PREDICTING FINANCIAL RISK TOLERANCE AND RISK-TAKING BEHAVIOUR: A COMPARISON OF QUESTIONNAIRES AND TESTS

John E. Grable*, Amy Hubble, Michelle Kruger, Melissa Visbal

* Corresponding author

John E. Grable
205 Dawson Hall, The University of Georgia, Athens, GA 30602
Tel: +1 707 542 4758
Email: grable@uga.edu

ARTICLE INFORMATION

Article history:

Submitted: 24 March 2020

Revision: 28 August 2020

Acceptance: 19 Spetember 2020

Key words:

Risk-taking, risk preference, risk tolerance, economic theory, classical test theory, psychometrics

ABSTRACT

The purpose of this study was to compare and contrast the concurrent, convergent, and predictive validity of a variety of risk tolerance tests and questionnaires. The tested measures represent tests and scales derived from economic and psychometric theory. It was determined that tests based on economic theory had similar predictive power, implying that economic measures provided some degree of convergent validity across measures. Only the psychometric risk tolerance measure, however, was found to be correlated to other indicators of risk tolerance and risk-taking. Results from this exploratory study suggest that a questionnaire developed using psychometric theory appears to offer superior predictive insight into financial risk-taking, at least when compared across the measurement techniques examined in this study.

Acknowledgements: The authors wish to sincerely thank Nicki Potts for her support and encouragement in conceptualizing and completing this research project.

Introduction

Financial planners and advisors are tasked with accurately assessing the risk tolerance of clients before making investment or financial recommendations or implementing recommendations.[1] Although numerous assessment approaches are used in practice, nearly all financial planners and advisors use some type of risk tolerance questionnaire or test (Moore 2018). Questionnaires and tests can be generally classified as based either on economic or psychometric theory.

Those who advocate the use of measures derived from economic theory assert that 'the only rigorous theoretical analyses relating risk tolerance to optimal portfolios are based on the economic concept of risk aversion' (Hanna & Lindamood 2004, p. 27), which is premised on the notion of expected utility. In practice, economic theory approaches estimate risk aversion (the inverse of which is risk tolerance) using an investor's responses to income, asset, or other gamble/lottery choice questions (Barsky et al. 1997) in an effort to derive an approximation of a person's revealed-preference (Frey et al. 2017). Regardless of the type of question asked, or the manner in which a choice scenario is framed, nearly all economic theory measures rely on choice scenarios that require a person to choose between two options—one with a guaranteed outcome and the other with a 50% chance of success and a 50% chance of failure. Barsky et al. (1997) argued that if enough gamble/lottery questions are asked, it is possible to obtain an estimate of a person's gamma coefficient, which can be used to construct a utility function where U is the utility function and c is permanent consumption or wealth. The risk aversion coefficient (λ) derived from a revealed-preference test represents the rate at which an individual will give up a higher expected income (or other asset) in exchange for less uncertainty. Theoretically, '[an] expected utility maximizer will choose the 50-50 gamble of doubling lifetime income as opposed to having it fall by the fraction $1 - \lambda$ if $\frac{1}{2}U(2c) + \frac{1}{2}U(\lambda c) > U(c)$' (Barsky et al. 1997, p. 540).[2]

Estimates of risk aversion (risk tolerance) developed from measures based on economic theory—referred to as tests of revealed-preference—are thought to provide insights into the way people make risky choices. Those with low risk aversion (high risk tolerance) are known to be more likely to engage in sensation seeking activities like smoking, drinking heavily, and using illicit drugs. There is some evidence to suggest that low risk aversion is also associated with being self-employed and holding risky financial assets.

The economic-based approach to measuring risk aversion (risk tolerance) is not without its critics (see Dow & Werlang 1992). Hanna, Gutter, and Fan (2001) noted that some revealed-preference tests fail to provide enough context to capture a person's true level of risk tolerance. Barsky et al. (1997) noted that the question response options imbedded in 50-50 gamble/lottery scenarios may be too complex for many people to answer accurately. In effect, choice scenarios may lead to random guessing.

[1] In the United States, the Securities and Exchange Commission recommends and FINRA requires financial advisors who provide investment advice to retail clients to evaluate each client's risk tolerance prior to making an investment recommendation. The Certified Financial Planner Board of Standards, Inc. also mandates that a 'CFP profession must act with the care, skill, prudence, and diligence that a prudent professional would exercise in light of the client's goals, *risk tolerance*, objectives, and financial and personal circumstances (2018, p. 2). In Australia, the Corporations Act 2001 places strict requirements on financial advisors to 'know your client' (KYC). KYC rules require that clients receive appropriate advice and that only suitable products are recommended. Imbedded in KYC rules is a requirement to assess a client's risk appetite and risk tolerance, with a client's risk tolerance guiding the selection of recommended products and services (ASIC 2011; ASIC 2017).

[2] See Hanna and Lindamood (2004) for a complete description of the way in which constant relative risk aversion can be calculated.

Outside of economics, psychologists and behaviorists frame the measurement of risk attitudes differently. Instead of relying on choice scenarios in which the probabilities of choice outcomes are always based on predetermined probability estimates, psychometricians generally assume that choices are based on subjective probability estimates made by individuals at the time a decision is made. This assumption has led to the development of robust questionnaires based primarily on classical test theory. Outcome scores from such questionnaires are sometimes referred to as propensity measures (Frey et al. 2017).[3] Classical test theory is premised on the notions of reliability and validity. A robust questionnaire based on psychometric principles is one in which random error is minimized across questions. To do this, a test developer focuses on asking appropriate questions that help uncover a person's attitudes and future behavior. If developed properly, questionnaire items can be combined into a preference scale or index (Faff, Hallahan & McKenzie 2009).

