J Risk Uncertain (2010) 41:219–243
DOI 10.1007/s11166-010-9103-z

Eliciting risk preferences: When is simple better?

Chetan Dave · Catherine C. Eckel ·
Cathleen A. Johnson · Christian Rojas

Published online: 22 September 2010

Abstract We study the estimation of risk preferences with experimental data and focus on the trade-offs when choosing between two different elicitation methods that have different degrees of difficulty for subjects. We analyze how and when a simpler, but coarser, elicitation method may be preferred to the more complex, but finer, one. Results indicate that the more complex measure has overall superior predictive accuracy, but its downside is that subjects exhibit noisier behavior. Our main result is that subjects' numerical skills can help better assess this tradeoff: the simpler task may be preferred for subjects who exhibit low numeracy, as it generates less noisy behavior but similar predictive accuracy. For subjects with higher numerical skills, the greater predictive accuracy of the more complex task more than outweighs the larger noise. We also explore preference heterogeneity and provide methodological suggestions for future work.

Keywords Risk aversion · Experiments · Elicitation methods · Heterogeneity

JEL classification C90 · C81

C. Dave
Department of Economics, New York University (Abu Dhabi),
19 W. 4th Street, 6FL, New York, NY 10012, USA
e-mail: cd86@nyu.edu

C. C. Eckel
School of Economic, Political and Policy Sciences, University of Texas at Dallas,
800 W. Campbell Rd., Richardson, TX 75080, USA
e-mail: eckelc@utdallas.edu

C. A. Johnson
Norton School of Family and Consumer Sciences, University of Arizona,
650 N. Park Avenue, Tucson, AZ 85719, USA
e-mail: cathleen@email.arizona.edu

C. Rojas (✉)
Department of Resource Economics, University of Massachusetts Amherst,
219 Stockbridge Hall, Amherst, MA 01003, USA
e-mail: rojas@resecon.umass.edu

Things should be made as simple as possible, but not simpler. *Albert Einstein*

In experimental decisions, individuals are assumed to reveal their preferences as long as the performed tasks have salient payoff consequences. A number of methods can be employed to elicit such preferences and the resulting data can be used to estimate the parameters of a utility function. Since many economic models of individual and family decision making rely on a parameterized utility function to make precise predictions, being able to effectively measure individual-specific risk preferences is important for predicting and understanding outcomes and preference heterogeneity across individuals.

In this paper we explore several issues involved in estimating the underlying risk preferences of individuals using experimental data. First, we study the trade-offs that arise when choosing between two different elicitation methods that have different degrees of difficulty for subjects. We place a particular emphasis on the population where task complexity may be more critical: people with limited mathematical skills. A brief description of the data should make this objective clearer. As part of a larger study, observations were collected on two experimental measures[1] of risk preferences for nearly 900 adults, together with an extensive survey and a widely-validated measure of mathematical skills (Johnson et al. 2003). The experimental measures differ in that one (developed by Holt and Laury 2002—HL hereafter) is more complex, involving ten decisions between gambles with probabilities ranging from 0.1 to 0.9, and allows categorization of decision makers into 10 categories, while the other (developed by Eckel and Grossman 2002, 2008—EG hereafter) is simpler, involving a single choice among 6 gambles, all with 0.5 probability of winning a higher prize, but only allows categorization of decision makers into 5 risk categories. The EG task is similar to Binswanger (1980, 1981); the primary difference with the EG protocol is that Binswanger's task includes dominated gambles, a nonlinear relationship between risk and return, and the presentation of tasks in a pair-wise format. HL has the advantage of providing a finer classification of subjects, but has the disadvantage of greater complexity, which may introduce errors into the subjects' choices. Greater complexity makes it more likely that the subjects fail to understand the task. The simpler EG method relies on more intuitive alternatives, with 50/50 gambles that are more easily understood by subjects who might struggle with descriptions of gambles involving varying probabilities and payoffs. Its disadvantage is the coarser classification of subjects. It is unclear which method dominates the other, and the question of which method to use to elicit risk preferences is an open one.

