Elicitation of risk preferences through satisficing

Kavitha Ranganathan [a,*], Tomás Lejarraga [b,c]

[a] *T A Pai Management Institute, India*
[b] *Universitat de les Illes Balears, Spain*
[c] *Max Planck Institute for Human Development, Germany*

Abstract

Financial institutions across the globe are now required to measure how much risk their clients are willing to accept. Despite its importance, there is no consensus on how to assess risk attitudes in providing adequate and legally compliant financial services. The standard approaches have been challenged. Recently, financial regulators have begun to focus on the worst possible scenario: Retail investors should be inquired about how much loss they are willing and able to bear. We examine the satisficing method, a recent approach that brings the worst possible scenario to the center of the risk preference assessment. This method involves asking participants to state the minimum returns they are willing to accept given a portfolio comprising a safe and a risky prospect. The stated minimum returns are a measure of risk preference. We observe that this measure correlates well with existing measures of risk preference, has high test–retest reliability while capturing high response variation, and predicts investments in stock or mutual funds.

1. Introduction

Since the 2008 financial crisis, financial regulators have tightened the requirements that financial institutions must follow to offer advice and protection of the retail investor. They are now required to measure how much risk their clients are willing to accept and, importantly, how much loss they are willing to bear in the worst possible scenario. It is not clear, however, how this assessment should be performed. In Europe, the Markets in Financial Instruments Directive (MiFID) requires financial institutions to measure client risk preference and offer products suitable to the client's risk profile, in particular "[t]he firm should specify the percentage of losses target clients should be able and willing to afford (for example, from minor losses to total loss)" (European Securities and Markets Authority, Guidelines on MiFID II, 2018, p. 6).

Similar measures were taken in the United States by the Financial Industry Regulatory Authority (FINRA), a non-governmental organization that regulates its member brokerage firms and exchange markets. One of FINRA's suitability rules requires firms to obtain information about customers' risk tolerance, defined as "ability and willingness to lose some or all of [the] original investment in exchange for greater potential returns" (2014, p. 4). The Securities and Exchange Board of India (SEBI) also requires investment advisers and sellers of financial services to look after investor interests. As MiFID and FINRA, SEBI centers the attention in the worst-case scenario: risk profiling tools should "assess client's capacity for absorbing loss" and "identify whether client is unwilling or unable to accept the risk of loss of capital". SEBI regulation goes further: "any questions or description in any questionnaires used to establish the risk a client is willing and able to take are fair, clear and not misleading, and should ensure that: (i) questionnaire is not vague or use double negatives or in a complex language that the client may not understand..." (SEBI, 2013, p. 10).

Two lessons can be drawn from current financial regulations on investor protection. First, financial institutions must assess risk preference to provide adequate and legally compliant financial services and advice. Second, in assessing risk preference, clients must bear in mind the consequences of a worst-case scenario, and report how much loss they are willing to accept in their investments.

In this paper, we examine the properties of a novel tool to elicit risk preference: the satisficing tool (hereafter SAT; Berg et al., 2018). SAT gives focal importance to the maximum loss an investor is willing to bear in her investment portfolio consisting of a risky and a riskless assets. We refer to this magnitude as the *worst-case* aspiration. The worst-case aspiration implies an upper bound on the best-case portfolio gains, resulting in an optimal

* Correspondence to: T A Pai Management Institute (TAPMI), Manipal, 576104, India.
E-mail address: kavitha.r@tapmi.edu.in (K. Ranganathan), tomas.lejarraga@uib.eu (T.Lejarraga).

asset allocation. SAT is grounded in Herbert Simon's (1955) idea of satisficing, according to which people choose options that surpass a minimum acceptable threshold.[1] Thus, with SAT we infer individual risk preference by inquiring about the worst possible consequence an investor is willing to bear. This approach meets the recommendations on risk profiling by financial regulators around the world.

Although there is a variety of methods available to measure risk preference, there is no consensus on which one is best. Such variety of methods can be broadly classified into largely unrelated measurement traditions of revealed preference and stated preference (Frey et al., 2017). The revealed-preference methods include incentivized tasks that entail a single choice between gambles (Binswanger, 1981; Eckel and Grossman, 2002; Dave et al., 2010); tasks framed as an investment decision (Gneezy and Potters, 1997; Charness and Gneezy, 2010; Berg et al., 2018); tasks framed as games, such as the Balloon Analogue Risk Task (Lejuez et al., 2002; Crosetto and Filippin, 2013), and more complex tasks where people make several decisions between pairs or sets of risky lotteries presented in a structured manner (Holt and Laury, 2002; Andersen et al., 2006). Stated-preference methods involve self-reports to questionnaires and harness people's introspective abilities (Weber et al., 2002; Dohmen et al., 2011).

We examine how risk measures from SAT compare with other three well-documented risk-elicitation methods from different traditions: the multiple-price list by Holt and Laury (2002, HL), the Eckel and Grossman (2002) method (EG), and a general self-report measure concerning participants' willingness to take financial risk (hereafter SR, Kaufmann et al., 2013). Specifically, in a within-subjects design, we examine SAT's relative performance in terms of consistency across elicitation methods (rank-order correlation with other risk-elicitation methods), test–retest reliability over four months, and its ability to predict whether or not people hold stocks or mutual funds. We also assess participants' confidence in their own responses in each risk-elicitation task.

Although scholars have examined and compared the properties of various risk elicitation tasks (Anderson and Mellor, 2009; Dave et al., 2010; Crosetto and Filippin, 2016); for an extensive discussion, see, Charness et al. (2013), the recently introduced SAT remains to be contrasted against the standard methods. Our overall assessment of the results indicates that SAT correlates well with EG and SR, two measures belonging to different traditions in the study of risk elicitation. In contrast, SAT does not correlate highly with HL. Low correlation between HL and other measures has also been observed by Frey et al. (2017). In addition, relative to other methods SAT measures are able to capture wider variation in risk preference without losing temporal stability SAT predicts better than other tools whether an individual participates in the stock market. Participants' confidence in their responses did not differ across risk elicitation methods or between the test and retest phases.

The paper is organized as follows. In Section 2, we briefly describe the literature on measurement traditions and various risk elicitation methods; Section 3 formally introduces the SAT task and its benchmarks: the HL, EG and self-report tasks. Sections 4 and 5 describes the comparison measures and the methodology employed. Section 6 reports the findings of the study, followed by a conclusion and discussion in Sections 7 and 8.

2. Risk preferences

In modeling economic behavior formally, researchers tend to assume that people are risk averse (Bernouilli, 1738). With the advent of behavioral economics, researchers have tried not to assume a general form of risk preference, but to examine preferences directly (Holt and Laury, 2002). To do so, researchers have adopted the *revealed-preference* approach (Samuelson, 1938), which assumes that people reveal what they truly prefer only when they make decisions with real consequences. Thus, individuals' risk preferences in this tradition are inferred from their choices among risky gambles with monetary consequences. Although several methods have been proposed to infer risk preference from consequential choices (Crosetto and Filippin, 2013; Gneezy and Potters, 1997; Lejuez et al., 2002), the most prevalent is HL (Holt and Laury, 2002), which is the second most highly-cited article in the American Economic Review since 2002 according to ISI's Web of Knowledge, and has been cited more than 6400 times according to Google Scholar (as of March 22, 2021). This method involves making 10 choices, each between two gambles with constant payoffs and varying probabilities. The resulting risk score is often used in experimental studies to measure and control for individual variation in risk preference (Masclet et al., 2009; Charness et al., 2018).

In spite of its prevalence, studies that have examined the efficacy of the HL method have raised questions about its validity. Anderson and Mellor (2009) showed that the risk measures inferred with HL have a low correlation with survey responses on the frequency of real risk taking, and Lönnqvist et al. (2015) found that HL correlates "at best only weakly" with self-report surveys. More recently, Ert and Haruvy (2017) examined the effect of repeating the HL elicitation method on a group of experimental participants. They observed that repeating HL changes risk scores, moving preferences from risk aversion to risk neutrality. Finally, in one of the most exhaustive studies on risk assessment comparing 39 methods on 1507 participants, Frey et al. (2017) observed that risk measures elicited with HL were uncorrelated with all self-report and frequency measures, and were only weakly correlated with other risk measures involving incentivized choices between gambles. Moreover, HL's test–retest reliability was among the lowest, a result that was consistent with Ert and Haruvy's (2017) observation that repeating HL does not guarantee similar scores. For Charness et al. (2018), "... the complexity and structure" of HL "can affect measured risk preferences" (p. 2).

