Insights from Psychology and Psychometrics on Measuring Risk Tolerance

by Michael J. Roszkowski, Ph.D.; Geoff Davey; and John E. Grable, Ph.D., CFP®

Executive Summary

- Despite some arguments to the contrary, a client's financial risk tolerance can be measured accurately by a questionnaire, provided that the questionnaire has been developed in accordance with psychometric principles.
- The science of psychometrics has a set of standards by which to judge the quality of a questionnaire. These standards deal with the processes used to create the questionnaire as well as the characteristics of the results produced by the questionnaire.
- In questionnaire creation, the questions should be evaluated for their understandability and answerability, and their ability to differentiate between individuals with different levels of risk tolerance. Moreover, the questionnaire in its entirety should be subjected to an evaluation of its adequacy. Adherence to these principles can ensure that the questionnaire's results are both reliable and valid.
- Validity and reliability determine quality. A reliable questionnaire measures consistently, with known accuracy. A valid questionnaire measures what it claims to measure. The publisher of a questionnaire should provide evidence of the questionnaire's reliability and validity.
- Unfortunately, questionnaires commonly used by financial planners do not adhere to psychometric standards. They are generally too brief (a reliability problem) and contain too many "bad" questions (a validity problem).
- Bad questions are those dealing with constructs other than risk tolerance, such as risk capacity (how much risk the client can afford to take), time horizons, liquidity, and goals. Although important to the financial planning process, these issues are not part of the construct of risk tolerance. Questions that require explanation are also bad questions.
- Many of the commonly used "investor risk" questionnaires are actually asset allocation calculators mislabelled as risk tolerance tests.
- While few planners have the resources to develop and maintain a psychometrically sound questionnaire, all planners should know how to do due diligence on any questionnaire they use.

Michael Roszkowski, who holds a Ph.D. in educational psychology, is director of institutional research at La Salle University in Philadelphia, Pennsylvania. Previously, he was an associate professor of psychology at The American College in Bryn Mawr, Pennsylvania. His e-mail address is roszkows@lasalle.edu.

Geoff Davey is co-founder and CEO of FinaMetrica Limited in Sydney, Australia, provider of a psychometric risk-profiling system. Previously, he pioneered financial planning in Australia. He can be reached at geoff.davey@finametrica.com.

John E. Grable, Ph.D. CFP®, serves as program director for the financial planning program at Kansas State University in Manhattan, Kansas. He is best known for his risk tolerance research. Send correspondence to grable@humec.ksu.edu.

Can risk tolerance be measured with questionnaires? The ubiquity of risk tolerance questionnaires would suggest a definitive yes. According to Droms and Strauss (2003), the first financial risk tolerance questionnaire was published in 1984, and in the ensuing two decades their use has become increasingly more frequent and accepted. In fact, Cochran (2002) offers the following advice to financial planners: "If you do not have a risk tolerance questionnaire, develop one, and use it to help structure your clients' portfolios" (p. 2 of downloaded article).

But two recent articles question whether questionnaires can truly assess a client's risk tolerance. Bouchey (2004) devised a ten-question risk tolerance survey that he believed typified the questions used by financial planners and found that the questionnaire did not predict respondents' actual investment behavior, while Yook and Everett (2003) reported the disturbing finding that six "investor risk tolerance" questionnaires failed to correlate highly (correlations ranging from .31 to.78, with an average of .56).

Both papers reach some legitimate conclusions, and their authors are to be commended for raising concerns about current assessment practices in the industry. Bouchey is correct in concluding that his short homegrown test was a poor measure of risk tolerance. Likewise, Yook and Everett rightly contend that a majority of risk tolerance questionnaires in current use fail to provide a consistent picture of the same investor, and that this could lead to different recommendations depending on which test was used. We have no quarrel with these two conclusions.

However, we must take issue with the explanations provided for these findings, and the resultant implications. Yook and Everett maintain that the problem with such questionnaires lies in "the artificiality inherent in the risk-questionnaire design" (p. 50). According to Bouchey, "(t)he key weakness appears to be that traditional risk tolerance questionnaires are trying to get an answer to what is a technical question, one that is difficult for the average investor to comprehend." Both articles seem to imply (perhaps unintentionally) that questionnaires consequently cannot ever be valid measures of risk tolerance. For instance, Bouchey recommends that "planners may want to look at some other ways to guide them in drawing up portfolios for their clients."

