity) of the other test. Thus, of the answer options listed, construct validity is the most direct choice (i.e., for simplicity, convergent/divergent validity were not listed as separate options under construct validity). NOTE: If you chose “criterion validity,” it is true that scores on one test help to predict scores on the other test (any time a significant correlation between X and Y exists, this is the case). However, in practice people don’t use the ACT and SAT to predict each other’s scores: they use these tests to predict college achievement. Researchers would correlate these two tests to help establish the tests’ construct validity.
4. After many administrations, researchers administering a polygraph test begin to worry that the machine is actually measuring anxiety and not dishonest responses. This illustrates a concern about potentially low construct validity, because the concern is that the instrument (the polygraph) does not appear to be measuring the desired construct (dishonesty), but is instead measuring a different theoretical construct (anxiety).

5. A personality test that helps to predict the development of schizophrenia consists entirely of items such as “What is your favorite color?” and “Are red apples better than green apples?”
   In this hypothetical example, the personality test has low face validity, because the items on the test seem to be unrelated to the construct (schizophrenia) being measured. What on Earth do favorite colors and “red versus green apples” have to do with schizophrenia? But even though the items might look silly or irrelevant, the more important issue (in terms of developing psychological tests) is that the test has high criterion validity: based on the information provided in the example, the test helps to predict the development of schizophrenia. (“Low face validity” and “high criterion validity” would be the best answers here)

6. Individuals that score high on a questionnaire measuring racism on Tuesday morning are likely to score high on the same scale one month later. This illustrates high reliability because multiple administrations of the questionnaire are yielding similar results. Note that the questionnaire’s high reliability does not indicate anything about its validity. The questionnaire may have low or high validity – we need more information to determine this.

7. An observational coding system for marital conflict and unhappiness correctly predicts which couples will get divorced 95% of the time. This coding system has high criterion validity, because it is highly successful in predicting future divorce. (Ultimately, if supported by other types of evidence, this high criterion validity will help to establish the general construct validity of this coding scale. In other words, it will help to establish that this coding system really is measuring marital conflict and unhappiness).

8. Dr. Smith develops a personality test to measure a new psychological construct that he calls “ego self-centeredness.” His theory predicts that “ego self-centeredness” is not the same construct as “narcissism.” Research shows that his test of ego self-centeredness correlates .90 with a well validated test of narcissism. This finding decreases the construct validity of Dr. Smith’s test (and of his new concept of ego self-centeredness). Dr. Smith’s claim that ego self-centeredness and narcissism are different constructs would be supported by demonstrating discriminant (divergent) validity. Unfortunately, the very high correlation does not support the test’s divergent validity. Thus, it undermines the test’s construct validity. Instead, the findings tentatively suggest that Dr. Smith’s test may simply be another test of narcissism and that his proposed new construct of “ego self-centeredness” may simply be a re-labeled version of “narcissism.”
9. As theoretically predicted, scores on a new personality test that has been developed to measure shyness correlate only weakly with scores on a validated test of self-esteem.

The new test was designed to measure the construct of shyness. The researcher’s theory predicts that shyness is not the same construct as self-esteem. Therefore, the two tests should not correlate strongly. The fact that their correlation is weak supports the discriminant (divergent) validity of the new shyness test. In turn, good discriminant validity directly supports construct validity (which is the appropriate answer among the options listed).

10. Students in Professor Jones' Geography 215 class are assigned to read Chapters 1, 2, 3, and 4 for the first exam. All the chapters are similar in length and amount of material. In lecture Professor Jones conducts 3 lectures on the topics in each chapter. Students are told to study all chapters and lecture notes for their first exam. On the first exam, however, 90% of the exam questions are based on the material in Chapter 3 and Chapter 4, and only 10% of the questions are based on material in Chapter 1 and Chapter 2.

Most directly (the best answer) this example illustrates that Professor Jones' exam has low content validity. The sample of questions contained in the exam poorly represents the domain of material that students were asked to read about and which they learned about in lecture. Roughly 50% of the material covered in class and the text focused on concepts related to Chapters 1 and 2, but only 10% of the exam questions focused on topics from these two chapters. In addition, the poor content validity probably will cause many students to feel that this was not a "fair exam." If so, then poor content validity would lead to poor face validity.