A similar but simpler method was proposed by Eckel and Grossman (2002). Participants in EG make one choice between 6 gambles with equiprobable outcomes. Dave et al. (2010) argued that these gambles are more intuitive for people than gambles with varying probabilities, particularly for people with low mathematical skills (Charness et al., 2013). Although risk measures elicited with EG correlate significantly with those elicited through the other standard methods (Reynaud and Couture, 2012), its simplicity comes at the cost that EG is not able to distinguish between small differences in risk-seeking preference. We return to EG and describe it in more detail below.

In the *stated-preference* tradition, participants report their propensity using response scales. For example, the general risk item from the German Socioeconomic Panel asks "Are you generally a person who is willing to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means not at all willing to take risks and the value 10 means very willing to take risks". (Dohmen et al., 2011). People respond to this question by recounting their own behaviors and experiences, particularly those behaviors that were voluntary and consequential (Arslan et al., 2020). In contrast to the revealed-preference approach, responses in the stated preference tradition have no

---

[1] Artinger et al. (2021) provide a detailed review on Simon's original visions of satisficing and the interpretation that followed among the two research traditions in behavioral economics.

consequences for the responder. Thus, these responses are often considered "cheap talk" by proponents of the revealed-preference approach (see Camerer and Hogarth, 1999, for a discussion).

The relative strengths and weaknesses of these measurement traditions continue to be debated (Charness et al., 2013; Crosetto and Filippin, 2016; Friedman et al., 2014; Mata et al., 2018). Behavioral measures are easy to translate into numeric indices assuming specific utility functions. These indices, however, do not correlate highly with other measures in the revealed-preference tradition (Lönnqvist et al., 2015; Guan et al., 2020). In contrast, self-report measures are difficult to translate into numerical parameters, but they converge well with other stated-preference methods and exhibit temporal stability and predictive validity for real-world behavior (Szrek et al., 2012; Frey et al., 2017).

Against this background, Berg et al. (2018) proposed a novel method to elicit risk preferences based on Herbert Simon's concept of *satisficing*. According to Simon (1956), individuals do not trade off risk and reward optimally, maximizing the value of their options. Instead, they first set an aspiration level, then search through alternatives, and choose the first option that surpasses their aspiration level. Assuming this process, Berg et al. (2018) asked participants to invest an amount of money between a risky and a safe option, but instead of asking them to allocate their money between the two options, participants had to state the minimum outcome they could tolerate in the worst case. Setting their *worst-case aspiration* implies an allocation between the two options. A participant willing to minimize the possibility of losing under the worst scenario gives up high potential returns, because her portfolio would consist of a higher proportion of the safe option. In contrast, a participant willing to bear a large loss under the worst scenario embraces the possibility of even superior returns, because her portfolio would consist of a higher proportion of the risky option, exposing her to a broader range of possible returns.

To illustrate, consider a person who must allocate USD 1500 between a risk-free bond earning constant return of 1.12 and a risky investment with returns of either 0.48 or 2.40 with equal chance. A person with high aspirations, who would only tolerate losing USD 100 in the worst-case scenario—i.e., only willing to accept a final amount of USD 1400—would allocate USD 1062 to the risk-free bond and only USD 438 to the risky option. These high aspirations give her a chance to earn USD 2241. However, a person with lower aspirations, who could tolerate a potential loss of USD 400—accepting a final amount of USD 1100—would allocate only USD 594 to the risk-free bond and USD 906 to the risky asset, and take a chance at earning USD 2840. In this investment decision, people's worst-case aspirations reveal their willingness to take risks. Higher worst-case aspirations imply lower potential gains and reveal a lower willingness to take risks. Berg et al. (2018) showed "analytically and empirically that choosing a most preferred worst-case aspiration maps into a logically equivalent—but psychologically distinct—process of expected utility maximization" (p. 127). We describe the method formally in the next section.

SAT follows the design presented in Güth (2007) and further used in other studies (Güth et al., 2008; Fellner et al., 2009) in that only aspiration levels are inquired, avoiding having the participant ponder probabilities. While Güth's approach inquires participants about worst *and* best-case aspirations, SAT simplifies the approach by inquiring only about the worst-case aspiration. In referring to this approach, Brown and Sim (2009) state that, "One important advantage of this approach is that aspiration levels are often very natural for investors to specify, whereas traditional models based on risk measures or utility functions depend critically on tolerance parameters, which are often difficult for investors to intuitively grasp and even harder to appropriately assess (certainly, at least, relative to assessing aspiration levels)" (p. 71). In sum, SAT asks investors to state the loss they are willing to bear in the worst possible case. This simple task measures risk preference in a way that is consistent with retail-investor protection regulations in Europe, the US, and India.

3. Satisficing tool (SAT)

In SAT, participants are endowed with 1500 units of experimental currency (EC) to be allocated between a risk-free bond returning 12% (r = 1.12) and a risky gamble with equiprobable returns of +140% in a *high* state (h = 2.40) and −52% in a *low* state (l = 0.48) for one year.[2] The portion of the endowment e allocated to the risky asset is denoted i. Then participants are asked: "Choose the minimum acceptable worst-case value for your portfolio in the event the low-state return is realized". Their response is *worst-case aspiration*, denoted A_1, expressed in EC corresponding to Indian rupees, INR.[3] Participants enter A_1 within the admissible range for the worst-case portfolio outcomes[4] (i.e. the riskiest to safest portfolio outcomes range from 720 EC to 1680). Then, participants learn that "Choosing a value of A_1 determines the best possible portfolio value that can be achieved when the risky asset achieves the high outcome, and subsequently the amounts to be invested in the risky asset and the risk-free bond. Go ahead and experiment with different values and hit 'Proceed Further' when you are satisfied with your choice of worst-case portfolio value". Thus, based on A_1, the elicitation tool computes the *best-case aspiration,* denoted A_2, and the corresponding values invested in the risky asset (i) and risk-free bond ($e - i$). For example, specifying 1200 EC as A_1, the implicit portfolio is i = 750 EC in the risky asset and ($e - i$) = 750 EC in the risk-free bond, which implies that A_2 = 2640 EC (conditional on A_1). SAT automatically assigns the maximal A_2 conditional on A_1. This differs from Güth's (2007) approach, where participants set worst *and* best case aspirations, potentially resulting in inconsistent aspirations. Although similar methods have used allocation decisions between a risky and a safe option to measure risk preferences (e.g., Gneezy and Potters, 1997), SAT's key difference is the focal importance of worst-case aspiration (A_1).

Given the allocation in the risk-free and risky options, the standard mean–variance approach poses a trade-off between risk and expected return to derive the measure of risk preference. For example, an investor's optimal portfolio maximizes the utility function: $U = r_f + i[E(r_p) - r_f] - \frac{1}{2}Ai^2\sigma_p^2$, where U is the appropriate utility function (i.e. CARA and CRRA utility functions), r_f is the certain return from the risk-free option, $E(r_p)$ and σ_p^2 is mean and variance of the portfolio, and A is the coefficient of risk aversion (Bodie et al., 2014). Thus, given the allocation in the risk-free and risky options, the level of risk aversion can be inferred.

Berg et al. (2018) shows that the tradeoff between more favorable worst-case and best-case aspirations elicits a risk measure that is algebraically equivalent to the optimal capital allocation under the assumption of EU maximization. The elicited worst-case aspiration and implied upper bound on the high state portfolio return together produce an optimal portfolio that is a revealed preference for portfolios with lower standard deviations and expected values. Berg et al. (2018) argue that SAT is intuitive. "The

---

[2] The parameters of the gambles used in this example and in the experiment described hereafter differ from that parameters used in Berg et al. (2018), because we created a version of SAT that is comparable (in potential payoffs) to HL and EG.

[3] USD 1 = INR 21.253, OECD (2020).