While we agree that poor risk tolerance questionnaires are rampant in the financial services industry, we don't believe in a blanket condemnation of the questionnaire *as a method* for measuring risk tolerance. Our position is that appropriately designed questionnaires can validly and reliably assess risk tolerance provided that (1) *no* inappropriate questions are asked and that (2) *enough* appropriate questions are asked. In fact, we would go further and say that best practice requires the use of a valid and reliable questionnaire.

The problem with nearly all so-called risk tolerance questionnaires is that they have been constructed without regard to psychometrics. Commonly, they contain too many "bad" questions and not enough "good" questions. As a consequence, the results produced by such questionnaires are neither valid nor reliable. Psychometrics, a blend of psychology and statistics, is the measurement science for attributes such as risk tolerance. In psychometric terms, a valid test is one that measures what it purports to measure and a reliable test is one that does so consistently (with known accuracy).

In this article, we introduce the reader to basic concepts of good measurement principles by describing how inadequate risk tolerance test design could lead to the results observed by Bouchey and Yook and Everett. Although some statistical formulas will be discussed, we will refrain from presenting the formulas in their traditional mathematical format or delving into their derivations and proofs. Rather, we will take the reader through a step-by-step process to obtaining the final result. As part of this discourse, we will also canvas more general issues relating to the use of risk tolerance tests in the financial planning process.

As a result of this article, we don't expect all financial planners to be able to design their own psychometrically valid risk tolerance questionnaire. But we hope the psychometric principles presented here will help the financial planning profession design better tests, and that all readers will be better able to assess the validity of third-party questionnaires or their own questionnaires they may want to use on their clients.

Risk Tolerance, Risk Attitude, and Risk Capacity

Because the terms used to describe risk-related constructs are not always used with the same meanings, we begin by clarifying our use of the terminology. Some commentators (for example, Boone and Lubitz 2003) do not

talk about risk tolerance but rather talk of risk attitude (how much risk I *choose* to take) and risk capacity (how much risk I can *afford* to take). For others (for example, Cordell 2002), risk tolerance is a composite of risk attitude and risk capacity.

We agree that planners must understand their client's risk attitude (a psychological attribute) and risk capacity (a financial attribute). In this paper, we talk about risk tolerance and risk capacity, but use "risk tolerance" to mean the psychological attribute. We believe the majority of clients and planners use "risk tolerance" in this sense. We see risk tolerance being the client's emotional comfort with financial risk—how psychologically receptive an individual is to situations involving financial risk. Risk capacity, on the other hand, is about the extent to which the client's finances can sustain a financial setback.

Risk tolerance and risk capacity act as two unrelated constraints, which should not be combined into an amalgam but rather kept separate so that alternatives can be compared against each.

The Science of Psychometrics

Since the late 19th century, psychologists and statisticians have been developing techniques to quantify and assess psychological constructs such as risk tolerance. While this development has not been free of controversy, there is now a widely accepted discipline–psychometrics–dealing with psychological testing and assessment. Today, the technical quality of any psychological assessment device (which includes questionnaires) can be measured against internationally agreed psychometric standards.[1]

To meet these standards, a test must go through a rigorous development process. First, a large pool of questions is created and tested on representative samples of the population for which the test is intended, to see if the question is understandable and answerable by this audience. Questions that seem straightforward are often revealed to have poor understandability or answerability. Note that even though Bouchey believed he had eliminated technical jargon and made the questions short and simple, his respondents, who were "fairly well versed in financial investments," still informed him that they found some of the questions confusing.

Next, questions with apparent promise, based on their understandability and answerability, are tested on further representative samples using statistical criteria. The results are examined to determine if the statistical characteristics of the questions and the scoring algorithm are proper. Upon testing, questions that at first appear insightful are often revealed to have little or no statistical value in differentiating one respondent from another. Typically, question development requires multiple loops through both trial processes.

Reasons Why Risk Tolerance Questionnaires Can Fail to Correlate

The usefulness of a test is indicated by its validity and reliability. Validity is the extent to which a test actually measures what it claims to measure. Reliability indicates how consistent the results from the test will be. A test that is not reliable can't be valid, although a reliable test is not necessarily a valid one because it could be measuring the wrong thing consistently. Risk tolerance questionnaires may fail to correlate for two primary reasons:

1. The tests are really assessing different constructs (that is, at least one of them is not a valid test).
2. The tests are measuring the same construct, but at least one of them has low reliability, so the signal is lost because of the noise in the measures.