Economists, behaviorists, and commercial software testing firms typically adopt one testing tradition when developing measurement tools. There is little agreement among proponents of one approach or another regarding test methodologies. Each assessment procedure offers users advantages and disadvantages. For example, the revealed-preference approach has the advantage that scores can be mapped to a mean variance optimized portfolio recommendation. A disadvantage, however, is that the questions used to estimate risk tolerance may not be valid with some test takers or populations. For example, traditional risk tolerance measurement techniques require a test taker to exhibit relatively high cognitive skills (Charness, Gneezy & Imas 2013). The psychometric approach has the advantage of providing users with statistically robust measures of reliability and validity. A disadvantage is that it is difficult to map psychometric risk tolerance scores to a portfolio or other financial recommendation.

To date, there have been few attempts to compare assessment methods. This study was undertaken to address this gap in the literature by addressing the following questions: (1) do measures based on economic theory and psychometric classical test theory exhibit concurrent validity;[4] (2) do measures based on economic theory and psychometric classical test theory exhibit convergent validity;[5] (3) do measures based on economic theory and psychometric classical test theory exhibit predictive validity; and (4) which measurement approach (i.e., revealed-preference testing or psychometric scaling) provides the clearest insight into investor risk-taking tendencies? As will be shown in this paper, while both approaches offer unique advantages and disadvantages, scores from a psychometric risk tolerance measure appear to be more valid in describing financial risk-taking behavior. The remainder of this paper reviews relevant literature associated with the research questions. This is followed by an explanation of the research methodology, with a special emphasis on describing the different measures of financial risk tolerance evaluated in the study. The paper concludes with a report of findings and an applied discussion of results.

[3] Other psychometric theories can be used to frame questions, including item response theory (IRT) and Rasch modelling.

[4] Concurrent validity refers to the degree to which a measure 'agrees with another contemporaneous measure of the same concept' (Remler & Van Ryzin 2015, p. 110).

[5] Convergent validity is measured as a correlation between a measure and an outcome one would expect the measure to be related to in practice (Remler & Van Ryzin 2015).

Literature review

Among test developers, two traditions dominate the way financial risk tolerance assessments are created: measures of revealed-preference and psychometric assessments. Economists tend to utilize revealed-preference assessment techniques, whereas psychologists and those who construct tests using classical test theory utilize psychometric tools (Frey et al. 2017). Both revealed-preference and psychometric approaches stand in contrast to the use of stated preferences (see Adamowicz, Louviere & Williams 1994), which is a measurement technique often employed by financial planning practitioners when attempting to measure the extent to which someone is willing to take a financial risk in which the outcomes of the risk are potentially negative and unknown.[6]

The methodology used by a particular test developer tends to be driven by the person's academic training. In order to understand the test development process in the context of financial risk tolerance, it is first useful to review the difference between risk and uncertainty because these two constructs often shape the manner in which tests are developed. According to Knight (1921) and described by Weber and Johnson (2009), '*risk* refers to situations where the decision-maker knows with certainty the mathematical probabilities of possible outcomes of choice alternatives. *Uncertainty* refers to situations where the likelihood of different outcomes cannot be expressed with any mathematical precision' (p. 132). Those who advocate the use of revealed-preferences argue that 'uncertain situations can be reduced to risky situations' (Weber & Johnson 2009, p. 132). Ellsberg (1961) found fault with this argument by noting that people exhibit what he called ambiguity aversion, which makes it difficult to simplify uncertainty into a risk analysis.

The revealed-preference tradition, however, offers several unique advantages, not the least of which is the notion that revealed-preferences are easy to identify. Within the domain of financial risk tolerance, a person's revealed-preference can be assessed by asking a test taker to choose between and among choice alternatives where monetary incentives (real or hypothetical) are provided (Eckel & Grossman 2008). The notion underlying this assessment approach is that what people say they will or will not do is often different from the choices made when faced with a choice that involves relatively little variance in monetary outcomes compared to a choice alternative with more variance (Samuelson 1948). Sometimes referred to as a behavioral score, revealed-preferences are generally assessed by documenting actual engagement in behavior (e.g., how often a person engages in or chooses a particular risky behavior) or as an evaluation of hypothetical monetary gambling/lottery choices. According to Frey et al. (2017), revealed-preference measures are generally designed to 'capture specific cognitive processes, such as the integration of gains and losses or the role of learning and experience' (p. 1). Proponents of the revealed-preference tradition in relation to financial risk tolerance assessment often argue that this measurement technique provides the most direct link to identifying a person's utility function when investment choices are being made (Houthakker 1950; Richter 1966; Weber & Johnson 2009).

Objections have been raised about the wide use of revealed-preference methodologies. Frey et al. (2017) stated that measurements of revealed-preference may actually be capturing situational characteristics that help a person adapt to a particular situation (i.e., states) (Buss 1989) rather than traits, which can be thought of as preferences that exhibit consistency across time (Josef et al. 2016; Reynaud & Couture 2012). Others have noted that questions used in revealed-preference measures,

[6] An example of a stated preference item is, 'If the markets were to fall by 10 per cent later this week, what would you do?'

particularly assessment techniques that rely on a test taker to choose between gambles or lotteries, demand a cognitive ability that falls outside the norm for generalized assessments (Dave et al 2010), which can create noisy data (Charness, Gneezy & Imas 2013). Corter and Chen (2006) reported that familiarity with concepts related to expected value and probability outcomes can create a familiarity bias that distorts the way some people answer choice dilemmas. In other words, answers derived from revealed-preference measures may not capture a decision-maker's true willingness to take a financial risk. Another criticism is that to fully capture a revealed-preference, a test developer must assume that a test taker has access to full information and uses such information. As noted by Fischhoff et al. (1978), these assumptions are typically not observed and are unlikely in practice.