[4] Note that the admissible range is not presented directly to participants. Instead, feedback is given stating "value out of range" before being prompted to re-enter a valid value.

tradeoff between worst-case and best-case aspirations is likely to be more salient because: (i) the currency units measuring levels of payoffs in the worst-case and best-case aspirations avoid the unfamiliar statistical concepts of mean and standard deviation; (ii) no weighted averaging (i.e., multiplying payoffs times probabilities) is required; and, perhaps most importantly, (iii) because the magnitude of the slope in the relationship between A_1 and $A_1 \mid A_2$ is substantially greater than for the linear tradeoff between expected return and reductions in standard deviation''. (p. 130). Participants with poor knowledge of probability and statistics may find it easier to tradeoff best-case and worst-case payoffs with a probability fixed at 0.5 than trading off standard deviations and expected values of lotteries.

The portfolio's risk weighting $\frac{i}{e} = \frac{A_2 - A_1}{(h-l)e}$ is the measure of risk preference. This risk measure can be interpreted as the proportion of the maximal best-to-worst case range ($he - le$) that the participant chooses as her portfolio's best-to-worst-case aspiration ($A_2 - A_1$). Therefore, the satisficing tool draws attention to the willingness to pay for risk reduction. Moreover, the participant's choice of A_1 can be interpreted in terms of mean and variance of the portfolio (Eqs. (4) and (5) in Berg et al., 2018, p. 130). Also, the aspiration set ($A_1, A_2 \mid A_1$) can be used to estimate risk preference parameters of individuals' utility functions, such as CARA or CRRA functions. The advantage of SAT is that the risk measure elicited is continuous (unlike HL and EG, which are categorical) which allows fine distinctions among participants who would otherwise be classified as falling in the same category. However, one limitation of SAT relative to HL and EG is that the instantiation of the tool examined here is unable to capture risk-seeking preferences.[5]

3.1. Benchmarks

3.1.1. Holt and Laury price list (HL)

In HL, participants see 10 pairs of gambles on a computer screen and choose their preferred gamble from each pair (Table 1). Both gambles offer two possible outcomes. The gamble pairs are constructed such that the low outcome in Gamble A is always higher than the low outcome Gamble B. Specifically, in all pairs, A offers 2400 EC or 1920 EC, whereas B offers 4620 EC or 120 EC. To make the task meaningful to Indian participants, and for consistency with SAT and EG, we converted the original HL payoffs to experimental currency (15 EC = 1 Indian Rupee) and multiplied by 1200. In each pair of gambles, A is safer than B, because the range of outcomes in B is wider. The critical feature of HL is that gamble pairs differ in their probabilities. In the first pair, both in A and B, the probability of the high outcome is 1/10, and the corresponding probability of the low outcome is 9/10. Across the list of 10 pairs, the probability of the high payoff increases by 1/10 with each gamble. Choosing B is risky when the probability of the high outcome is low, in the first few pairs of gambles, but as the probabilities gradually change, B becomes more attractive, as its expected value increases relative to A. After the fifth pair, continuing to choose A reflects risk aversion.

Assuming constant relative risk aversion (CRRA) $U(x) = \frac{x^{1-r}}{1-r}$, Holt and Laury (2002) classify risk preferences depending on the number of safe (A) choices, with $r = 0$ exhibiting risk neutrality, $r < 0$ exhibiting risk seeking, and $r > 0$ exhibiting risk aversion. This classification remains constant with our transformed payoffs. The switch point determines the number of safe choices and the corresponding risk aversion parameter range. The result of HL is that participants are classified into one of 10 risk categories.

One advantage of HL over SAT is that HL is able to capture risk-seeking preferences, while SAT is restricted to risk aversion and risk neutrality. One downside of HL, however, is that it provides a categorical measure of risk preferences (Andersen et al., 2006), which limits its potential to distinguish between participants whose risk preferences differ only slightly. SAT provides a continuous measure, which allows it to capture wider variation in risk preferences, although constrained within the domain of risk aversion. Another disadvantage of HL is that participants can—and often do—make inconsistent choices, by switching from A to B and back to A. This inconsistency complicates the inference of risk preferences, but it serves to identify participants who may not have understood the task. The standard procedure is to exclude these participants from the analysis (Dave et al., 2010; Charness et al., 2013; Crosetto and Filippin, 2016),[6] as we do here.

Finally, HL differs from SAT in another important aspect. Probabilities are central in HL, as they are only aspect that changes down the gamble list. In SAT, however, probabilities play no central role, as they remain constant at 0.5 across different options. Much evidence from the behavioral sciences has highlighted people's poor ability in processing probabilities (Tversky and Kahneman, 1974) and that payoffs easier to process (Dave et al., 2010; Charness et al., 2013). This difference makes SAT potentially easier to complete than HL.

3.1.2. Eckel and Grossman price list (EG)

In EG, participants see six gambles and choose the one that they prefer. All gambles involve a 50/50 chance of a low or high outcomes. Gamble 1 is safe, as the two outcomes offer the same amount. Gambles 2 to 5 increase in expected return and risk, with expected return increasing linearly with standard deviation. Gamble 6 involves only an increase in risk from Gamble 5, keeping the expected return constant. To be able to compare EG with other measures, we multiplied Dave et al.'s (2010) payoffs by 60 (Table 2). Similarly, the range of payoffs in our EG and HL is comparable to that of the SAT task of from 720 EC to 3600 EC. As HL, EG is captures risk-seeking preferences, while SAT is restricted to risk averse and risk-neutral preferences. Each choice in EG corresponds with the range of coefficients of relative risk aversion, under the assumption of CRRA, shown as 'r' in Table 2. Since we scale the payoffs of Dave et al. (2010), the implied range for the CRRA parameters remains constant.

The EG task provides a coarser classification of risk preference relative to HL and SAT, as participants are classified into six categories, ranging from most risk-averse to risk-neutral to risk-seeking preferences. The advantage of EG over HL is that EG involves fewer gambles to consider and a single choice to make, sidestepping the problem of inconsistent choices present in HL. Arguably, EG's gambles are simpler than HL's, because probabilities in EG are constant across options with varying payoffs, while in HL payoffs are constant across options with varying probabilities.

3.1.3. Self-Report measure (SR)

We use the self-reported risk measure from Kaufmann et al. (2013). We ask participants: "Please estimate your willingness to take financial risk (1 = not willing to accept any risk; 5 = willing to accept substantial risk to potentially earn a greater return)".

---

[5] A tool that captures risk-seeking preferences is possible in principle but would be counterintuitive given the built-in assumption in the SAT tool that risk and reward are positively correlated.

[6] An alternative to excluding subjects with inconsistent choices is to assume stochasticity in responses while estimating the parameters of interest (Harrison and Rutström, 2008; Jacobson and Petrie, 2009). To avoid the possibility of inconsistencies, some experimenters enforce unique switching points in the experimental software (Tanaka et al., 2010). Arguably, excluding inconsistent choices is more conservative than modeling noise, because only participants who are left in the analysis can be thought to have expressed their true preferences (Andersen et al., 2006; Dave et al., 2010; Charness et al., 2013).

Table 1
Ten Pairs of Gambles from Holt and Laury (2002) Adapted to Experimental Currency (15 EC = 1 INR).

Option A (in EC)	Option B (in EC)	EV difference	Safe choices	Implied CRRA range	CRRA code
1/10 of 2400, 9/10 of 1920	1/10 of 4620, 9/10 of 120	1398	0	$r < -1.71$	RL4
2/10 of 2400, 8/10 of 1920	2/10 of 4620, 8/10 of 120	996	1	$-1.71 < r < -0.95$	RL3
3/10 of 2400, 7/10 of 1920	3/10 of 4620, 7/10 of 120	594	2	$-0.95 < r < -0.49$	RL2
4/10 of 2400, 6/10 of 1920	4/10 of 4620, 6/10 of 120	192	3	$-0.49 < r < -0.14$	RL1
5/10 of 2400, 5/10 of 1920	5/10 of 4620, 5/10 of 120	−210	4	$-0.14 < r < 0.15$	RN
6/10 of 2400, 4/10 of 1920	6/10 of 4620, 4/10 of 120	−612	5	$0.15 < r < 0.41$	RA1
7/10 of 2400, 3/10 of 1920	7/10 of 4620, 3/10 of 120	−1014	6	$0.41 < r < 0.68$	RA2
8/10 of $2400, 2/10 of 1920	8/10 of 4620, 2/10 of 120	−1416	7	$0.68 < r < 0.97$	RA3
9/10 of 2400, 1/10 of 1920	9/10 of 4620, 1/10 of 120	−1818	8	$0.97 < r < 1.37$	RA4
10/10 of 2400, 0/10 of 1920	10/10 of 4620, 0/10 of 120	−2220	9–10	$1.37 < r$	RA5

Table 2
Six Gambles from Eckel and Grossman (2002, 2008) Adapted to Experimental Currency (15 EC = 1 INR).