The failure to find high correlations between the six questionnaires studied by Yook and Everett can probably be attributed to these two causes, as discussed below.

Problems Associated with Risk Tolerance Questionnaires

Years ago it was not uncommon to find questions relating to physical risk tolerance in questionnaires designed to measure financial risk tolerance. Today, the more prevalent problem is that many risk tolerance questionnaires deal with financial matters that are not really part of the construct of risk tolerance. This is a legacy from the ubiquitous asset allocation calculators which were often incorrectly described and mistakenly thought of as testing risk tolerance (see Droms and Strauss 2003.) Even though Yook and Everett treated the questionnaires used in their study as though each assessed an individual's risk tolerance, a review of the actual questionnaires suggests that this assumption is unwarranted. For example, the Vanguard questionnaire "...makes asset allocation suggestions based on the information you enter about your investment objectives and experience, time horizon, risk tolerance and financial situation."

Likewise, at least half of the questions in Bouchey's questionnaire are not measures of risk tolerance. For example, the following has nothing to do with risk tolerance: "I make withdrawals from my investments to cover my living expenses." It may provide clues to a client's risk capacity or investment goals, but not to the client's risk tolerance. Another question in Bouchey's questionnaire is, "I do not plan to make withdrawals from this investment over the next several years." Questions about a client's time horizon (or age or stage of life), while valid for making investment recommendations, are invalid questions for assessing risk tolerance. A financial planning proposal is a recommendation about behavior–that a client should (or should not) do something. Behavior will be a function of goals, perceived risk, risk tolerance, and risk capacity, as well as other factors (Trone, Allbright, and Taylor 1996). Time horizon is relevant in a strategy-selection context but not in a risk-tolerance-assessment context. The expectation that a risk tolerance questionnaire should include risk capacity, time horizon, and other non-risk-tolerance questions is a consequence of familiarity with asset allocation calculators mislabeled as risk tolerance tests.

Mixing questions about more than one construct in a single brief questionnaire will almost invariably lead to an inaccurate assessment of all the constructs because none can be measured adequately due to the brevity of the questionnaire. Bouchey observed this himself, noting that "(b)y failing to answer just one of the questions correctly, a respondent moves closer to the middle, or moderate position. When two or three of the questions are incorrectly answered, the effect is magnified. Unless the respondents are totally consistent–and accurate–in all of their answers, therefore, chances are strong that just a few misinterpreted questions will change the entire thrust of their response."

Modeling packages often have "risk tolerance questionnaires" built into them but in most instances they do a shabby job of measuring this construct. Generally, the questionnaires are simplistically short or they require a level of investment-risk understanding beyond the vast majority of clients. In some cases these risk tolerance questionnaires are no more than re-labeled asset allocation calculators. Some financial planning firms have developed their own risk tolerance questionnaires and processes, of varying degrees of sophistication and sensibility, but all, to our knowledge, without regard to psychometrics principles.

Bouchey is again quite correct in stating, "One way to improve the reliability of a risk tolerance questionnaire might be to introduce more science into the process and enlist the help of psychologists or sociologists. These professionals have been trying to elicit answers from people for a long time and understand how to quantify them in ways that are more statistically valid than a random set of questions like those most planners use."

So, the first problem with industry-standard questionnaires is one of invalid questions dealing with capacity, time horizon, and other non-risk-tolerance issues.

The second problem relates to questions that require explanation. Such questions arise out of the misguided concept that the client should complete the questionnaire with the help of the planner. Once a planner plays an active role in the completion of a questionnaire, the results will be influenced and the objectivity of the test will be compromised. Surveys of the public (for example, Cutler and Devlin 1996) reveal a low level of financial literacy

and sophistication. Therefore, high-school-standard, plain English should be the order of the day. Financial terminology should be avoided if one aims for high understandability. Even something as straightforward (from a planner's perspective) as "bonds" could cause difficulties. Similarly, questions involving percentage rates of return are problematic. If inflation is not mentioned, some respondents will have difficulty answering this question because they want to know whether the return is before or after inflation. Yet once a question mentions inflation, the majority finds the question too difficult. (For examples of people's difficulties in comprehending and estimating inflation, see Bolton, Warlop, and Alba 2003; Hudson 1989; Krause and Granato 2003). As for questions involving means and standard deviations, they might as well be in another language (which, in reality, they are!).