As in much of our earlier field work, we noted that in the above described experiment our (non-conventional) subjects (i.e. non-college students) had a more difficult time grasping the HL task, demonstrated by their notable level of confusion and their frequent requests for clarification, as well as more frequent inconsistent decisions.[2] This observation led us to conjecture that important tradeoffs may be involved in developing experimental tasks to measure risk attitudes. We employ an expected utility framework to investigate these tradeoffs: We assume that the HL and

[1] We use the terms "elicitation methods" and "experimental measures" interchangeably throughout the text.

[2] When subjects fail to understand a task in the field, they tend to resort to alternating between options (for example, alternating between the risky and safe gambles in the HL task), and describe their choices with phrases such as "I just wanted to try it."

EG tasks explore the curvature of individuals' utility functions and that the data from these tasks can be used to estimate the parameter(s) of such utility functions. Specifically, and following Holt and Laury (2002), we estimate a constant relative risk aversion (CRRA) utility function that allows for a parameter that captures noisy behavior. The attractiveness of such specification is that the noise parameter accommodates the possibility that subjects make decisions with some degree of error.[3, 4]

A significant part of our analysis compares the estimated noise parameter between the two risk elicitation methods, so it is important to describe the interpretation of this parameter in this particular specification. A small noise parameter indicates that subjects' decisions conform closely to the predictions of expected utility theory, whereas a large noise parameter is considered as evidence of a high degree of randomness (i.e., noisy behavior) in subjects' decisions. The advantage of the cognitively more difficult measure (HL) is that it produces a finer categorization; as a result, estimates from the HL measure may have higher predictive accuracy (as measured by the percentage of choices that are correctly predicted by the estimates). However, complexity comes at a cost: subjects' decisions exhibit larger noisy behavior, especially among the low math skills population. Conversely, the disadvantage of the simpler measure is that it produces a coarser categorization (and thus lower predictive accuracy), but decisions are substantially less noisy. A central result in our analyses is that the simpler measure appears to be unambiguously superior for subjects with low math skills as it generates smaller noise and equal predictive accuracy as compared to the complex measure.

We also investigate heterogeneity in preferences by allowing the risk preference and noise parameters to vary by individual characteristics: gender, income and age (in addition to the math skills measure discussed earlier). As in earlier studies (see Eckel 2007, Eckel and Grossman 2008, and references therein), we find a higher degree of risk aversion for women; this effect is strongly present in our data as it appears in all specifications and in both risk elicitation methods. Low math ability is associated with greater risk aversion (especially in EG), and a larger noise parameter (especially in HL). Interestingly, the more complex measure shows a greater level of overall risk aversion. We discuss alternative interpretations of these findings in the results and conclusion sections.

Our overall assessment of the results is to indicate that the type of measurement technique may be critical in recovering reliable estimates of risk preferences. Economists often neglect the importance of how a task is presented to subjects—in particular, the difficulty of the task—but focus instead on the theoretical properties of such tasks. Complexity seems to be an important issue when designing an appropriate method, especially for low math skills populations. Thus, care should be taken to develop experimental measures of preferences that have desirable properties for the population under study.

[3] The existence of a well-behaved utility function (with precise decision making skills on the part of subjects) has been criticized, for example, by researchers that have found that subjects' preferences are inconsistent across similar tasks (e.g., Slovic 1962; Isaac and James 2000; Berg et al. 2005; Peters et al. 2006).

[4] To ensure that this is a sensible specification, we also investigate results of alternative specifications of the utility function in Sections 3 and 4 below.

In Section 1 we link our study to previous literature; Section 2 provides a brief description of the experimental data and Section 3 describes the methodology employed. Section 4 reports model estimates and Section 5 provides a conclusion and discussion.