Choice	Roll	Payoff (EC)	Expected return	Standard deviation	Implied CRRA range	CRRA code
Gamble 1	Low	1820	1820	0	$3.46 < r$	RA1
	High	1820				
Gamble 2	Low	1560	1950	390	$1.16 < r < 3.46$	RA2
	High	2340				
Gamble 3	Low	1300	2080	780	$0.71 < r < 1.16$	RA3
	High	2860				
Gamble 4	Low	1040	2210	1170	$0.50 < r < 0.71$	RA4
	High	3380				
Gamble 5	Low	780	2340	1560	$0 < r < 0.50$	RN
	High	3900				
Gamble 6	Low	130	2340	2210	$r < 0$	RL
	High	4550				

Risk preferences stated with this measure are consistent with the literature (Nosić and Weber, 2010; Van Rooij et al., 2011) and are good predictors of risk taking in the context of investing (Kaufmann et al., 2013). SR provides five categories of risk preference, ranging from most risk-averse to most risk-seeking.

4. Comparison of risk measures

Consistency across elicitation methods. We examine to what extent measurements of risk preferences are consistent across methods in terms of rank-order correlation of individual measurements. Although a person's absolute risk preferences may change across elicitation methods, that person's position relative to others may remain constant.

Temporal stability. Although risk preference can change across an individual's life span (Mata et al., 2016) and be influenced by specific life events such as experiencing an economic shock (Lejarraga et al., 2016) or a natural disaster (Cameron and Shah, 2015), useful measures of risk preference should be temporally stable: a measurement taken today should not differ substantially from a measure taken in the near future. We compute the rank-order correlation between participants' risk scores in a first measurement and a second measurement four months later.

Predictive Validity. We examine how measures of risk propensity elicited in the laboratory relate to financial risk taking outside the laboratory. Specifically, we fit a generalized linear model (logistic regressions) for each risk-elicitation task, using risk scores as the main independent variable to explain participants' *yes* or *no* answer to the question on whether they ''owned stocks or mutual funds''.

Confidence. Finally, we examine participant self-reports of confidence in their own responses for each risk-elicitation task. Convergence and temporal stability could be facilitated by more confident responses, whereas less confident responses could promote temporally unstable measurements.

5. Method

The goal of this study is to examine SAT against HL, EG, and SR in terms of consistency, temporal stability, predictive validity, and confidence. This study involved a test phase and a retest phase, both conducted in an experimental laboratory.[7] In the test phase, participants responded to four risk-elicitation tasks, namely, HL, EG, and SAT presented in random order, followed by SR. Subsequently, participants reported their age, gender, academic qualifications, and marital status. Four months after the test phase, participants returned to the laboratory to complete the retest phase. They again completed the four risk-elicitation methods in a new random order.

---

[7] Data and experimental materials are available at https://osf.io/4gqub/?view_only=07c95ab233e54d339f4800a017c1884f and http://www.riskitude.com.

Table 3
Summary of sample size across risk elicitation tasks and demographics.

	Test (N = 112)	Retest (N = 103)	Test-retest (N = 103)
	N valid	N valid	N valid
SAT-range	112	103	103
SAT-CARA	111	103	103
SAT-CRRA	111	103	103
EG	112	103	103
HL	94	93	82
SR	112	103	103
SAT-confidence	112	103	–
EG-confidence	112	103	–
HL-confidence	94	93	–

5.1. Participants

Participants were recruited via e-mail after ethics approval from the ethics committee of T A Pai Management Institute (TAPMI). E-mails were sent to approximately 300 people. A hundred and twelve people took part in the study: 69 were students pursuing the MBA program at TAPMI and 43 were staff working in clerical positions at the business school, including employees at the department of human resources, computer maintenance, general administration, and cleaning staff. Some participants were excluded from the analysis: One participant gave a boundary response in SAT, leading to underdetermined risk aversion scores in the SAT-CARA and SAT-CRRA, and 18 participants (16%) made inconsistent choices in HL, because they switched more than once or switched in the incorrect direction (from Gamble B to A). The proportion of inconsistent choices is similar to that found in other studies.[8] Hence, the final sample of participants for the test phase is different across elicitation methods, and detailed in Table 3.

For the retest phase of the experiment, an invitation e-mail was sent again to all 112 participants from the test phase. Four MBA students and five staff did not return to the laboratory to complete the second part of the study. Hence, a total of 103 participants were analyzed in the test–retest analysis for SAT, EG, and SR tasks. Out of 103 participants in the retest phase, only 93 participants completed the HL task consistently. Therefore, the test–retest analysis includes 103 participants for all tasks except HL, which includes 82 subjects. Thus, 9 participants were inconsistent in either the test or retest phases. Table 3 summarizes sample size at each stage of the analysis, and Table 4 describes participant demographics.

Participants who completed all study phases were 60% male. Most participants were younger than 30 years old (82%), with participants between 31 and 40 years amounting to 9%, and participants between 41 and 50 years old also amounting to 9%. Only 4% were older than 51. Most participants were undergraduates (71%), followed by post-graduates (24%), and a few who only completed high-school (3%) or had higher professional degrees (3%). Participants were predominantly single (74%). Around 54% of participants reported not owning any financial assets, such as mutual funds or stocks.

5.2. Procedure

An invitation e-mail was sent to both groups of participants, and individuals from each group volunteered to report at a specified time and location. Consent forms were distributed, signed, and collected before the experimental session began. Participants were directed to independent cubicles in the computer laboratory, and received written instructions on their computer screen. Instructions included information about the project, the time commitment, compensation scheme, and confidentiality of responses. The instructions were also read aloud in vernacular language (Kannada) for the benefit of the staff group. Participants completed the three risk-elicitation methods (HL, EG, and SAT) in random order. The SR question was presented at the end, together with the demographic questions.[9]

5.3. Incentives

Participants were informed that each task involved "experimental currency", with 15 EC = 1 INR. Participants were endowed with 1500 EC to perform the SAT task. The range of outcomes (in EC) across EG, HL, and SAT were made comparable.[10]

6. Results

We first describe how the individual risk measures were computed in each risk-elicitation method. We then report the relative performance of each method in terms of consistency across risk-elicitation methods, temporal stability, predictive validity, and confidence.

6.1. Risk measures

SAT offers three measures of risk: SAT-range is a risk measure based on the portfolio's risk weighting $(\frac{i}{e})$, interpreted as the proportion of the participant's choice of best-to-worst-case aspiration $(A_2 - A_1)$ to the maximal best-to-worst case range of portfolio outcomes $(he - le)$. SAT-range captures only risk-averse to risk-neutral preferences. We transform the SAT variable such that it can be interpreted in the same manner as other measures. Hence, higher SAT-Range indicates greater risk aversion, and the value of 0 can be interpreted as risk-neutral as shown in Eq. (1).

$$\left(1 - \frac{i}{e}\right) = 1 - \frac{A_2 - A_1}{(h-l)e} \tag{1}$$

Two other measures can be derived from SAT assuming an expected-utility approach. The participant's choice of A_1 can be

---

[8] Other studies report similar levels of inconsistent choices: 17.1% in Filippin and Crosetto (2016) using a sample of 6315 subjects in 54 published papers; 8.5% in Dave et al. (2010); 22% in Crosetto and Filippin (2016). The proportion of inconsistent choices is even higher in samples from developing countries: 55% among Rwandan adults (Jacobson and Petrie, 2009).