Common methods of assessing risk tolerance share a third problem–namely, relying too heavily on questions overly focused on investment issues (another consequence of the ubiquity of asset allocation calculators). Financial planning is not just about investment advice but financial issues in general, and risk tolerance is relevant to all financial decisions.

What Is a 'Good' Risk Tolerance Question?

While it is possible to do a quick scan of a questionnaire for "bad" questions using the problems listed above as a checklist, determining what is a "good" question is not as easy. Users of risk tolerance questionnaires must look critically at the questions being asked of their clients. A question that appears to be suitable may not be, for reasons that only become apparent when it is subjected to psychometric scrutiny, and this process must be conducted on the questionnaire as a whole. Examples of "good" questions devoted solely to the assessment of financial risk tolerance can be found at www.risk-profiling.com/downloads/Sample.pdf. Table 1 provides further examples of "good" and "bad" questions.

TABLE 1

Examples of 'Good' and 'Bad' Risk Tolerance Questions

'Good' Questions[1]	'Bad' Questions[2]
1. When you think of the word "risk," which of the following words comes to mind first? a. Danger b. Uncertainty c. Opportunity d. Thrill	1. Do you anticipate having a large cash need within the a. Next year b. Next 2 to 3 years c. Next 4 to 7 years d. Next 8 or more years
2. Compared to other people you know, how would you rate your ability to tolerate the stress associated with important financial matters? a. Very low b. Low c. Average d. High e. Very high	2. How much discretionary income do you expect to have available in the next three years compared to today? a. Substantially less b. About the same c. Substantially more

1. Adapted from the Investment Risk Tolerance Questionnaire published by the American College, 1992.
2. These are examples of situational (rather than attitudinal) questions. They are relevant to financial planning decisions but not risk tolerance.

Designing Effective Questionnaires

Research indicates that planners should be concerned about the accuracy of any client questionnaire (test) they are considering. But how many questions would suffice and what level of accuracy is feasible? In psychometrics, these questions are answered through consideration of a test's "reliability." So let's turn our discussion to what constitutes psychometric reliability and what it means in terms of a test's performance.

Reliability

The score on any test, including questionnaires purporting to measure risk tolerance, consists of two parts: a true score and an error (that is, obtained score = true score ± error of measurement). All psychometric tests have

some margin of error, so it is a matter of degree. Reliability can be conceptualized as the ratio of the true score to the obtained score. In other words, reliability tells us what proportion of the test is non-error. If the error component is large, then the test is unreliable and will fail to give consistent results from one testing to the next, even if the client's risk tolerance has not changed. The error generally comes from sources in the test itself (such as ambiguous wording), but it also can be due to random situational factors, like the test-taker being anxious or tired the day the questionnaire is administered. Other situational factors include motivation, fluctuations in attention or memory, and recent experiences.

Correlation coefficients–statistics that range in value from 0 to 1—are used extensively in psychometrics. A correlation coefficient indicates how closely two things relate to each other (that is, "go together"). A correlation of 0 means that there is no relationship whatsoever, so knowing the value of the one thing tells us absolutely nothing about the value of the second thing. Conversely, a correlation of 1 indicates a perfect relationship. In a perfect relationship, knowing the value of one variable allows one to perfectly predict the value of the second variable. In real life, most correlations fall in between these extremes.

Standard Error of Measurement

The reliability of a test can be thought of as the correlation coefficient between the true score and the score as tested. *Reliability tells the planner the band in which the client's true risk tolerance score is located.* It is possible to estimate the typical margin of error in a test if two things are known: (1) the reliability of the test and (2) the standard deviation of the scores in the sample on which the test is normed. This statistic, called the standard error of measurement (SE^m), is obtained as follows:

Step 1. Subtract the correlation coefficient from 1. Let's use a reliability correlation coefficient of .53 as an example. So we have 1 – .53 = .47.
Step 2. Take the square root of the value from Step 1. In our example, √.47 = .6856.
Step 3. Multiply the value in Step 2 by the standard deviation. Let's assume that the standard deviation of the sample scores was 10 points. So, 10 x .6856 = 7, when rounded to a whole number.

Now we know that the SE^m is 7. So what? Well, with this information we can come up with an idea of the band in which the client's "true" risk tolerance score is located given the margin of error inherent in the test due to unreliability. This band is sometimes called the confidence interval. We can be 95 percent certain that the client's true risk tolerance lies in a range that is 1.96 times the SE (because 95 percent of a normal distribution lies within 1.96 standard deviations of the mean). In our example, the confidence interval is 7 x 1.96 = 13.72, or 14 points when rounded to a whole number. Thus, if the client scored a 60 on this risk tolerance test, his or her true level of risk tolerance is somewhere between the observed score of 60 and plus or minus 14 points. That is, the true risk tolerance score is a figure between 46 (60 – 14) and 74 (60 + 14). One would be correct in concluding that this is quite a wide spread.