1 Literature review

Several experimental procedures have been developed for measuring risk aversion.[5] A widely-used means of measuring risk attitudes is one based on the two-stage preference-revelation mechanism developed by Becker et al. (1964). This procedure requires subjects to choose a selling price for a specified lottery. A randomly drawn value then determines whether the subject sells the lottery (if the drawn price exceeds the subject's price) or plays the lottery. Examples of its use include Harrison (1986), Kachelmeier and Shehata (1992), Eckel (1999) and Eisenberger and Weber (1995). However, this mechanism has come under scrutiny because of its sensitivity to seemingly irrelevant parameters such as the upper range of possible valuations (Bohm et al. 1997) and the low incentives to value low-probability lotteries accurately (Harrison 1992). Karni and Safra (1987) and Horowitz (2006) point out that it may not even be incentive compatible for certain preferences and goods.

Several additional measures involve choices between or among lotteries. Binswanger (1980, 1981) developed a measure for use in rural India, which asks subjects to make binary choices between pairs of 50/50 gambles; in this procedure, gains in expected value can be had only with an increase in risk (standard deviation). Hey and Orme (1994) also use a lottery-choice approach, with more complex probabilities. Harbaugh et al. (2002) compare several different ways of eliciting risk preferences, including valuation of lotteries and choices between lotteries, and find different patterns of errors depending on whether subjects are presented with valuation or choice tasks. In Harbaugh et al. (2010) they continue this line of research, and their results suggest that failing to account for errors may significantly affect estimated risk preference parameters in ways that are task-dependent.

We adopt two risk elicitation methods: Holt and Laury (2002) and Eckel and Grossman (2002). Both tasks involve choices among lotteries. As stated earlier, Eckel/Grossman is similar in approach to Binswanger (1980, 1981). Recent implementations of Binswanger's protocol adopt a format very similar to theirs, and have been used successfully in the field with less literate populations (e.g., Barr and Genicot 2008).

Our work fits within the larger methodological effort to improve the measurement of risk aversion and our understanding of the factors that determine risk preferences. We build on earlier work, including Andersen et al. (2008), who show how ignoring individuals' risk preferences may lead to biased estimates of inter-temporal preferences. Finally, a recent literature focuses on studying whether observed deviations from risk neutrality are related to cognitive ability and task complexity (e.g. Huck and Weizsacker 1999; Burks et al. 2009).

[5] Starmer (2000) surveys the symbiotic theoretical and empirical developments in analyzing decision making under risk.

2 Description of experiments and data

2.1 Implementation

From May 2002 to March 2003, 881 Canadian residents participated in 102 experimental sessions. The sample was drawn from both urban and non-urban sites across Canada and was made up mainly of adult labor market participants. The sample includes three age groups: young adults, age 18–24, including 80 high school students; adults aged 25–44 who had some labor-force attachment (including both employed and unemployed samples); and older adults aged 45–55. Table 1 gives sample details.

Unemployed participants were recruited by local Human Resource Centre of Canada (HRCC) staff and using pamphlets placed in HRCC waiting areas. Other participants were targeted through newspaper advertising in popular daily newspapers, and TV and radio announcements on community stations. Site visits to local businesses and community organizations also proved to be effective recruiting tools, especially for younger participants. As expected, the recruited subjects belong to a lower income level than the typical Canadian (see bottom of Table 1), although the average adult age 45 and older in our sample is comparable to the average Canadian adult in terms of education attainment.

Recruitment materials included information about the project, time commitment, a show-up fee (CAD$20), the potential for more earnings, and confidentiality of responses. Participants volunteered by calling a toll-free number or accessing a web page, where they completed a demographic questionnaire to determine eligibility.

Table 1 Sample summary

Demographic	# of Participants	
Age 18–24	170	
Age 25–44	438	
Age 45–55	193	
Male	347	
Female	454	
High school student sample	80	
Post-Secondary Student	101	
Unemployed	162	
Part-time employed	168	
Full-time employed	260	
Total	881	
Aggregate Demographics		
	Our Sample	Canada
Median Income	CAD 10,000-20,000	CAD 54,000
Average years of schooling (45 and older)	11.93	11.60

Authors' data and Organization for Economic Co-Operation and Development

The experimental sessions were conducted in controlled environments including classrooms, boardrooms and hotel conference facilities. Upon arrival, the experimenter reminded participants of the confidentiality of the data, and provided participants with appropriate details of the potential earnings, including the possibility of cash payments. The maximum number of participants in any session was 30. Care was taken to ensure that subjects understood the decisions they were to make. Because these decisions were unfamiliar, thirteen practice examples representing all types of decisions were demonstrated to ensure that subjects understood the nature of the decisions and how payment was linked to their choices.