The primary alternative to a revealed-preference assessment is a psychometric test, which is sometimes referred to as a propensity measure. Psychometric tests are designed to assess a test taker's attitudes in a way that uncovers an underlying trait.[7] Psychometric tests are widely used to assess intelligence, personality, and other psychological constructs.[8] An advantage associated with psychometric measurements is that a well-designed test can account for a test taker's deeply held feelings of regret, fear, greed, and happiness associated with financial decision-making. Rather than a test taker's score being derived primarily from cognitive appraisals, psychometric tests allow for the incorporation of emotional factors. The primary argument against the use of psychometric tests is that what a person states may not be related to the person's ultimate behavior. Fischhoff et al. (1978) argued that this criticism is likely overstated. They noted that 'attitudes elicited in surveys often correlate highly with behavior . . . Furthermore, they elicit present values rather than historical preferences (p. 130). Research conducted over the past decade has confirmed Fischhoff et al.'s findings. Several researchers have recently noted the superiority of psychometric tests in the domain of financial risk-taking. For example, Dohmen et al. (2011) compared self-reported attitudinal risk-taking questions to hypothetical gambling questions. Dohmen et al. reported that the self-rating items did a better job of predicting actual financial risk-taking behavior. A similar finding was reported by LÖnnqvist et al. (2015).

In addition to direct comparisons of measurement tools, some researchers have attempted to determine how well revealed-preference and psychometric questionnaires and tests correlate with each other. In general, and particularly in the context of financial decision-making, associations between measures have been weak. Consider the following conclusion from Frey et al. (2017): '. . . measures from the propensity and behavioral measurement traditions cannot be used interchangeably to capture risk preference' (p. 8). This implies that those who rely on scores from risk assessments (e.g., financial planners, investment advisors, and other professionals who counsel individuals on day-to-day financial matters) cannot assume consistency across measurement techniques, nor can users of assessments presuppose that all assessment tools are equal. In effect, a decision must be made as to the validity of one measurement technique over the other. The findings from this study provide additional information to help guide the choice of financial risk tolerance assessment tools for practice.

[7] Indicators that a test is likely built on psychometric principles include reports of reliability and validity. Within classical test theory, reliability refers to the 'consistency of individuals' responses to an instrument across measurement occasions and is a descriptive statistic designed to capture how much measurement error is in a variable (Beauchamp, Cesarini & Johannesson 2017, p. 205). Validity refers to how well a measurement tool actually measures what it is purported to measure.

[8] It is possible to compute the reliability coefficient for a revealed-preference measure. One approach involves using a test-retest reliability estimate, which essentially involves calculating polychoric correlations. Beauchamp et al. (2017) estimated the reliability coefficient for the Barsky et al. (1997) revealed-preference test to be approximately .59, which falls below the typical criterion of .70 (see Nunnally & Bernstein 1994).

Methods

Recruitment for participation in the study was conducted during late 2017. Given the exploratory nature of the project, recruitment flyers and emails were distributed in one college community in a southeastern US state. Based on institutional IRB restrictions, potential participants were screened by age. Specifically, in order to be eligible for participation, a participant needed to be at least 21 years of age at the time of the study. The recruitment period was open for six weeks. At the end of the recruitment period, 164 participants were selected for the study. Each participant was sent an email survey link using Qualtrics. Participants were offered a $US10 gift card upon completion of the survey and related tasks. A unique code was assigned to each participant who finished the survey.

Participant characteristics

The general characteristics of participants in this study were skewed towards those who were young with a high level of attained education. This was not surprising given the general demographic profile of the community in which this study was conducted. Slightly more than 48% of participants were female. The average age of participants was close to 26 years of age, with the majority (56%) being unmarried. Other marital statuses included married (13%), divorced (2%), and other, including those who were widowed or not married but living with a significant other (29%). The majority of participants lived in a household of two or fewer people. Slightly more than three out of four participants were currently employed, with the remainder being retired or students. Only 4% of participants indicated being self-employed. Approximately 77% of participants indicated being Caucasian/White, whereas 10% of participants indicated being Asian or Pacific Islander. Other racial/ethnic categories included African-American/Black (6%), Hispanic (6%), and Native American (1%). The median personal income of participants fell between $US30,001 and $US40,000, whereas median household income fell between $US40,001 and $US50,000. The majority of participants held at least a college degree level of education.

Outcome measures

Participants received an email with a link to the survey. The survey comprising attitudinal, demographic, and risk tolerance measures (both revealed-preference tests and psychometric questions). Data collection occurred over a three-week period. At the end of the survey data gathering phase of the study, 40 participants were randomly chosen to complete a financial risk-taking activity. The activity comprised the following steps:

(1) Each participant was asked to visit the research team's lab to receive a participation incentive ($US10). Once in the lab, the participant was asked if they would be interested in an opportunity to win an additional $US10 or $US20 by playing a simple game of chance (i.e., a monetary risk-taking activity). The question was asked as the participant and interviewer stood next to a Las Vegas style craps table.[9]

(2) Those who indicated no interest were then asked to draw a winning ball from an opaque jar. This task was manipulated so that each participant "won," meaning that everyone who opted out of the task left the activity with $US30 in gift cards.

[9] Although not widely used in personal and household finance studies, the use of dice tasks as a measure of decision-making under risk and uncertainty is common in the gaming and risk-assessment literature (see Brand et al. 2005; Schiebener& Brand 2015).

(3) Those who indicated an interest in playing the game were read the following script:

Here is how the game works. You will be given dice to roll. You must wager your $US10 gift card. In order to win $US10, you must roll a 5, 6, 8, or 9. If you roll any other number you will lose $US10. In order to win $US20, you must roll a 2, 3, 4, 11, or 12 to win; if you roll any other number you will lose $US10. Which game would you like to play?

(4) Each participant's choice was recorded. Participants were then allowed to take a practice roll of the dice. This was followed by the interviewer reading the following script:

Okay, before you roll, we would like to share with you the true odds associated with your choice. The odds of rolling a 5, 6, 8, or 9 to win $US10 is 50% or 1 out of 2. The odds of rolling a 2, 3, 4, 11, or 12 to win $US20 is 25% or 1 out of 4.

Knowing this information, would you like to change your decision? You may also still choose to withdraw from the game and leave with your $US10.

(5) Each participant's choice was recorded. Those who decided to leave the game were then asked to draw a winning ball from an opaque jar. This task was manipulated so that each participant who opted out "won," meaning that those who exited the game left the activity with $US30 in gift cards.