Now let's suppose that the reliability of the test is higher, say .85 rather than .53. What impact will this have on the margin of error? Intuitively, the margin of error, as indicated with the SE^m, should be smaller at a reliability of .85. Let's do the math and see what it comes out to be exactly by plugging this value into our three-step formula. If we use the same standard deviation as before (10), the answer is 3.87, which we can round to 4. Thus, with a client scoring 60 on our test, we can be 95 percent confident that his or her true risk tolerance is within about 8 points (1.96 x 4) of the observed score. That is, the score is no lower than 52 and no higher than 68. This is a much smaller confidence interval than the 46 to 74 that we observed previously with a reliability of .53 (that is, 16 points versus 28 points). As is now evident, the smaller the SE^m, the more accurate the observed measure of risk tolerance becomes. All other things being equal, the SE^m depends on the reliability of the test: the higher the reliability, the smaller the SE^m.

A common question is, "What does it mean to be 95 percent confident?" Another way of understanding the concept of the SE^m and the confidence interval is to think of all the people who have been tested for risk

tolerance with a particular test. Given the SE^m of 4 discussed in the previous paragraph, it means that for 95 percent of them, their observed score (what they got on the test) and their true score would be within 8 points of each other (1.96 x 4).

What if you retested a large number of these people–does the SE^m mean that 95 percent of them would have differences between their two scores that are within 8 points? The answer is no. The interpretation of the SE^m presented above deals with differences between observed and true scores. The answer to the new question posed here involves what's been called the "reliability limits of agreement" by some statisticians and "repeatability" by others. It requires multiplying the SE^m by a value that is higher than 1.96. Specifically, we use 2.77. In our example of an SE^m of 4, the reliability limits of agreement would be 2.77 times 4, or about 11 points.

If the second risk tolerance score was somewhere between —11 and +11 points of the first score, it would not be considered unusual because it is within what would be expected given the reliability of the test. But a score higher or lower than this would be considered unusual enough to suggest that a real change in risk tolerance had occurred. An equivalent statement would be to say that for 95 percent of the people who took the risk tolerance test twice, their two scores should be within 11 points of each other if no change in risk tolerance took place.

The next question that often occurs at this point is, "What level of reliability should be expected in a risk tolerance test?" The recommendations vary depending on the type of test, but generally speaking, tests with reliabilities below .70 should not be used to make decisions about individuals because the margin of error is too large. Correlations of .80 to .89 are typically acceptable, and ones of .90 and above are excellent, but may be hard to achieve for personality measures (Heilbrun 1992, Nunnally and Bernstein 1994).

What Makes a Test Reliable?

Other things being equal, the more questions of the same type one asks, the more reliable an instrument becomes (Krus and Helmstadter 1987). Using an equation called the Spearman-Brown Prophecy formula,[2] we can come up with a satisfactory estimate of what the reliability would be if we increased the length of a risk tolerance test by a certain proportion. Let's consider a five-question test with a reliability of .44 and another one with a reliability of .53 as examples. How many questions would it take to make the first test reach a reliability of .80? The Spearman-Brown formula tells us that the answer is 25. What if we increased the length of the second test, the one with reliability of .53, to 25 questions? The reliability of that questionnaire would also go up–to .85, to be exact. Table 2 shows how the reliability of the two instruments would be increased by increasing the number of questions in steps of five.

TABLE 2

Estimated Reliability of Two Tests with Increasing Number of Questions

Number of Questions	Test 1	Test 2
5	.44	.53
10	.62	.69
15	.71	.77
20	.75	.81
25	.80	.85

The questions added to the risk tolerance measures must be "good" questions. Adding "bad" questions to the test will actually lower its reliability. There is a formal procedure called item (question) analysis to tell if the questions one contemplates adding to a test are good or bad, and what impact adding a particular question will have on the test's overall reliability. In essence, one determines whether the question works the same way as the overall test. One method of checking this is by looking at how strong the correlation is between answers to the question and answers to the overall test (the total score on the risk tolerance questionnaire). To achieve a given

level of reliability, we will need to ask fewer questions if the answers to the questions correlate highly with each other. Conversely, we will need to ask more questions if they correlate poorly with each other. Generally speaking, questions that have correlations below .30 with the overall risk tolerance score should be eliminated because they hurt the reliability of the questionnaire (Nunnally and Bernstein 1994).