[9] The reason for this methodological choice is that the self-report question had no implications on the final payment for participants, so we decided to keep the three incentivized tasks together for practical purposes.

[10] Participants were instructed that their compensation depended on their choices. In SAT, it was explained to participants that, depending on whether the low-state or high state occurred (by flipping a fair coin), the corresponding portfolio outcome would be paid out. We used EC because it allows us to work with high and round amounts and can be converted to other currencies without altering the experimental materials. In EG, the earnings would depend on one of six gambles that the participant selected during the task. A fair coin would then be tossed to determine whether the low or high state payout would be operational in the selected gamble. For HL, the description of the compensation scheme read: "To determine your earnings we will first use a ten-sided dice, whose faces are numbered from 1 to 10. After you have made all the choices, we will roll the ten-sided dice to select one of the ten gambles to be used. For the second time, we will roll a two-sided coin to determine what your payoff is for the option you chose, Gamble A or Gamble B, for the particular gamble selected in the first round". It was made clear that one of these tasks would be selected for the final payout, and the outcome in one task would be independent of the outcome in another task: "At the end of the entire experiment, one of the three tasks will be randomly picked and based upon your gain or loss for that task, you will be paid that amount of money in INR (after converting the experimental currency)". The average payout across both groups in the test and retest phases was INR 150 and INR 130, and everyone was paid the amount of money they earned, privately, in cash. In terms of Purchasing Power Parity, these amounts translate to USD 7,05 and USD 6,11, respectively (according to an exchange rate of USD 1 = INR 21.253, OCDE, 2020).

Table 4
Socio-demographic description of the participants.

		Test	Retest
Group	MBA	69 (62%)	60 (58%)
	Staff	43 (38%)	43 (42%)
Gender	Male	68 (60%)	64 (62%)
	Female	44 (40%)	39 (38%)
Age	Below 30 years	88 (78%)	85 (82%)
	31 to 40 years	10 (9%)	8 (8%)
	41 to 50 years	10 (9%)	7 (7%)
	51 to 60 years	4 (4%)	3 (3%)
Education	School Final	3 (3%)	1 (0.00%)
	Under-graduate	80 (71%)	71 (71%)
	Post-graduate	26 (23%)	25 (24%)
	Professional degree (CA, ICWA, CFA etc.)	3 (3%)	3 (3%)
Marital status	Married	29 (26%)	27 (26%)
	Single	83 (74%)	76 (74%)
Ownership of financial assets	Yes, I own mutual funds or stocks currently	52 (46%)	55 (53%)
	No, I do not own mutual funds or stocks currently	60 (54%)	48 (47%)

mapped to a CARA utility function defined as $u(x) = 1 - e^{-kx}$ where x denotes wealth and k is the coefficient of absolute risk aversion. Expected utility for the participant's choice of aspiration set $(A_1, A_2 \mid A_1)$ can be defined as, $u(i) = p[1 - e^{-kA_1}] + (1 - p)[1 - e^{-kA_2}]$. Assuming the first-order condition is satisfied, in Eq. (2) we compute the value of the coefficient of absolute risk aversion, k. In case of SAT-CARA, higher k indicates higher risk aversion.

$$k = \left\{ \frac{\log[(1-p)(h-r)] - \log[(r-l)p]}{(h-l)(er - A_1)} \right\} (r - l) \tag{2}$$

Similarly, if we assume a CRRA utility function, then $u(x) = x^\alpha$, $\alpha > 0$. Expected utility for the participant's choice of aspiration set $(A_1, A_2 \mid A_1)$ can be defined as $u(i) = (1 - p)(A_2)^\alpha + p(A_1)^\alpha = E[u(iX + (e - i)r)]$. Assuming the first-order condition for A_1 holds and solving for α provides the coefficient of relative risk aversion, then higher values of $(1 - \alpha_i)$ imply higher risk aversion, as in in Eq. (3).[11]

$$(1 - \alpha_i) = \frac{\log\left\{ \left[\frac{h-r}{r-l}\right]\left[\frac{1-p}{p}\right] \right\}}{\log\left[\frac{A_2}{A_1}\right]} \tag{3}$$

We compute a SAT-range, SAT-CARA, and SAT-CRRA for each individual in our sample, both in the test and retest phases.

The derivation of a risk-aversion measure from the HL task assumes CRRA, defined as $u(x) = x^{1-r}$, and provides 10 categories of risk preference depending on the switch point and the number of safe options the participant chooses. We use the corresponding risk aversion parameter range to classify the participant in one of the 10 risk categories, where higher r values imply higher risk aversion.

Similarly, the EG task provides six categorical measures of risk, and each participant's choice of gamble corresponds with the range of coefficients of relative risk aversion, r. Here again, higher r values imply higher risk aversion.

Finally, the SR measure provides five levels of risk preference, ranging from most risk-averse (1) to most risk-seeking (5). For consistency with other measures, we transform the SR score (by computing 6 minus the individual response) to interpret higher scores as indicative of higher risk aversion.
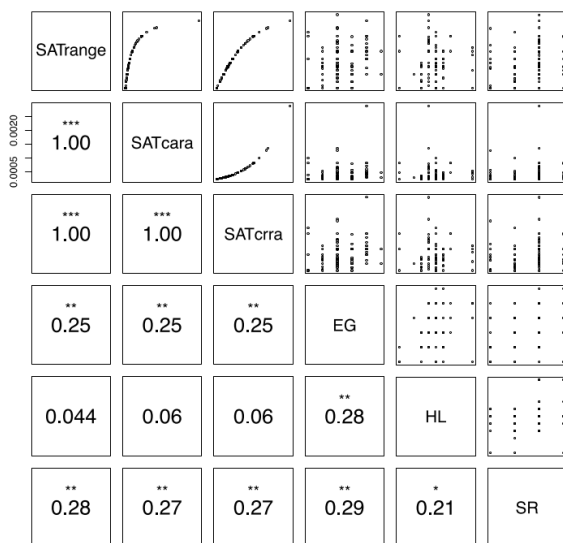

Fig. 1. Relationship of risk measures in the test phase. The lower section of the matrix indicates Spearman's rank correlation between measures, with significance indicated as "***" for $p < .001$, "**" for $p < .01$, "*" for $p < .05$, and "." for $p < .1$.

6.2. Consistency

To examine how consistent different elicitation methods are, we compute Spearman's rank-order correlations across measures. Fig. 1 shows that higher scores in the three SAT measures (SATRange, SAT-CARA, and SAT-CRRA) are strongly and significantly correlated with higher scores in EG (Spearman's $\rho = 0.25$, ($p = .01$), 0.25 ($p = .01$), 0.25 ($p = .01$) respectively) and SR (Spearman's $\rho = 0.28$ ($p = .00$), 0.27 ($p = .00$), 0.27 ($p = .00$) respectively), but not with HL (Spearman's $\rho = 0.04$ ($p = .67$), 0.06 ($p = .57$), 0.06 ($p = .57$) respectively). Also, EG correlates significantly with HL and SR (Spearman's $\rho = 0.28$ ($p = .01$), 0.29 ($p = .00$) respectively), and SR does not correlate with HL (Spearman's $\rho = 0.21$ with $p = .04$).