Subjects completed a series of experimental decisions involving choices between two or more alternatives. After all decisions were made, one decision was selected at random (from all decisions) for payment using a bingo ball cage, where each decision number was matched with one corresponding bingo ball number.[6] Each had an equal probability of being selected, making decisions independent of each other. Subjects were paid in private, and average earnings were CAD$165. Each session included exactly the same set of decisions, in the same order, and took between 1.5 and 3 h to complete from instruction to payoff.

2.2 Risk preferences

To measure risk preferences, participants completed two sets of decisions involving choices among cash gambles (see Appendix). Subjects' choices and details of the gamble sets are shown in Tables 2 and 3. In the first experimental measure, developed by Eckel and Grossman (2002, 2008), subjects choose from among six possible gambles the one they would most prefer to play.[7] To determine payment, the subject plays the chosen gamble by rolling a die.

Table 2 contains the gambles in the choice set, all of which involve a 50/50 chance of a low or high payoff. The range of gambles includes a safe alternative involving a sure payoff of $28 with zero variance. From here, the gambles increase in both expected return and risk (standard deviation) moving from Gamble 1 to 5; expected return increases linearly with standard deviation. Gamble 6 involves only an increase in variance, with the same expected return as Gamble 5. More risk-averse subjects would choose lower-risk, lower-return gambles; risk-neutral subjects would choose Gamble 5 or 6, which have the highest rate of return; only risk-seeking subjects would choose Gamble 6. This task was designed to be as simple as possible, while retaining a reasonable range of risky choices, and takes only a few minutes to explain and implement. Subjects rarely have any difficulty understanding it.

[6] There were 8 types of decisions (three used here), and 100 decisions in total. All subjects were paid at the end of the experimental session, after all decisions were completed, including the (unpaid) math literacy task that we describe below.

[7] Our design does not allow us to investigate order effects directly since all subjects completed the EG task prior to the HL task. A potential drawback for our results is that we cannot determine whether the additional noise observed in HL is due in part to this particular ordering. We deem this possibility as somewhat unlikely as it is not clear how a single decision on a prior gamble (recall that it is only one choice that subjects make in EG) can increase noise later in HL (i.e. one can argue that it may decrease it as subjects are "more focused" on analyzing gambles after seeing the EG task).

Table 2 Eckel-Grossman gamble choices (subjects choose which gamble to play)

Choice (50/50 Gamble)	Low Payoff	High Payoff	Expected Return	Standard Deviation	Implied CRRA[a] Range	Fraction of Subjects (%)
Gamble 1	28	28	28	0	$3.46<r$	10.7
Gamble 2	24	36	30	6	$1.16<r<3.46$	11.2
Gamble 3	20	44	32	12	$0.71<r<1.16$	39.2
Gamble 4	16	52	34	18	$0.50<r<0.71$	16.8
Gamble 5	12	60	36	24	$0<r<0.50$	11.5
Gamble 6	2	70	36	34	$r<0$	10.7

[a] Coefficient of relative risk aversion

Table 2 also includes ranges of coefficients of relative risk aversion implied by each possible choice, under the assumption of constant relative risk aversion (CRRA), shown as 'r' in the table. Each range was calculated by comparing the gamble to its neighbors, and calculating the value of r that generates the same utility level for the payoffs associated with each adjacent gamble.[8] A choice of Gamble 6 implies risk-seeking, with $r<0$. The distribution of choices shown here is very similar to samples of university students with stakes 1/3 to 1/2 this size (e.g., Eckel and Grossman 2008). While not completely comparable with the parameters of our study, earlier versions of this task (Binswanger 1980, 1981) also appear to have generated distributions of choices that are similar to ours.