(6) Those who remained in the game were asked to confirm their risk-taking choice and to roll the dice. Those who won an additional $US10 gift card where then asked to draw a winning ball from the same opaque jar. This task was again manipulated so that each participant "won." In this case, the participant received another $US10, bringing the total incentive to $US30 in gift cards. Those who won the $US20 gamble were congratulated and given $US30 in gift cards. Participants who lost the game were then asked to draw a winning ball from the same opaque jar. As with the other drawing scenarios, the task was manipulated so that each participant "won," thus leaving the study with $US30 in gift cards.

Participants did not know prior to the game that they were guaranteed to leave with $US30. Participant behaviors related to the game were used as an indicator of risk preference and behavior.

Economic theory measures

As noted above, the use of 50-50 gambles and lottery choice scenarios is the primary way someone's revealed-preference is measured within economically-derived models. In this study, the widely used Barsky et al. (1997) test of risk aversion was used as one economic choice indicator (heretofore referred to as the Barsky test). Barsky and his associates published the following series of questions to classify individuals into one of four categories of risk aversion (i.e., high, above-average, below-average, and low):

Question 1: Suppose that you are the only income earner in the family, and you have a good job guaranteed to give you your current (family) income every year for life. You are given the opportunity to take a new and equally good job, with a 50-50 chance it will double your (family) income and a 50-50 chance it will cut your (family) income by a third. Would you take the new job?

If the answer to this question was 'yes,' the participant was then asked:

Question 2: Suppose the chances were 50-50 that it would double your (family) income, and 50-50 that it would cut it in half. Would you still take the new job?

If the answer to the first question was 'no,' the participant was then asked:

Question 3: Suppose the chances were 50-50 that it would double your (family) income and 50-50 that it would cut it by 20 percent. Would you then take the new job?

Participants who answered 'no' to the first and third questions were classified as having high risk aversion (i.e., low risk tolerance). A participant who answered 'no' to the first question and 'yes' to the third question was classified as having above-average risk aversion. A participant who answered 'yes' to the first question and 'no' to the second question was classified as having below-average risk aversion. Those who answered 'yes' to the first and third questions were classified as having low risk aversion (i.e., high risk tolerance).

A similar measure was used as a second economic choice indicator: the Hanna and Lindamood (2004) subjective risk tolerance test (heretofore referred to as the H&L test). This measure used income choice scenarios that require participants to choose between pensions with 50-50 answer options.[10] The questions and scoring are shown below:

1. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension B has a 50% chance your income will be double your preretirement income, and a 50% chance that your income will be 20% less than your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, go to #2. If B, go to #5.*
2. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension C has a 50% chance your income will be double your preretirement income, and a 50% chance that your income will be 10% less than your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, go to #3. If C, your subjective risk tolerance is moderate.*
3. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension D has a 50% chance your income will be double your preretirement income, and a 50% chance that your income will be 8% less than your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, go to #4. If D, your subjective risk tolerance is low.*
4. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension E has a 50% chance your income will be double your preretirement income, and a 50% chance that your income will be 5% less than your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, your subjective risk tolerance is extremely low. If E, your subjective risk tolerance is very low.*

[10] This study used only the question narrative from the Hanna and Lindamood (2004) test. The original framework included visual representations of the 50-50 choice scenarios; these were not shown to participants in this study.

5. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension F has a 50% chance your income will double your preretirement income, and a 50% chance that your income will be one third less than your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, your subjective risk tolerance is moderately high. If F, go to #6.*
6. *Suppose that you are about to retire, and have two choices for a pension. Pension A gives you an income equal to your preretirement income. Pension G has a 50% chance your income will be double your preretirement income, and a 50% chance that your income will be half your preretirement income. You will have no other source of income during retirement, no chance of employment, and no other family income ever in the future. All incomes are after-tax. Which pension would you choose? If A, your subjective risk tolerance is very high. If G, your subjective risk tolerance is extremely high.*

Psychometric theory measures

In this study, the Grable and Lytton (1999) multi-dimensional risk tolerance scale was used as a proxy for a variety of psychometric risk tolerance measures. The choice of this scale was based on the open access nature of the scale and the questionnaire's wide use in previous research. The Grable and Lytton (1999) measure comprises 13 multiple choice items.[11] The scale was designed to be multi-dimensional and focused on measuring someone's willingness to make investments, comfort and experience taking risks, and disposition to taking speculative risk. The literature suggests that scale scores are positively associated with risky asset holdings and other consumer choices that entail risk (see Kuzniak et al. 2015; Rabbani et al. 2017).

The scale's Cronbach's alpha has ranged from .70 to .80 in most published studies. The highest reliability estimates tend to be associated with those with higher incomes and those who are older (see Kuzniak et al. 2015). Although certainly not the only (or even the best) risk tolerance scale available to researchers (e.g., Weber, Blais & Betz 2002), the Grable and Lytton scale (heretofore referred to as the propensity scale) has been used in numerous studies, with a general pattern of lower (higher) risk tolerance scores being associated with less (more) equity portfolio holdings and more conservative (aggressive) financial decisions (Kuzniak et al. 2015).

Participants were asked to answer a second measure of subjective risk tolerance. The single-item investment risk aversion measure from the Survey of Consumer Finances was included in the study. The question asks:

Which of the following statements below comes closest to the amount of financial risk that you are willing to take when you save or make investments?

1. *No financial risk.*
2. *Average financial risks expecting to earn average returns.*
3. *Above-average financial risks expecting to earn above-average returns.*
4. *Substantial financial risks expecting to earn substantial returns.*

[11] Examples of items asked in the scale include: (1) In general, how would your best friend describe you as a risk taker? a. A real gambler b. Willing to take risks after completing adequate research c. Cautious d. A real risk avoider; and (2) You are on a TV game show and can choose one of the following. Which would you take? a. $1,000 in cash b. A 50% chance at winning $5,000 c. A 25% chance at winning $10,000 d. A 5% chance at winning $100,000. The scale can be found at: http://pfp.missouri.edu/research_IRTA.html.

As coded, a high score represents low risk aversion (i.e., high risk tolerance), whereas a low score indicates high risk aversion (i.e., low risk tolerance). Scores were used to test the concurrent validity of the other risk tolerance measures.