In the retest phase (Fig. 2), the directions of association between risk measures remain unchanged, but the correlations increased markedly; SAT-Range, SAT-CARA and SAT-CRRA show significant correlations with G (Spearman's $\rho = 0.41$ ($p = .00$) 0.41 ($p = .00$), 0.41 ($p = .00$) and SR (Spearman's $\rho = 0.34$ ($p = .00$), 0.34 ($p = .00$), 0.34 ($p = .00$), respectively). Similar to the

---

[11] Using CARA and CRRA assumptions introduces another limitation to the SAT measure. Risk scores using these functions are not well-behaved at the boundary. For example, if a participant reports a worst-case aspiration (A_1) of 1680 EC, this would entail the entire endowment of 1500 EC being invested in the risk-free bond and 0 in the risky asset. Therefore, in both the CARA and CRRA risk measures, the denominator tends to 0.
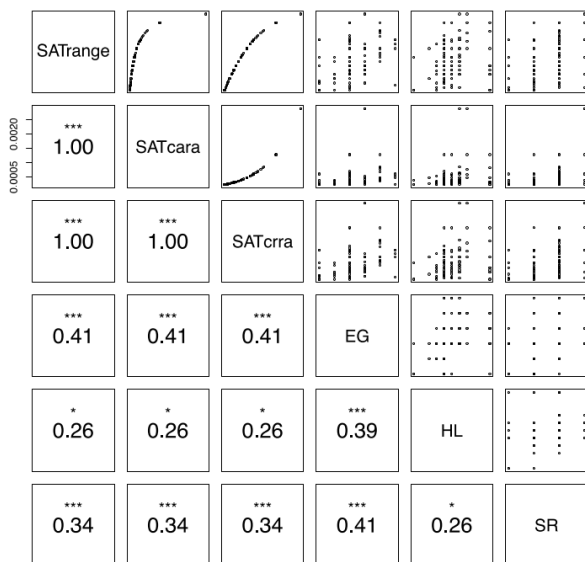
test phase, SAT and SR measures show weakest association with HL (Spearman's $\rho = 0.26$ ($p = .01$), 0.26 ($p = .01$), 0.26 ($p = .01$) for SAT-Range, SAT-CARA and CRRA, respectively, and Spearman's $\rho = 0.26$ ($p = .01$)). Why did correlations increase? At least three explanations are possible. SAT measures could have converged to the other measures; the other measures could have converged to SAT measures; or both measures have changed in the direction of each other. We return to this puzzle after discussing the temporal stability of the measures.

In sum, SAT scores converge well with EG and SR, two measures derived from two different traditions in the study of risk elicitation. In contrast, HL does not show significant correlation with SAT measures or SR, an observation consistent with Frey et al. (2017). Our findings conform to results found in the literature that measures of risk preference are positively and statistically significantly correlated, yet in most cases only to a low degree (Deck et al., 2013; Crosetto and Filippin, 2016; Menkhoff and Sakha, 2017).

6.3. Temporal stability

We assess temporal stability by computing Spearman's rankorder correlation between the scores in the test and retest phases. The sample subject to analysis consists of 103 participants for SAT, EG, and SR tasks, and 82 participants for the HL task, who returned for the retest phase of the experiment at an interval of four months. Fig. 3 shows, on the x-axis, the test–retest reliability of the risk scores as measured by Spearman's rank-correlation coefficients: SAT-Range, SAT-CARA, SAT-CRRA, and SR have the highest temporal stability (Spearman's $\rho = 0.36$ ($p = .00$), 0.35 ($p = .00$), 0.35 ($p = .00$) and 0.33 ($p = .00$), respectively), followed by EG ($\rho = 0.24$ ($p = .02$). In contrast, HL does not exhibit significant rank-order stability ($\rho = 0.20$ with $p = 0.07$).
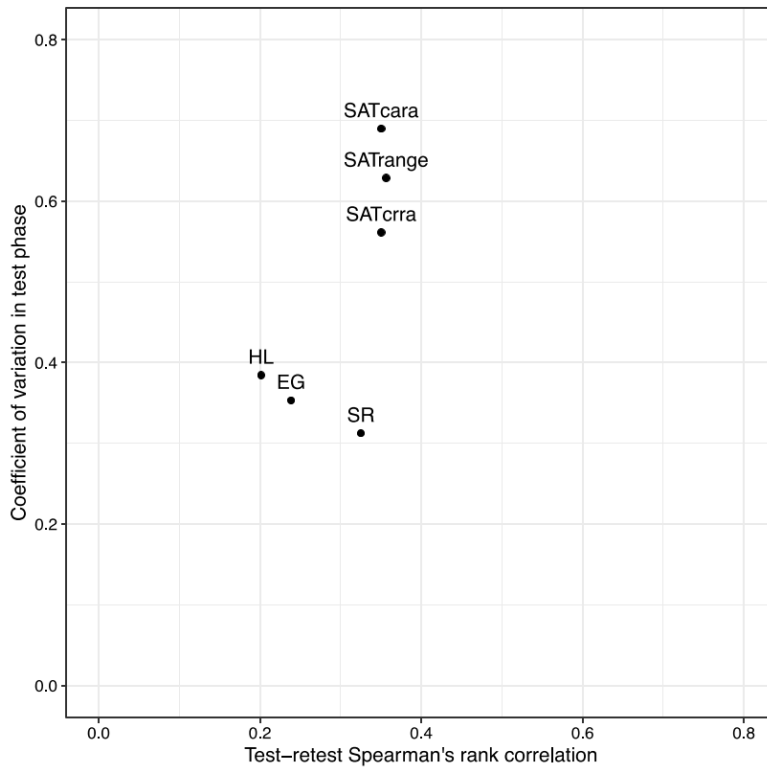
Fig. 2. Relationship of risk measures in the retest phase. The lower section of the matrix indicates Spearman's rank correlation between measures, with significance indicated as "***" for $p <$ .001, "**" for $p < .01$, "*" for $p < .05$, and "." for $p < .1$.

Comparing test–retest reliability across continuous and categorical measurements, as well as across categorical measurements with different numbers of categories, is potentially misleading because a ranking of just a few categories (such as EG and HL) is less likely to change across measurements than a ranking of multiple categories or continuous measurements (such as SAT). Therefore, Fig. 3 plots, on the y-axis, the coefficient of variation of each measure—computed as the ratio of standard deviation across participants and the mean in the test phase—such that test–retest reliability can be interpreted in relation to the individual variation that each measure is able to capture. Useful measures should achieve high test–retest reliability while being able to capture high levels of variation—i.e., be located in upper-right corner of Fig. 3.

The three SAT measures (SAT-Range, SAT-CARA, and SATCRRA) show the highest temporal stability. Importantly, SAT measures also show the highest coefficient of variation compared to the other measures. These results suggest that the SAT measures are able to capture wider variation in risk preference than the other measures examined, and do so without losing temporal stability. Consistent with Frey et al. (2017), SR tends to be highly stable across measurements, but this seems facilitated by its poor potential to capture individual variation. HL and EG show the lowest temporal stability. One potential reason for why HL stands out as having low temporal stability—and low convergence with other risk measures—is that HL is the only measure with strong emphasis on probabilities, and people's ability in processing probabilities is known to be poor (Charness et al., 2013; Dave et al., 2010; Tversky and Kahneman, 1974).

The temporal stability of the measures sheds some light on the increased correlations between measures from the test phase to the retest phase. Results indicate that SAT measures have the highest temporal stability, suggesting that the convergence across measures was not caused by changes in the SAT measurements, but more likely caused by changes in measurements from EG and HL. Indeed, previous evidence suggests that HL measurements change with experience, and in particular, change toward risk-neutrality (Ert and Haruvy, 2017), potentially indicating that second measurements are more reliable than first ones.

6.4. Predictive validity

We examine to what extent these measures of risk preference predict whether or not participants take financial risk. Specifically, we examine the relation between risk measures and whether or not individuals own financial assets, such as stocks or mutual funds. The prediction is that higher levels of risk aversion should be associated with lower probability of owning financial assets. Indeed, in the test phase, we find that the three SAT scores and SR scores predict ownership of financial assets significantly (Fig. 4), controlling for age, gender, academic qualifications, and marital status, as indicated by the 95% confidence intervals of the estimates that do not include 0. In contrast, EG and HL do not show significant correlations with ownership of financial assets. Among the control measures, being male was significantly associated with a lower probability of owning financial assets across elicitation methods, except in SR.

6.5. Confidence

Fig. 5 shows individual confidence in each risk-elicitation task in the test and retest phase. The means and corresponding 95% confidence intervals show that confidence was constant across risk-elicitation methods and in the test and retest.