Comparing Two Tests

In comparing tests, Yook and Everett found correlations ranging from .31 to .78, with an average of .56, and interpreted these correlations as evidence that the tests are measuring different constructs. This explanation is plausible, but another reason for the size of the correlations could be the reliability of the tests.

The correlations need to be seen in the context of what would have been reasonable given the reliability of the tests. Specifically, the maximum theoretical correlation of two tests is the square root of the product of their individual reliabilities. Take two tests with quite acceptable reliabilities of .8 and .9. The maximum theoretical correlation between them is $\sqrt{.8 \times .9} = .85$ (not 1). If you correlate results from the two tests, the best you could hope for is .85.

So even if both instruments were valid measures of risk tolerance, and each had acceptable levels of reliability, the correlation would still not be perfect because neither test is measuring the construct of risk tolerance with 100 percent reliability. In the language of psychometrics, the correlation between the observed measurements remains attenuated because of unreliability.

Now suppose the two tests have lower reliabilities, .5 and .6. Theoretically, the highest possible correlation between them should be approximately .55. Suppose the observed correlation was only .55. Despite the low correlation, it is still possible that they are measuring the same thing, although unreliably. We can examine the likelihood that both tests are measuring the same construct using a formula called "correction for attenuation."

To do this, the actual correlation between test results is divided by the maximum theoretical correlation. Dividing .55 by .55 gives us 1, a perfect correlation. With a value of 1, it is not so easy to conclude that the two tests are measuring different constructs as it was with a value of .55. So, the low correlations uncovered by Yook and Everett could be due to low reliability rather than the tests measuring different constructs.

What Makes a Test Valid?

Broadly defined, a valid test is one that actually measures what it purports to measure. There are various aspects of validity that can be considered in the development of a test, of which content validity and criterion-related validity are the most frequently reported. If a test has good content validity, the questions it asks are seen to be very relevant by those with expertise in the field (Anastasi and Urbina 1997).

Criterion-related validity is expressed as a correlation coefficient for the relationship between the test score and a separate measure of behavior related to the construct being tested (the criterion). In the case of risk tolerance assessment, the criterion would be actual behavior reflecting risk-taking propensity (for example, the proportion of stocks owned within a portfolio). If the criterion is collected at the same time the test is administered, it is called *concurrent* validity; if the criterion does not materialize until some later time, it is called *predictive* validity. Although stock ownership can be attributed to a variety of reasons, people who own stocks are generally more risk tolerant than people who do not own stocks. (Of course, no test can be expected to perfectly differentiate between owners and non-owners of stocks because more than risk tolerance is involved.) One should expect a useful risk tolerance questionnaire to correlate reasonably highly (.30 or greater) with stock/equity ownership and other forms of financial risk taking.

Generally, a lower value for a validity coefficient is more acceptable than for a reliability coefficient. Validity

coefficients as low as .40 are considered good (Heilbrun 1992). The reason is that most complex behavior is determined by more than one factor, so we can explain only part of the behavior in terms of any one construct, such as risk tolerance. For example, the correlation between the SAT (scholastic aptitude test) and college grades is about .40, yet most colleges find that the SAT is useful in making selection decisions. Similarly, the average validity coefficient between aptitude and job proficiency is only about .22 (Ghiselli 1973).

How Should Planners Assess Their Clients' Risk Tolerance?

Anyone can develop a questionnaire. The question is, "Does it work?" As should now be obvious, this question can be answered only by determining whether the questionnaire meets psychometric standards and thereby predicts how clients actually behave.

Psychologists divide behavior into cognitive (intellectual) and affective (emotional) domains. Years of research have shown that ordinarily it takes fewer questions to reliably assess cognitive traits than affective ones. Unfortunately for those who want a quick assessment, risk tolerance falls into the affective domain. To do the job correctly, a reasonable amount of time needs to be allotted to measuring it. Financial planners who seek a five- to ten-question test that is 100 percent accurate will be disappointed because no such instrument can ever be developed. Even without knowing anything about psychometrics, one should be skeptical about brief risk tolerance tests on the basis of pure logic. Think about it: on a five-question test, each question constitutes 20 percent of the total score. Changing just one answer could put the client into an entirely different risk tolerance category. This is far less likely on a 25-question test, where each question is only 4 percent of the total score.

