Valid and Reliable Risk Tolerance Assessment

By Geoff Davey

"Know your client" is the cornerstone that anchors all financial advice. With new rules 2090 and 2111, the Financial Industry Regulatory Authority (FINRA) has attempted to define what it means to know your client and to have a reasonable basis for your advice.

Client information can be categorized as hard, firm, or soft. Age, gender, assets and liabilities, insurances, and the like are hard data. Firm data are situational and aspirational, such as marital status or a savings goal. Soft data include personality, attitudes, values, motivations, etc. Soft data make two otherwise very similar individuals quite different clients and can be the most difficult to assess.

The most critical piece of soft data is risk tolerance.

The new FINRA rules became effective on July 9, 2012, and are attracting considerable attention. Under rule 2111 FINRA has defined "suitability" and goes so far as to describe the categories of information that must be obtained to ascertain the investment profile and to evaluate the suitability of an investment recommendation. This information includes risk tolerance specifically.

Advisors may have preferred to rely on their interviewing skills to determine a client's risk tolerance, but psychology tells us that tests are a better methodology and that the optimal methodology is to use a test as an input to a focused discussion. In this context, a test is a scored questionnaire.

It has become standard industry practice to use a risk questionnaire as a starting point for suitability with regard to risk. However, not all questionnaires are created equal. The scientific discipline called psychometrics tests soft attributes such as risk tolerance. Psychometrics is a blend of psychology and statistics, and it provides methodologies for developing tests and standards against which the efficacy of tests can be evaluated (Roszkowski et al. 2005).

A good test will be both valid and reliable: A valid test measures what it purports to measure and a reliable test measures consistently, with a known and acceptable level of accuracy.

In simple terms, a valid test will be on target and a reliable test will be accurate, as illustrated in figure 1.

Valid

Various aspects of validity 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 are seen to be very relevant by those with expertise in the field. 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). 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.

The Scholastic Aptitude Test (SAT) is a psychometric test with both content- and criterion-related validity. However, it is only a moderately good predictor of college success. SAT scores and college grades correlate at around

FIGURE 1: VALID AND RELIABLE

Reliable, not Valid | Valid, not Reliable | Neither Valid, nor Reliable | Both Valid and Reliable
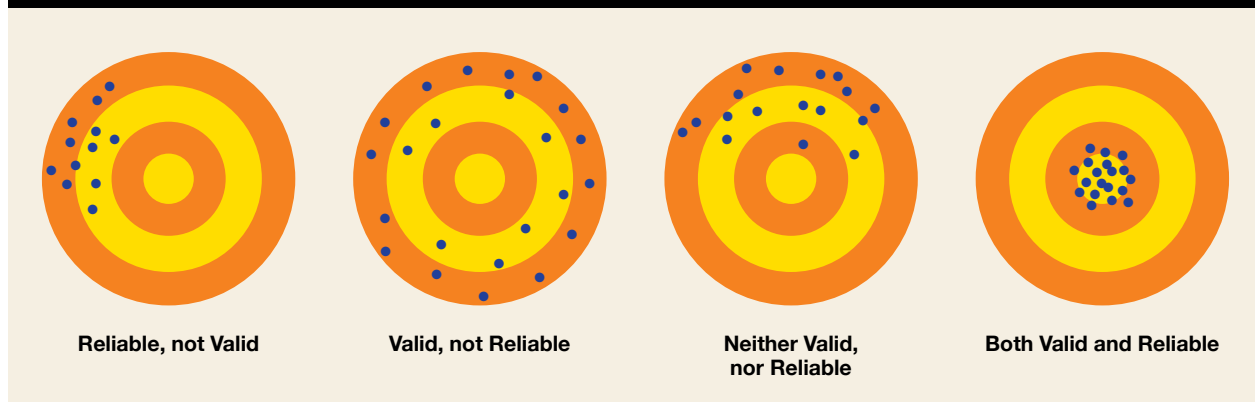
0.4. Given that college grades will be a function of more than just scholastic aptitude, this low correlation is only to be expected.

Similarly, risk tolerance is a predictor of financial risk actually taken, such as the percentage of stocks in a portfolio. But, again, because factors other than risk tolerance affect portfolio selection, risk tolerance scores and percentage of stocks correlate at around 0.4. However, an accurate understanding of a client's risk tolerance is a critical ingredient in the process of arriving at an optimal investment strategy.

In an optimal investment strategy, investment risk should be consistent with the investor's risk tolerance. Too much risk is likely to result in much unhappiness and possibly a panicked sale when risk becomes a reality. Too little risk means that the investor is missing out on opportunities.

Reliable

The score on any psychometric test consists of two parts: a true score and an error (that is, test 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 correlation between the true score and the test 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.

The error generally comes from sources in the test itself (such as ambiguous wording or too few questions), but it also can be due to random situational factors such as 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.

There are various psychometric techniques for determining a test's reliability, which will be expressed as a

number between 0 and 1. Acceptable levels of reliability range upward from 0.6 depending upon the purpose for which the test is to be used.

Reliability determines accuracy. Reliability tells you the band in which the true score is located by using a statistic called the standard error of measurement (SE^m) which is calculated as follows:

$$SE^m = \sigma \sqrt{(1 - r)}$$

where σ = the standard deviation of the scores of the sample upon which the test is based and r = the reliability of the test.