7. Conclusion

Despite the importance of assessing risk preference in providing adequate and legally compliant financial services, there is no consensus on how to do it. Empirical evidence suggests that standard methods do not converge well with one another or with real-world behavior, and have poor temporal stability. Financial regulators and risk management practitioners are now advised

Fig. 3. Coefficient of variation in the test phase as a function of test–retest reliability measured as Spearman's rank correlation.
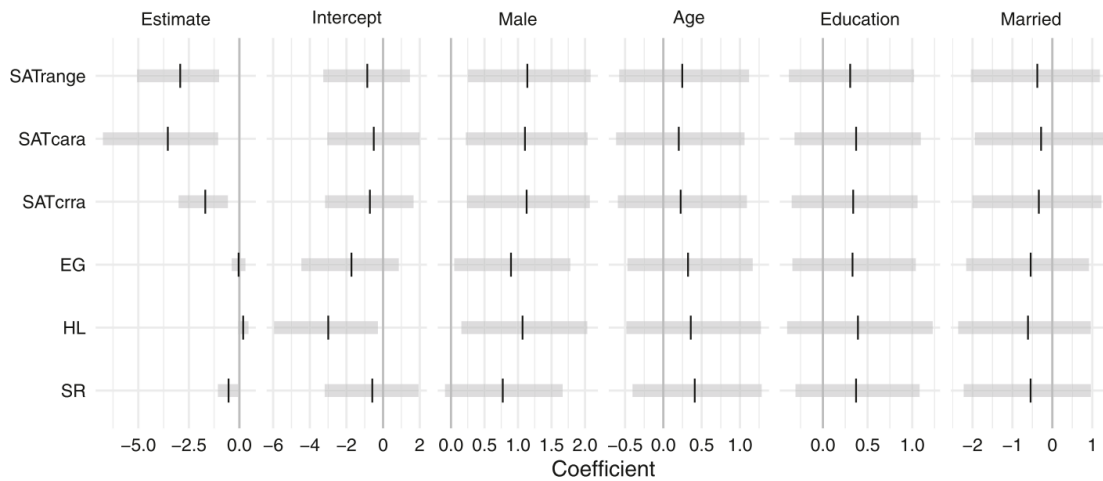

Fig. 4. Logistic regressions for each risk elicitation method, controlling for gender, age, education, and marital status. The vertical line indicates the estimated coefficient, and the gray bar indicates the 95% confidence interval. Intervals that cross 0 in the *x*-axis indicate statistically significant relationships. Values of SATCARA were multiplied by a constant to yield an estimated coefficient in the same order of magnitude as the other elicitation methods. This transformation does not alter the sign or significance of the estimates.

to ask retail investors how much loss they are willing and able to bear. The idea of a minimum aspiration is central in Herbert Simon's (1955) satisficing, a process that describes how people make decisions under uncertainty.

This study examined a risk-elicitation method based on satisficing (Berg et al., 2018). The method involves asking participants to state the minimum return they are willing to accept given a portfolio comprising a safe and a risky prospect. The resulting amount is a measure of risk preferences, which can be converted into other measures assuming CARA and CRRA utility functions.

We find that this measure has desirable characteristics. It correlates well with existing measures of risk preference (except with measurements elicited through Holt and Laury, 2002), it predicts better than other tools whether an individual participates in the stock market, and it has high test–retest reliability while capturing high response variation. Moreover, participants who used the tool report having similar confidence in their responses using this tool as in their responses using other risk-elicitation tools.
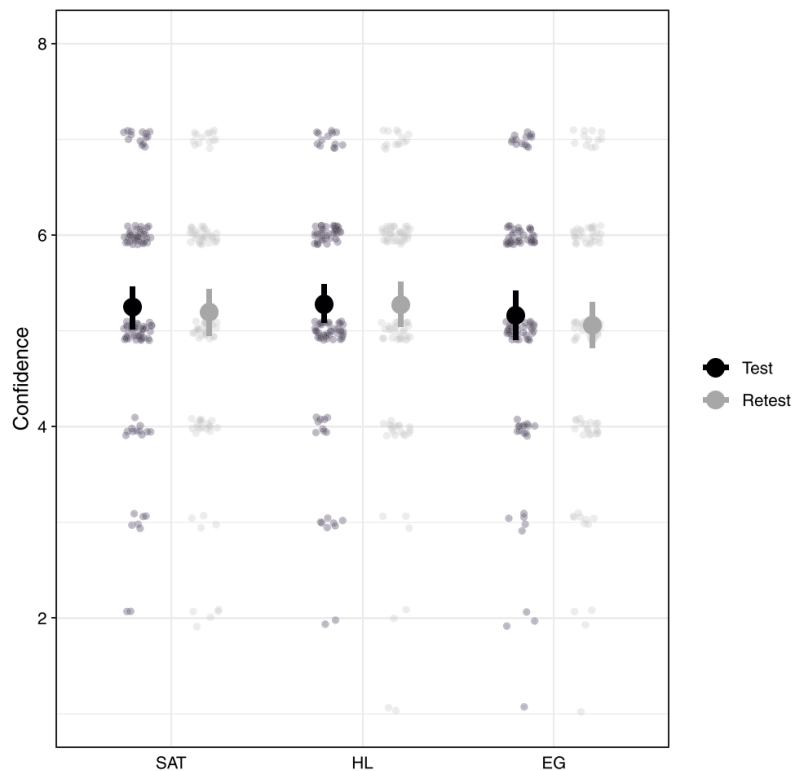
Fig. 5. Confidence in each of the three main risk-elicitation methods, in the test and retest phases. Each dot indicates an individual confidence measurement. Bigger dots indicate the mean accompanied by its 95% confidence interval.

8. Discussion

While using aspiration levels to assess risk preference is rare in the study of risk taking, the use of aspiration levels is common in techniques for risk management across a broad range of industries. For example, farmers minimize the probability of falling below the survival level by apportioning their land into "safe" crops and "risky" crops (Lopes, 1987), cabdrivers have a daily income target (Camerer et al., 1997), used car dealers set prices according to their minimum aspiration (Artinger and Gigerenzer, 2016), and investment managers try to meet a target return (Payne et al., 1980). Surveys of management practice provide additional evidence on the use of aspiration levels. For example, financial executives understand "investment risk" as: "the prospect of not meeting the target rate of return" (p. 343, Mao, 1970), and portfolio managers emphasize the significance of setting aspiration levels in behavioral portfolios (Shefrin and Statman, 2000). Although the study of aspiration levels has received little attention in the finance research, the use of aspiration levels in practice seems prevalent when making decisions under uncertainty (Artinger et al., 2021).

Because risk preferences are sensitive to the context (Blais and Weber, 2006), the propensity to take risks should be elicited in the context in which it is to be interpreted (Zhou and Hey, 2018). This is a recommendation that Loomes and Pogrebna (2014) made also for researchers: "... researchers who wish to take some account of and/or make some adjustment for risk attitude in their studies should take care to pick an elicitation procedure as similar as possible to the type of decision they are studying". (p. 592). We extend this recommendation to financial professionals, who must assess the risk tolerance of their clients for their client's benefit, but also to comply with recent regulatory requirements of investor protection.

All measures of risk propensity have advantages and limitations. SAT is a simple tool to measure risk preference in the context of financial behaviors that heeds the call from regulatory agencies to emphasize the worst possible consequences for investors. Admittedly, due to its investment framing, it is an open question whether SAT is a good measure of risk propensity beyond financial matters. But because the SAT task admits different framings, it remains to be examined in future research whether a framing can be devised that provides more general risk measurements.

Appendix A. Supplementary data

Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0009-22CC-B.

References

Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2006. Elicitation using multiple price list formats. Exp. Econ. 9 (4), 383–405.

Anderson, L.R., Mellor, J.M., 2009. Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. J. Risk Uncertain. 39 (2), 137–160.

Arslan, R.C., Brümmer, M., Dohmen, T., Drewelies, J., Hertwig, R., Wagner, G.G., 2020. How people know their risk preference. Sci. Rep. 10, 15365.

Artinger, F.M., Gigerenzer, G., 2016. Heuristic pricing in an uncertain market: Ecological and constructivist rationality. Available at SSRN: https://ssrn.com/ abstract=2938702.

Artinger, F., Gigerenzer, G., Jacobs, P., 2021. Satisficing: Integrating two traditions. J. Econ. Lit. in press.

Berg, N., Prakhya, S., Ranganathan, K., 2018. A satisficing approach to eliciting risk preferences. J. Bus. Res. 82, 127–140.