A third measure of self-assessed financial risk tolerance was assessed by asking each participant to 'Rate yourself as a financial risk taker.' A 10-point scale was used, with 1 indicating the lowest level and 10 indicating the highest level.

Validity measures

Several questions were asked to gauge the degree of validity associated with the risk tolerance measures. Knowledge of casino gambling was assessed by asking, 'How knowledgeable are you about casino games?' A 10-point scale was used, with 1 indicating not at all knowledgeable and 10 indicating very knowledgeable. The likelihood of gambling was measured by asking, 'How likely is it that you would bet a day's income at the horse races?' Another 10-point scale was used, with 1 meaning very unlikely and 10 meaning very likely. Financial decision-making experience was assessed by asking, 'How much experience do you have making financial decisions?'. A 10-point scale was used, with 1 representing very little and 10 representing a lot. Participants were also asked to rate their investing knowledge on a 10-point scale with 1 indicating the lowest level and 10 indicating the highest level. Asset allocation data from each participant was assessed by asking, 'Suppose that you were to take a snapshot of your current financial position. Approximately what per cent of your total savings and investments are in the categories below?' Six asset categories were provided: (a) cash, (b) fixed-income, (c) equities, (d) business ownership, (e) real estate, (f) hard assets.

Risk tolerance scores from the measurements used in this study were also evaluated against five demographic characteristics. The gender of participants was assessed by asking each participant to self-identify as either male (coded 1) or female (coded 2). Approximately 60% of participants were female. Age was measured by asking each participant to indicate their age in years. Personal and household income was assessed using the following 11 categories: (1) none, (2) less than $US20,000, (3) $US20,001 to $US30,000, (4) $US30,001 to $US40,000, (5) $40,001 to $US50,000, (6) $US50,001 to $US60,000, (7) $US60,001 to $US70,000, (8) $US70,001 to $US80,000, (9) $US80,001 to $US90,000, (10) $US90,001 to $US100,000, and (11) above $US100,000. Median personal income fell between $US30,001 and $US40,000. Median household income fell between $US40,001 and $US50,000. Educational attainment was measured using the following six categories: (1) some high school or less, (2) high school graduate, (3) some college/trade/vocational training, (4) Associate's degree, (5) Bachelor's degree, and (6) graduate or professional degree. The majority of participants held at least a college degree level of education.

Analytical methods

The research questions were tested using a variety of statistical techniques. The core validity tests were assessed using non-parametric correlation coefficients. The use of non-parametric statistics was based on the categorical manner in which some of the variables were coded. ANOVA, *t*, and Kruskal-Wallis tests were used to evaluate the third research question, with each of the *t* and ANOVA tests using a bootstrap methodology. An answer to the fourth research question was based on a review of results from each of the tests.

Results

The first research question asked whether the measures based on economic theory and psychometric classical test theory exhibited concurrent validity. A preliminary answer to this question can be found in Table 1. Table 1 shows the mean and standard deviation for each of the risk tolerance measures used in this study, as well as the non-parametric correlation coefficients among the measures. One would expect that each of the measures should be correlated. This was generally the case. The Barsky and H&L tests were statistically related. Both measures were also positively correlated with the propensity scale. Curiously, however, H&L test scores were not correlated with scores from the SCF risk item.

Table 1: Descriptive and correlational data for the risk rolerance measures

	Range	Mean	S.D.	Barsky Corr.	H&L Corr.	Propensity Corr.	Self-Rated Risk Tolerance Corr.	SCF Risk Item Corr.
Barsky	1 – 4	2.20	0.88	1.00				
H&L	1 – 7	4.21	1.30	0.26**	1.00			
Propensity	13 - 44	26.10	4.66	0.21**	0.28**	1.00		
Self-Rated RT	1 – 10	4.11	2.19	0.18*	0.18**	0.42**	1.00	
SCF Risk Item	1 – 4	1.58	0.50	0.17*	0.14	0.44**	0.44**	1.00

Note: *p < .05 **p < .01

Results from tests undertaken to answer the first (concurrent validity) and second (convergent validity) research questions can be found in Table 2. Table 2 shows the mean and standard deviation for the gambling, gambling likelihood, financial experience, and investment knowledge questions, as well as the non-parametric correlation coefficients for each item linked to the Barsky, H&L, and propensity measures. Curiously, scores from the Barsky and H&L tests were not correlated with the casino gambling items. However, as suggested by Ahmad et al. (2011), Barsky and H&L test scores were correlated with investing knowledge, and for the Barsky test, a correlation with financial decision-making experience was noted. Scores on the propensity scale were correlated across the items.[12]

Table 2: Descriptive and correlational data for the knowledge, likelihood, and experience items

	Mean	S.D.	Barsky Corr.	H&L Corr.	Propensity Corr.
Casino Knowledge	4.13	2.63	0.11	0.10	0.26**
Gambling Likelihood	2.25	2.14	0.08	0.01	0.18**
Financial Decision Experience	5.70	2.22	0.16*	0.05	0.12*
Investing Knowledge	4.71	2.46	0.15*	0.17*	0.34**

Note: *p < .05 **p < .01

[12] Although not shown, self-rated risk tolerance and scores from the SCF risk item were statistically correlated with the items.

Table 3 provides a description of the portfolio allocation categories described by participants and the non-parametric correlation coefficients each asset class had with scores from the risk tolerance measures. Similar to the results shown in Table 2, neither the Barsky nor H&L tests were correlated to holdings in the asset classes. This is a curious finding in that theoretically the correlations should have been large and significant; however, given the relatively young average age of the sample, this may not be a surprise. It is possible participants did not understand the questions and/or held limited risky assets. The correlations were stronger for the propensity scale. Those who exhibited a higher tolerance for financial risk held less cash, more equities, and more hard assets (e.g., gold). No statistical significance was found between the propensity scale and fixed-income, business ownership, and real estate assets.