Table 3 provides the r ranges and the frequency of safe choices for the risk aversion measure developed by Holt and Laury (2002). This widely-used experimental measure involves a set of ten binary choices between a low risk gamble and a high risk gamble. The two gambles have the same probabilities but different low and high payoffs, making them relatively easy to compare. In our implementation Gamble A has payoffs of CAD$40 or CAD$32, and Gamble B has payoffs CAD$77 or CAD$2. For the first decision, the probability of winning the larger prize is 1/10, and most student subjects quickly see that they prefer Gamble A. This probability increases by 1/10 in each subsequent decision, so that at some point subjects switch from preferring A to B, and prefer B thereafter. The switch point determines the number of safe choices and, in turn, the risk aversion parameter range. A risk neutral subject would switch between decisions 4 and 5, making 4 safe decisions. Subjects in this sample tend to be risk averse, making more than 4 safe decisions.

Using the number of safe choices as an aggregate measure is not a fully accurate summary of the distribution of choices because some subjects (8.5% of the sample) make *inconsistent* decisions, either by switching more than once or by making "backwards" choices (switching in the other direction). The last column of the table removes these subjects; the resulting distribution is slightly more risk averse, overall. In the analysis below we distinguish the behavior of the subjects who make

[8] A person choosing Gamble 3, for example, would have a coefficient of relative risk aversion in the range 0.71–1.16: a person with $r=0.71$ would be just indifferent between Gambles 3 and 4, and a person with $r=1.16$ is just indifferent between Gambles 2 and 3.

Table 3 Holt-Laury gamble choices

Number of safe choices	Implied CRRA range	Fraction of choices (%)[a]	
		All subjects	Excluding inconsistent subjects
0	$r< -1.71$	1.25	1.36
1	$-1.71<r< -0.95$	0.34	0.37
2	$-0.95<r< -0.49$	0.45	0.25
3	$-0.49<r< -0.14$	3.75	3.35
4	$-0.14<r<0.15$	10.78	10.42
5	$0.15<r<0.41$	12.83	11.41
6	$0.41<r<0.68$	23.16	22.46
7	$0.68<r<0.97$	21.23	22.58
8	$0.97<r<1.37$	11.80	12.03
9–10	$1.37<r$	14.41	15.76

[a] Number of All Subjects=881, Number of Inconsistent=75. An inconsistent subject is one that, as he/she moves down the HL decisions, switches back to the safe gamble (A) after having chosen a risky gamble (B)

consistent decisions, but report regressions for both, since they constitute a significant portion of the population. Subjects may make inconsistent decisions for many reasons, not all of which involve confusion. For example, subjects may be indifferent among alternatives, or doubt their own ability to judge the correct decision and so include a mix of decisions. When questioned about their choices, inconsistent subjects tend to claim a reason for their choices and to stick with them.

The distribution of choices shown below is similar to those reported in Holt and Laury (2002) for university students making decisions at this stakes level (equivalent to their 20X Real treatment).

Comparing the two methods, the EG measure involves fewer (and simpler) gambles and a single choice, while the HL measure involves more complex gambles and more choices. Part of our purpose in including both tasks was to compare the two elicitation methods, and examine the tradeoff involved in having a coarser screen but an easier set of decisions for subjects to make. Our experience in the field and in the lab is that this difference in difficulty may not be an issue for university students, but less skilled subjects may find the EG method easier to understand.

2.3 Mathematical skills

A subject's ability to reason with numbers and math may affect his ability to understand and choose among risky alternatives, and hence the ability of the experimentalist to infer the degree of risk aversion. The study included an assessment that measures skill with numbers in everyday life (numeracy); we use it here to explore the relationship between noisy behavior and skill level. This numeracy assessment is a subcomponent of the Educational Testing Service's Adult Literacy and Lifeskills Survey (ALLS), a widely used and validated literacy test; for example, the score on the test is a strong predictor of income earning ability

(Statistics Canada and OECD 2003). The numeracy assessment consists of 31 problems involving the use of mathematics in real-life situations, and the results provide a gauge of an individual's competencies. Subjects completed this unpaid task after the experimental decisions, but before earnings are determined. The distribution of scores in this study is significantly higher than the general Canadian population.[9]

As an initial assessment of the potential importance of math skills for accurate measurement of risk attitudes, we calculated the number of people making inconsistent choices in HL by ability level. Figure 1 shows that the fraction of inconsistent subjects is approximately 4 times larger in the low math ability population (subjects whose math ability score is more than one standard deviation below the sample mean) than in the remaining population.