Lest planners be concerned that clients will find a 25-question psychometrically designed test onerous, it should be remembered that if the questionnaire has been designed appropriately, the understandability and answerability of all questions will have been assured and the process will therefore take less time than one may think. A 25-question psychometrically designed test should take approximately 15 minutes to complete. Further, the one thing we all want to know more about is ourselves, so the process should be an enjoyable one for most clients. Surveys of respondents who have taken a 25-question psychometric test[3] show that they consider it a worthwhile exercise, which leads to a better understanding of themselves in relation to financial risk (and, in couples, to one another). In fact, a good risk tolerance test should be a bright spot in the otherwise somewhat burdensome initial fact-finding experience.

Notwithstanding Cochran's (2002) advice that if you don't already have a risk tolerance test you should develop one, it is unlikely that individual planners or small planning firms will be able to cost-justify the effort involved in developing (and maintaining) their own psychometric risk tolerance test. Hence, planners should consider using a third-party test where the publisher can substantiate that the test meets psychometric standards. But be aware that the results of such tests should not be used prescriptively as a replacement of discussion between planner and client. Rather, tests should be an objective input to that discussion (LeBaron, Farrelly, and Gula 1989). Even a good test occasionally can produce inaccurate results. Planners should realize from what was said about the standard error of measurement that the completed questionnaire and test report should be discussed with the client to obtain their confirmation (or otherwise) of the test results. Such discussion will, as a byproduct, lead to a more in-depth understanding of the client and, in couples (each of whom should do an individual test), will clarify and quantify the almost invariable differences.

Lessons Applied

As this article makes clear, regardless of whether a planner designs his or her own questionnaire or uses a preexisting test, it is essential to evaluate the final product with psychometric principles firmly in mind. While few readers will actually go through the process of calculating a questionnaire's reliability, there are some lessons that can be applied in a planning practice now.

First, planners should not blindly discount the questionnaire method for use in assessing risk tolerance based on the Bouchey or Yook and Everett studies. Given the limited number of "good" questions and the inclusion of too many "bad" questions, it is not surprising that the questionnaires produced dubious results. The inclusion of "bad" questions puts in doubt the validity of the measures; that is, what was being measured was not risk tolerance. Additionally, the reliability of the instruments is unknown and may have been low. Under circumstances where the instruments being evaluated were clearly flawed, no sensible conclusions can be drawn about the efficacy of unflawed instruments. If anything, this suggests that planners should only consider questionnaires that have proven psychometric properties.

Second, for a number of reasons, it is important to assess risk tolerance (how much risk I *choose* to take) and risk capacity (how much risk I can *afford* to take) separately. For a questionnaire to have construct validity, the instrument should be as pure a measure of the construct as possible. Otherwise, one does not really know what he or she is measuring. Usually, if one tries to measure more than one construct in a short questionnaire, none of the constructs is measured adequately (reliably) because of the brevity.

Third, planners should be skeptical of short questionnaires. Although at face value they might look like they can do the job, the reliability of such instruments is typically low, which can cause a client's risk tolerance to be inaccurately classified. Short questionnaires can only provide "ballpark" answers at best.

Fourth, planners should also exercise caution when using questionnaires that focus entirely on investments and exclude other financial situations involving risk. Financial risk tolerance plays a role in the entire financial planning process–not just investment planning.

Summary

While some financial planners may find a discussion of psychometric principles a bit intimidating and overwhelming, it is important to at least be aware that such a field of knowledge exists to guide one in getting accurate assessments. We don't expect all financial planners to be able to design a psychometrically sound test themselves, but all planners should be able to at least tell if someone else designed one with these standards in mind. Not having an appreciation for the principles of good measurement can lead to (a) inaccurate client assessment and (b) faulty conclusions regarding the usefulness of risk tolerance questionnaires. Psychometric principles, properly applied, can ensure validity and reliability in risk tolerance test results. Unfortunately, financial planners are seldom exposed to psychometrics in their training, and without a basic knowledge of the topic, they have no way to differentiate between good and bad measures of risk tolerance. Hopefully, this article will serve as an introduction to psychometrics and to what to look for in a risk tolerance test.

It is our contention that, rather than there being doubt about the usefulness of risk tolerance questionnaires, a good questionnaire (that is, one that was designed to meet psychometric standards) is an essential ingredient of a best-practice process by which planners can reach a professional understanding of a critical planning variable–their clients' risk tolerance.