Table 3: Descriptive and correlational data for asset allocation items

	Mean	S.D.	Barsky Corr.	H&L Corr.	Propensity Corr.
Cash Assets	68.47	36.13	-0.01	-0.12	-0.23**
Fixed-Income	4.33	9.01	-0.05	0.01	0.07
Equities	16.99	27.58	0.01	0.08	0.20**
Business Ownership	2.53	9.90	0.01	-0.01	0.07
Real Estate	2.64	9.40	-0.03	-0.01	0.11
Hard Assets	5.07	13.39	0.05	0.05	0.14*

Note: *p < .05 **p < .01

The last test of convergent validity is shown in Table 4. Scores on the Barsky, H&L, and propensity measures were not found to be correlated with gender, age, personal or household income, or education. While not necessarily surprising given the homogenous nature of the participant sample, what is curious is that the direction of the correlation coefficients (although not statistically significant) varied across measures.

Table 4: Descriptive and correlational data for demographic items

	Mean	S.D.	Barsky Corr.	H&L Corr.	Propensity Corr.
Gender (1 = Male; 2 = Female)	1.60	0.49	-0.15	-0.15	-0.13
Age	25.94	6.48	0.06	0.05	-0.00
Personal Income	2.98	2.40	-0.01	0.01	0.04
Household Income	4.95	3.62	-0.06	0.03	0.08
Education	4.87	1.06	-0.04	-0.06	-0.00

Note: *p < .05 **p < .01

Data in Tables 1 through 4 represent cross-sectional data used to answer the first two research questions. The third research question asked if measures based on economic theory and psychometric classical test theory exhibit predictive validity to observed risk-taking behavior. In order to answer this question, data from the risk-taking game were evaluated.

Table 5 shows how well scores on the three risk tolerance measures predicted engagement in the risk-taking game. Of the 40 participants, 27 indicated a willingness to play, whereas 13 opted out of the game immediately. Excluding significance levels, scores from each measure were useful in predicting who would participate in the game, with higher scores predicting participation. However, when statistical significance was estimated, only scores from the propensity scale were predictive of game participation.

Table 5: Estimates of risk-taking by risk tolerance score

Would You Like to Participate in the Game?	N	Barsky M SD S.E. Mean	H&L M SD S.E. Mean	Propensity M SD S.E. Mean
Yes	27	2.26 0.90 0.17	4.22 1.48 0.28	26.52 5.13 0.99
No	13	2.08 0.64 0.18	4.15 1.07 0.30	22.92 3.71 1.03
t Test		$t_{38} = 0.65$, $p = .52$	$t_{38} = 0.15$, $p = .88$	$t_{38} = 2.26^{*}$, $p = .03$

Table 6 shows a more nuanced test of the predictive power of the risk tolerance scores. In this case, an analysis of variance test showed that propensity scale scores were useful in predicting three categories of participants: (1) those who opted out of the game, (2) those who opted in and chose the $US10 gamble, and (3) those who opted in and chose the $US20 gamble.[13] A post-hoc analysis using the Tukey post-hoc criterion for significance indicated that those who opted out of the game were similar to those who selected the $US10 gamble. Propensity scale scores for those who selected the $US20 gamble ($M = 27.69$, $SD = 4.90$) were significantly greater than the scores of those who opted out of the game ($M = 22.85$, $SD = 3.72$).

Table 6: Predictions of risk-taking by risk tolerance score

	N	Barsky M SD S.E. Mean	H&L M SD S.E. Mean	Propensity M SD S.E. Mean
No Gamble	13	2.15 0.69 0.19	4.15 1.07 0.30	22.85 3.72 1.03
$US10 Gamble	8	2.25 0.89 0.31	4.00 1.85 0.65	23.88 4.82 1.71
$US20 Gamble	19	2.21 0.92 0.21	4.32 1.34 0.31	27.69 4.90 1.12
F Test		$F_{37} = 0.04$, $p = .97$	$F_{37} = 0.30$, $p = .85$	$F_{37} = 4.94^{**}$, $p = .01$

[13] The sample size per group was based on a test power of .80, an effect size of .50, and a significance level of .05.

Another analysis of variance test (results not shown) indicated that propensity scale scores offered insights into the profile of participants who maintained their bet, compared to those who changed their bet, after learning about the true odds associated with the game. Three categories of risk-taking were examined: (1) those who opted out of the game, (2) those who retained their bet after hearing the odds, and (3) those who changed their bet after hearing the odds. A post-hoc analysis using the Tukey post-hoc criterion for significance indicated that those who opted out of the game were most similar to those who changed their bet and most dissimilar to those who retained their bet after hearing the true odds. In other words, participants with the highest risk tolerance scores on the propensity scale ($M = 27.48$, $SD = 5.39$) were more likely to participate in the game, and once a decision was made, they were the most likely to retain their choice. Neither Barsky nor H&L test scores were related to predicting the persistence of risk-taking choices.

Two robustness checks were made. First, a Kruskal-Wallis test was conducted to validate the risk-taking choice ANOVA findings (Table 6). Similar to an ANOVA analysis, the Kruskal-Wallis test is an omnibus test that indicates differences among groups. Because it is a non-parametric test, the analysis is based on median scores (Pett 2016). The results of the test matched the ANOVA findings. Specifically, neither Barsky nor H&L test scores were associated with risk choices. However, propensity scale scores were significant (χ^2 K-W = 7.20, $p = .03$). Post-hoc analyses using the Dunn procedure indicated that those who chose the highest risk choice had the highest propensity scale scores, with the highest risk group exhibiting a significantly different choice compared to those who opted out of the game. Second, an analysis of variance test was made using the variable titled, 'How knowledgeable are you about casino games' as a covariate in the model comparing (1) those who opted out of the game, (2) those who opted in and chose the $US10 gamble, and (3) those who opted in and chose the $US20 gamble. It was hypothesized that a participant's choice may have been influenced by the participant's familiarity with casino games, with those with little experience opting out of the game immediately. Even when accounting for this possibility, the core findings reported above were confirmed; propensity scale scores were predictive of risk-taking behavior ($F_{36} = 3.99$, $p = .02$), whereas Barsky and H&L test scores were statistically not significant.