For example, where standard deviation is 10 and the reliability is 0.9, SE^m = 3.2. With this information we can determine the band in which the client's "true" 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 true score lies in a range that is 1.96 times the SE^m (because 95 percent of a normal distribution lies within 1.96 standard deviations of the mean.) In our example, the confidence interval is $3.2 \times 1.96 = 6$ when rounded to a whole number. Thus, a test score of 60 means that we can be 95-percent confident that the true score lies between 54 and 66, which will be sufficiently accurate for most purposes. However, if the reliability of the test had been 0.6, then the 95-percent confidence interval would be 48 to 72, which will be too inaccurate for most purposes.

In the United States, psychometric testing for personality generally, and attributes such as risk tolerance specifically, is still in its infancy in financial services. To our knowledge our firm's risk-profiling system is the only psychometrically based tool available in the United States. Elsewhere, however, psychometric testing is becoming much more common.

In the United Kingdom, for example, in addition to the FinaMetrica risk-

profiling system that has been available there since 2004, the behavioral finance team at Barclays Wealth has developed a suite of psychometric tests and the Oxford Risk Research Institute, a spin-off of the University of Oxford, and Towers Watson have developed psychometric risk tolerance tests for several major institutions. The U.K. regulator, the Financial Services Authority, which is overseeing a comprehensive, process-oriented, and principles-driven transformation of the financial services industry that began early last decade, is the primary driver of higher standards in the United Kingdom. The FSA's March 2011 Guidance Paper set standards for advice suitability with regard to willingness and ability to take risk—risk tolerance and risk capacity, respectively.

Testing a Test

An advisor considering using a psychometric test will want to be satisfied with its credentials. The test's publisher should provide a technical manual that, among other things, demonstrates that the test is valid and reliable. Ideally, there also will be independent certification from a credible third party that the test meets psychometric standards for validity and reliability.

The basic building blocks of a robust test are good questions. Good questions are identified through a two-step process,

- Questions with good content validity are first tested for usability—ease of understanding and ease of answering, through surveys, focus groups, and interviews. (Usually, a large pool of questions is required because many are eliminated.)
- A questionnaire is then constructed from questions with high usability and tested through norming trials, which evaluate both the statistical qualities of the questions individually and together, and the scoring algorithms. Questions that have high usability often are found not

to work statistically, i.e., they do not distinguish between individuals sufficiently well in a norming trial.

Details of this two-step process should be set out in a test's technical manual. While it is not possible to identify a good question[1] without usability and norming trials, it is possible to identify some bad questions with a sight check. Bad questions that can be identified in this manner include the following:

Questions that assume financial literacy. Any technical term, even one as simple as "bond," causes problems. With questions related to rates of return, more-informed respondents will question if the rate is before or after inflation. On the other hand, any mention of "after inflation" in a rate of return question is too difficult for most. Questions involving means, standard deviations, or confidence levels and the like may as well be in another language, which in reality they are. Standard plain English is the order of the day. Questions with good content validity, i.e., seen as relevant by experts, often fail usability trials because they presuppose too much financial literacy.

Questions that are off the topic. Questions about time horizon, investment experience, and the like, while important to formulating advice, are not relevant to risk tolerance. Such questions in simple language may pass a usability trial, but they will fail a norming trial because they will be shown to have no statistical value in determining risk tolerance. Also, because risk tolerance is domain-specific, questions about other forms of risk-taking, such as physical risk-taking, are not appropriate (Weber et al. 2002).

A more general guide to the efficacy of a risk tolerance test is whether or not the results for an individual are stable over time. It is now well-established that risk tolerance is a psychological trait, i.e., a relatively enduring way one individual differs from another, which

is not subject to external influences such as market conditions (Davey 2012). A test that shows risk tolerance rising and falling with markets cannot be a valid test.

Conclusion

The obligation to "know your client" covers both quantitative and qualitative information. Qualitative information, such as risk tolerance, is best assessed by using a valid and reliable psychometric test as input to a focused discussion between advisor and client. Psychometrics is an established scientific discipline with standards against which the validity and reliability of a test can be evaluated. Advisors striving for best practice will be well-served by valid and reliable psychometric testing for qualitative client information. **I&W**

Geoff Davey, co-founder of FinaMetrica®, specializes in risk tolerance and its role in the investment advising process. He earned a BSc from the University of Melbourne. Contact him at geoff.davey@finametrica.com.

Endnote

[1] See a hard copy sample of a risk tolerance test where the questions have passed both usability and norming trials at http://tinyurl.com/8c2vsnh.

References

Davey, G. 2012. On the Stability of Risk Tolerance. http://riskprofiling.com/Downloads/On_the_Stability_of_Risk_Tolerance.pdf.

Financial Services Authority. 2011. Assessing suitability: Establishing the risk a customer is willing and able to take and making a suitable investment selection. http://www.fsa.gov.uk/pubs/guidance/fg11_05.pdf.

Roszkowski, M. J., G. Davey, and J. E. Grable. 2005. Questioning the questionnaire method: Insights on measuring risk tolerance from psychology and psychometrics. *Journal of Financial Planning* 18, no. 4: 68–76.

Weber, E. U., A. E. Blais, and N.E. Betz, N. E. 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making* 15: 263–290.