Endnotes

1. For example, see American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999), *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.
2. For the reader interested in knowing how we arrived at the values in Table 2, we present the steps involved in the Spearman-Brown formula for Test 2:
   Step 1. Determine the multiple by which the current number of questions in the test will be increased. For example, if we increase a 5-question test to 25 questions, we're increasing it by a multiple of 5, or an expansion factor of 5. If we increased the number of questions from 5 to 10, this expansion factor would

be 2.
Step 2. Multiply the expansion factor by the test's current reliability. For example, with 5 questions, Test 2 has a reliability of. 53. So, multiplying .53 by the expansion factor of 5 (going from 5 to 25 questions) gives us 2.65. This value is the numerator of the Spearman-Brown formula.
Step 3. Subtract 1 from the expansion factor, 5 – 1 = 4.
Step 4. Multiply Step 3 by the test's current reliability, 4 x .53, = 2.12.
Step 5. Add 1 to Step 4, 1 + 2.12 = 3.12. Now we have our denominator.
Step 6. Divide Step 2 (the numerator) by Step 5 (the denominator), 2.65/3.12 = .849359, or .85, when rounded to two decimals, the value shown in Table 2.

3. One such survey by a public Web site, Financial Passages, managed by an Australian ING-subsidiary from 1997 to 2002, where visitors were able to complete their FinaMetrica (formerly ProQuest) risk profile, can be downloaded from www.risk-profiling.com/downloads/Financial_Passages_Survey.pdf.

References

Anastasi, A. and S. Urbina. 1997. *Psychological Testing*. New Jersey: Prentice Hall.

Bouchey, P. 2004. "Questionnaire Quest: New Research Shows that Standard Questionnaires Designed to Reveal Investors' Risk Tolerance Levels Are Often Flawed or Misleading." *Financial Planning* July 1.

Bolton, L. E., L. Warlop, and J. W. Alba. 2003. "Consumer Perceptions of Price (Un)fairness." *Journal of Consumer Research* 29, 4 (March): 474–491.

Boone N. M. and L. S. Lubitz. 2003. "A Review of Difficult Investment Policy Issues." *Journal of Financial Planning* May: 56–63.

Cochran, R. A. 2002. "Trends to Watch in 2003: Enduring Lessons from FPA's Success Forum." Research December. http://www.researchmag.com/articles/pdf/rl02_07.pdf.

Cordell, D. M. 2002. "Risk Tolerance in Two Dimensions." *Journal of Financial Planning* May: 30–36.

Cutler, N. E. and S. J. Devlin. 1996. "Financial Literacy 2000."*Journal of the American Society of CLU & ChFC* 50, 4 (July): 32–37.

Droms, W. G. and S. N. Strauss. 2003. "Assessing Risk Tolerance for Asset Allocation." *Journal of Financial Planning* March: 72–77.

Ghiselli, E. E. 1973. "The Validity of Aptitude Tests in Personnel Selection." *Personnel Psychology* 26: 461–477.

Heilbrun, K. 1992. "The Role of Psychological Testing in Forensic Assessment." *Law and Human Behavior* 16: 257–272.

Hudson, J. 1989. "Perceptions of Inflation." in K. G. Grunert and F. Ölander (Eds.) Understanding Economic Behaviour. New York: Plenum Publishers, 77–91.

Krause, G. A. and J. Granato. 1998. "Fooling Some of the Public Some of the Time? A Test for Weak Rationality with Heterogeneous Information Levels." *Public Opinion Quarterly* 62, 2 (Summer): 135–151.

Krus, D. J. and G. C. Helmstadter. 1987. "The Relationship between Correlational and Internal Consistency Notions of Test Reliability." *Educational and Psychological Measurement* 47: 911–915.

LeBaron, D., G. Farrelly, and S. Gula. 1989. "Facilitating a Dialogue on Risk: A Questionnaire Approach." *Financial Analysts Journal* May–June: 19–24.

Nunnally, J. C., and I. C. Bernstein. 1994. *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

Trone, D. B., W. R. Allbright, and P. R. Taylor. 1996. *The Management of Investment Decisions*. New York: McGraw-Hall.

Yook, K. C. and R. Everett. 2003. "Assessing Risk Tolerance: Questioning the Questionnaire Method." *Journal of Financial Planning* August: 48–55.