Discussion

Results from the *t*, ANOVA, and Kruskal-Wallis tests were used to address the fourth research question, which asked what measurement approach (i.e., those based on economic utility theory or assessments based on classical test theory) provides the clearest insight into risk-taking tendencies? Before addressing this question directly, it is important to reconsider the purpose and sample of this study. Specifically, this study was designed to explore the relationships among the Barsky, H&L, and propensity measures with the intent of comparing and contrasting each method's predictive ability. In this regard, it was apparent that each measure offered unique advantages and disadvantages. The Barsky and H&L tests, for example, were very similar in terms of predictive power of participants' risk-taking behavior. Both measures were positively correlated with the propensity scale. This implies that these tests, based on economic theory, provide some degree of convergent validity across measures.

Curiously, the associations among the economic theory-based tests and the other measures of risk tolerance (i.e., the SCF risk question or self-assessed risk tolerance) were inconsistent, whereas scores from the scale developed using classical test theory were correlated consistently across measures. Similarly, only the psychometric scale was correlated with knowledge of casino games and the likelihood of gambling. None of the measures were found to be associated with participant demographic characteristics, which may have been due to the homogenous nature of the sample.

Of critical importance were the findings showing patterns of predictive accuracy among the risk tolerance measures. The psychometric scale was the only measure to predict who was more likely to participate in a game in which the participant was required to make a wager when the outcomes of the game were unknown and potentially negative. Similarly, scores from the psychometric scale were predictive of the level and persistence of risk-taking. Those who exhibited the highest scores on the psychometric scale were more likely to take the greatest risk in the risk-taking task, and when given the opportunity to change their wager, those with a high psychometric scale score were more likely to retain their bet. Robustness checks confirmed these findings. In matching tests, neither of the revealed-preference tests exhibited predictive ability to the level of the psychometric measure.

The results from this study provide support for what has often been reported in the literature. As noted by Frey et al. (2017), revealed-preference tests may be measuring situational characteristics rather than trait attributes (Buss 1989). The notion of having financial decision-makers choose between gambles or lotteries with probability outcomes known prior to the decision may not correspond to the cognitive demands placed on someone when a decision must be made when outcome probabilities are unknown. At a minimum, such assessments are likely to, as Charness et al. (2013) noted, cause noisy data. At their weakest, revealed-preference tests may not accurately capture a decision-maker's true preference for risk (i.e., risk appetite). Scores from a psychometric assessment appear to work more reliably because these tools are better able to account for a decision-maker's emotions (e.g., feelings of regret, fear, greed, and happiness). These tools appear to do a relatively good job of predicting future behavior (Dohmen et al. 2011; LÖnnqvist et al. 2015). When viewed holistically, the findings from this study provide support to the following comment made by Frey et al. (2017): '… measures from the propensity and behavioral measurement traditions cannot be used interchangeably to capture risk preference' (p. 8).

As such, a preliminary answer to the last research question—which measurement approach provides the clearest insight into risk-taking tendencies?—is that a questionnaire developed using principles from psychometric theory appears to offer greater validity when attempting to describe and predict financial risk-taking behavior, at least when compared across the measures examined in this study. The final choice of which assessment technique to use in practice should be based on the known strengths and weaknesses associated with each methodology. However, if the goal is to most accurately assess a decision-maker's willingness to engage in a risky financial behavior in which the outcome of the decision is both unknown and potentially negative, a psychometric assessment will likely provide more insights than a revealed-preference test.

Further research is needed to confirm this and the other results presented in this paper. In this regard, several limitations associated with the current study need to be acknowledged and addressed in future work. To begin with, the sample used in this study suffered from a potential

recruitment bias. It would be very useful to use a larger and more representative sample when replicating this study's methodology. Moving beyond an exploratory sample will provide additional insights into the usefulness of risk tolerance assessment tools. Additionally, future studies should attempt to align the use of incentives more closely with the demographic characteristics of participants. In this study, the maximum incentive was $US30, which was deemed to be meaningful to those who participated in the study. However, it is possible that some participants viewed the incentive as "found money," and as such, were willing to gamble even if this action was not in alignment with their stated or revealed risk tolerance. This possibility can be tested in future studies using larger incentives and categorizing future samples by income, net worth, financial knowledge, and similar characteristics.

To summarize, the results from this study are exploratory. This means that while the propensity scale, as a proxy for other questionnaires developed using principles from psychometric theory, showed the best concurrent, convergent, and predictive validity, this does not mean that tests based on economic theory are not valuable. It is possible that questionnaires based on economic theory work better at predicting behavior that does not involve a monetary risk (e.g., health, social, and physical risk-taking). Likewise, it is possible that economic theory tests and questionnaires, because of the types of questions asked, work best when administered to those with high cognitive ability. These and other potentialities need to be examined empirically before it will be possible to truly determine which measurement approach is the most valuable within the context of financial decision-making.

References

Adamowicz, W, Louviere, J & Williams, M 1994, 'Combining revealed and stated preferences methods for valuing environmental amenities', *Journal of Environmental Economics and Management*, vol. 26, pp. 271–292.

Ahmad, A, Safwan, N, Ali, M & Tabasum, A 2011, 'How demographic characteristics affect the perception of investors about financial risk tolerance', *Interdisciplinary Journal of Contemporary Research in Business*, vol. 3, no. 2, pp. 412–417.

ASIC 2011, *Review of financial advice industry practice*, Australian Securities & Investments Commission, Brisbane, QLD.

ASIC 2017, *Risk management systems of responsible entities*, Australian Securities & Investments Commission, Brisbane, QLD.

Barsky, RB, Juster, FT, Kimball, MS & Shapiro, MD 1997, 'Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study', *The Quarterly Journal of Economics*, vol. 112, pp. 537–579.

Beauchamp, JP, Cesarini, D & Johannesson, M 2017, 'The psychometric and empirical properties of measures of risk preference, *Journal of Risk and Uncertainty*, vol. 54, pp. 203–237.