Bernouilli, D., 1738. Specimen theoriae novae de mensura sortis. Comment. Acad. Sci. Imp. Petropolitanae 1738 (5), 175–192.

Binswanger, H.P., 1981. Attitudes toward risk: Theoretical implications of an experiment in rural India. Econ. J. 91 (364), 867–890.

Blais, A.R., Weber, E.U., 2006. A domain-specific risk-taking (DOSPERT) scale for adult populations. Judgm. Decis. Mak. 1 (1), 33–37.

Bodie, Z., Kane, A., Marcus, A.J., 2014. Investments, 10th Global Edition. McGraw-Hill, Berkshire.

Brown, D.B., Sim, M., 2009. Satisficing measures for analysis of risky positions. Manage. Sci. 55 (1), 71–84.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: One day at a time. Q. J. Econ. 112 (2), 407–441.

Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. J. Risk Uncertain. 19 (1–3), 7–42.

Cameron, L., Shah, M., 2015. Risk-taking behavior in the wake of natural disasters. J. Hum. Resour. 50 (2), 484–515.

Charness, G., Eckel, C., Gneezy, U., Kajackaite, A., 2018. Complexity in risk elicitation may affect the conclusions: A demonstration using gender differences. J. Risk Uncertain. 56 (1), 1–17.

Charness, G., Gneezy, U., 2010. Portfolio choice and risk attitudes: An experiment. Econ. Inq. 48 (1), 133–146.

Charness, G., Gneezy, U., Imas, A., 2013. Experimental methods: Eliciting risk preferences. J. Econ. Behav. Organ. 87, 43–51.

Crosetto, P., Filippin, A., 2013. The bomb risk elicitation task. J. Risk Uncertain. 47 (1), 31–65.

Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Exp. Econ. 19 (3), 613–641.

Dave, C., Eckel, C., Johnson, C., Rojas, C., 2010. Eliciting risk preferences: When is simple better? J. Risk Uncertain. 41 (3), 219–243.

Deck, C., Lee, J., Reyes, J.A., Rosen, C.C., 2013. A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. J. Econ. Behav. Organ. 87, 1–24.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. J. Eur. Econom. Assoc. 9 (3), 522–550.

Eckel, C.C., Grossman, P.J., 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. Evol. Hum. Behav. 23 (4), 281–295.

Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: experimental evidence. In: The Handbook of Experimental Economics Results, Vol. 1. pp. 1061–1073.

Ert, E., Haruvy, E., 2017. Revisiting risk aversion: Can risk preferences change with experience? Econom. Lett. 151, 91–95.

European Securities and Markets Authority, European Securities and Markets Authority, 2018. Guidelines on MiFID II product governance requirements. Retrieved from https://www.esma.europa.eu/sites/default/files/library/esma3543-620_guidelines_on_mifid_ii_product_governance_requirements_0.pdf.

Fellner, G., Güth, W., Maciejovsky, B., 2009. Satisficing in financial decision making—A theoretical and experimental approach to bounded rationality. J. Math. Psych. 53 (1), 26–33.

Filippin, A., Crosetto, P., 2016. A reconsideration of gender differences in risk attitudes. Manage. Sci. 62 (11), 3138–3160.

Financial Industry Regulatory Authority, FINRA Manual Rule 2111: Suitability 2014. Retrieved from https://www.finra.org/sites/default/files/NoticeDocument/p123701.pdf.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., Hertwig, R., 2017. Risk preference shares the psychometric structure of major psychological traits. Sci. Adv. 3 (10), e1701381.

Friedman, D., Isaac, R.M., James, D., Sunder, S., 2014. Risky Curves: On the Empirical Failure of Expected Utility. Routledge.

Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112 (2), 631–645.

Guan, M., Stokes, R., Vandekerckhove, J., Lee, M.D., 2020. A cognitive modeling analysis of risk in sequential choice tasks. Judgm. Decis. Mak. 15 (5), 823–850.

Güth, W., 2007. Satisficing in portfolio selection—Theoretical aspects and experimental tests. J. Socio-Econ. 36 (4), 505–522.

Güth, W., Levati, M.V., Ploner, M., 2008. Is satisficing absorbable? An experimental study. J. Behav. Finance 9 (2), 95–105.

Harrison, G.W., Rutström, E.E., 2008. Risk aversion in the laboratory. Res. Exp. Econ. 12 (8), 41–196.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Amer. Econ. Rev. 92 (5), 1644–1655.

Jacobson, S., Petrie, R., 2009. Learning from mistakes: What do inconsistent choices over risk tell us? J. Risk Uncertain. 38 (2), 143–158.

Kaufmann, C., Weber, M., Haisley, E., 2013. The role of experience sampling and graphical displays on one's investment risk appetite. Manage. Sci. 59 (2), 323–340.

Lejarraga, T., Woike, J.K., Hertwig, R., 2016. Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. Cognition 157, 365–383.

Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L...., Brown, R.A., 2002. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). J. Exp. Psychol.: Appl. 8 (2), 75.

Lönnqvist, J.E., Verkasalo, M., Walkowitz, G., Wichardt, P.C., 2015. Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. J. Econ. Behav. Organ. 119, 254–266.

Loomes, G., Pogrebna, G., 2014. Measuring individual risk attitudes when preferences are imprecise. Econ. J. 124 (576), 569–593.

Lopes, L.L., 1987. Between hope and fear: The psychology of risk. Adv. Exp. Soc. Psychol. 20 (3), 255–295.

Mao, J.C., 1970. Survey of capital budgeting: Theory and practice. J. Finance 25 (2), 349–360.

Masclet, D., Colombier, N., Denant-Boemont, L., Loheac, Y., 2009. Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. J. Econ. Behav. Organ. 70 (3), 470–484.

Mata, R., Frey, R., Richter, D., Schupp, J., Hertwig, R., 2018. Risk preference: A view from psychology. J. Econ. Perspect. 32 (2), 155–172.

Mata, R., Josef, A.K., Hertwig, R., 2016. Propensity for risk taking across the life span and around the globe. Psychol. Sci. 27 (2), 231–243.

Menkhoff, L., Sakha, S., 2017. Estimating risky behavior with multiple-item risk measures. J. Econ. Psychol. 59, 59–86.

Nosić, A., Weber, M., 2010. How riskily do I invest? The role of risk attitudes, risk perceptions, and overconfidence. Decis. Anal. 7 (3), 282–301.

OECD, 2020. Purchasing power parities (PPP) (indicator). http://dx.doi.org/10. 1787/1290ee5a-en, (Accessed on 05 2020).

Payne, J.W., Laughhunn, D.J., Crum, R., 1980. Translation of gambles and aspiration level effects in risky choice behavior. Manage. Sci. 26 (10), 1039–1060.

Reynaud, A., Couture, S., 2012. Stability of risk preference measures: results from a field experiment on French farmers. Theory and Decision 73 (2), 203–221. Samuelson, P.A., 1938. A note on the pure theory of consumer's behaviour. Economica 5 (17), 61–71.

SEBI (Investment Advisers) Regulations 2013 [Last amended on December 08, 2016]. Available at https://www.sebi.gov.in/sebi_data/attachdocs/ 1358779330956.pdf.

Shefrin, H., Statman, M., 2000. Behavioral portfolio theory. J. Financ. Quant. Anal. 35 (2), 127–151.

Simon, H.A., 1955. A behavioral model of rational choice. Q. J. Econ. 69 (1), 99–118.

Simon, H.A., 1956. Rational choice and the structure of the environment. Psychol. Rev. 63 (2), 129.

Szrek, H., Chao, L.W., Ramlagan, S., Peltzer, K., 2012. Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. Judgm. Decis. Mak. 7 (6), 716.

Tanaka, T., Camerer, C.F., Nguyen, Q., 2010. Risk and time preferences: Linking experimental and household survey data from Vietnam. Amer. Econ. Rev. 100 (1), 557–571.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185 (4157), 1124–1131.

Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. J. Financ. Econ. 101 (2), 449–472.

Weber, E.U., Blais, A.R., Betz, N.E., 2002. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. J. Behav. Decis. Mak. 15 (4), 263–290.

Zhou, W., Hey, J., 2018. Context matters. Exp. Econ. 21 (4), 723–756.