Brand, M, Fujiwara, E, Borsutzky, S, Kalbe, E, Kessler, J & Markowitsch, HJ, 2005a, 'Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: associations with executive functions', *Neuropsychology*, vol. 19, pp. 267–277.

Buss, AH 1989, 'Personality as traits', *American Psychologist*, vol. 44, pp. 1378–1388.

Certified Financial Planner Board of Standards Inc 2018, *Code of Ethics and Standards of Conduct*, CFP Board Washington, DC.

Charness, G, Gneezy, U & Imas, A 2013, 'Experimental methods: eliciting risk preferences', *Journal of Economic Behavior & Organization*, vol. 87, pp. 43–51.

Corter, JE & Chen, YJ 2006, 'Do investment risk tolerance attitudes predict portfolio risk?', *Journal of Business and Psychology*, vol. 20, pp. 369–381.

Dave, C, Eckel, CC, Johnson, CA & Rojas, C 2010, 'Eliciting risk preferences: when is simple better? *Journal of Risk and Uncertainty*, vol. 41, pp. 219–243.

Dohmen, T, Falk, A, Huffman, D, Sunde, U, Schupp, JU & Wagner, GG 2011, 'Individual risk attitudes: measurement, determinants, and behavioral consequences', *Journal of the European Economic Association*, vol. 9, pp. 522–550.

Dow, J & Werlang, SRDC 1992, 'Uncertainty aversion, risk aversion, and the optimal choice of portfolio', *Econometrica*, vol. 60, pp. 197–204.

Eckel, CC & Grossman, PJ 2008, 'Forecasting risk attitudes: an experimental study using actual and forecast gamble choices', *Journal of Economic Behavior & Organization*, vol. 68, pp. 1–17.

Ellsberg, D 1961, 'Risk, ambiguity and savage axioms', *Quarterly Journal of Economics*, vol. 75, pp. 643–679.

Faff, R, Hallahan, T & McKenzie, M 2009, 'Nonlinear linkages between financial risk tolerance and demographic characteristics', *Applied Economics Letters*, vol. 16, pp. 1329–1332.

Fischhoff, B, Slovic, P, Lichtenstein, S, Read, S & Combs, B 1978, 'How safe is enough? A psychometric study of attitudes towards technological risks and benefits', *Policy Sciences*, vol. 9, pp. 127–152.

Frey, R, Pedroni, A, Mata, R, Rieskamp, J, & Hertwig, R 2017, 'Risk preference shares the psychometric structure of major psychological traits', *Science Advances*, vol. 3, DOI: 10.1126/sciadv.1701381.

Grable, JE & Lytton, RH 1999, 'Financial risk tolerance revisited: the development of a risk assessment instrument', *Financial Services Review*, vol. 8, no. 3, pp. 163–181.

Hanna, SD & Lindamood, S 2004, 'An improved measure of risk aversion', *Journal of Financial Counseling and Planning*, vol. 15, no. 2, pp. 27–45.

Hanna, SD, Gutter, MS & Fan, JX 2001, 'A measure of risk tolerance based on economic theory', *Journal of Financial Counseling and Planning*, vol. 12, no. 2, pp. 53–60.

Houthakker, HS 1950, 'Revealed preference and the utility function', *Economica, New Series*, vol. 17, pp. 159–174.

Joseph, AK, Richter, D, Samanez-Larkin, GR, Wagner, GG, Hertwig, R & Mata, R 2016, 'Stability and change in risk-taking propensity across the adult life span', *Journal of Personality and Social Psychology*, vol. 111, pp. 430–450.

Knight, FH 1921, *Risk, Uncertainty, and Profit*, Sentry Press, New York.

Kuzniak, S, Rabbani, A, Heo, W, Ruiz-Menjivar, J & Grable, JE 2015, 'The Grable and Lytton risk tolerance scale: a 15-year retrospective', *Financial Services Review*, vol. 24, pp. 177¬–192.

LÖnnqvist, J-E, Verkasalo, M, Walkowitz, G & Wichardt, PC 2015, 'Measuring individual risk attitudes in the lab: task or ask? An empirical comparison', *Journal of Economic Behavior & Organization*, vol. 119, pp. 254–266.

Moore, S 2018, 'Communication task difficulty in investment risk profiling: a linguistic perspective', *Financial Planning Research Journal*, vol. 4, no. 1, pp. 33–52.

Nunnally, J & Bernstein, L 1994, *Psychometric Theory*, McGraw-Hill, New York.

Pett, JA 2016, *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions* (2nd ed.), Sage, Los Angeles.

Rabbani, AG, Grable, JE, Heo W, Nobre, L & Kuzniak, S 2017, 'Stock market volatility and changes in financial risk tolerance during the Great Recession', *Journal of Financial Counseling and Planning*, vol. 28, no. 1, pp. 140–154.

Remler, DK, & Van Ryzin, GG 2015, *Research Methods in Practice* (2nd ed.), Sage, Los Angeles.

Reynaud, A & Couture, S 2012, 'Stability of risk preference measures: results from a field experiment on French farmers', *Theory and Decision*, vol. 73, pp. 203–221.

Richter, MK 1966, 'Revealed preference theory', *Econometrica*, vol. 34, pp. 635–645.

Samuelson, PA 1948, 'Consumption theory in terms of revealed preference', *Economica, New Series*, vol. 15, pp. 243–253.

Schiebener, J & Brand, M 2015, 'Decision making under objective risk conditions—a review of cognitive and emotional correlates, strategies, feedback processing, and external influences', *Neuropsychology Review*, vol. 25, no. 2, pp. 171–198.

Weber, EU, Blais, A-R & Betz, NE 2002, 'A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors', *Journal of Behavioral Decision Making*, vol. 15, pp. 263–290.

Weber, EU & Johnson, EJ 2009, 'Decisions under uncertainty: psychological, economic, and neuroeconomic explanations of risk preference', in Glimcher, PW, Camerer, CF, Fehr, E & Poldrack, RA (eds.) *Neuroeconomics: Decision Making and the Brain*, Elsevier, Oxford, UK, pp. 127–144.