2.4 Coding choice data

To proceed with the econometric analysis, we convert choices in both elicitation methods to a binary format. In each of the 10 decisions of the HL method, the risky choice (B) takes a value of 1, and the less risky choice (A) takes a value of 0. Subjects in the EG method make only one choice; hence EG choices are not directly comparable to HL choices. To make data in both methods as comparable as possible, we transform the EG choice into data that has a format similar to that of HL. First we convert the six gambles into the five (HL equivalent) implicit binary decisions that result from comparing adjacent gambles:[10]

Decision 1: Gamble 5 v. Gamble 6
Decision 2: Gamble 4 v. Gamble 5
Decision 3: Gamble 3 v. Gamble 4
Decision 4: Gamble 2 v. Gamble 3
Decision 5: Gamble 1 v. Gamble 2

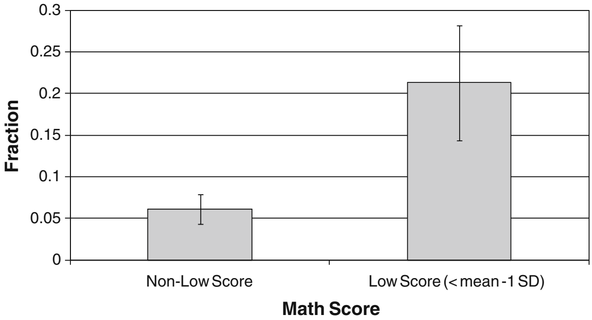
We have arranged these decisions to parallel the order of the decisions in HL: the more risky gamble is located on the right, and Decision 5 corresponds to subjects with the highest level of risk aversion. To illustrate the coding procedure, consider a subject who chooses Gamble 4 in the EG method. This implies a coding of 0, 0, 1, 1 and 1, for each of the five hypothetical decisions, respectively. Note that, as opposed to the HL method, this coding of the EG data does not allow subjects to be inconsistent (switching back and forth, or switching "backwards"). We address this issue in the estimation by analyzing the sensitivity of our results when inconsistent subjects in the HL method are excluded.

To illustrate that coding the EG data as pair-wise comparisons between adjacent gambles is not an unreasonable depiction, consider Table 2, and recall that expected return is linear in standard deviation (risk). Plotting the gambles with expected return

[9] The average raw score for our sample (out of a possible 500) is 281.25 (SEM=1.7). For the Canadian population the average is 272.3 (SEM=0.7) (Statistics Canada 2003, p. 49).

[10] Engle-Warnick et al. (2005) conduct experiments where subjects complete both a binary-choice version and a single-choice version of the EG task and report that most subjects make equivalent choices. We also have conducted both with a sample of students, and found about 85% consistent choices across the two presentations. (Data available on request.)

Fig. 1 Fraction of Subjects Making Inconsistent Choices in the HL Task*



Note: Vertical bars denote 95% confidence intervals.

*An inconsistent subject is one that, as he/she moves down the decisions on the HL task, switches back to the safe gamble (A) after having chosen a risky gamble (B).

on the vertical axis and standard deviation on the horizontal, the gambles trace an upward sloping line (until Gamble 6, which is on a straight horizontal line from Gamble 5). Risk-averse indifference curves in this space would be upward sloping, with a tangency on the upward sloping line representing the subject's preferred choice. Suppose a subject chooses Gamble 3, with a tangency at that point. For well-behaved utility functions, this implies that Gamble 2 is preferred to 1, 3 preferred to 2 and 3 is preferred to 4, 4 to 5, and 5 to 6.

2.5 Other variables and additional data details

We construct several variables to explore the issue of whether estimated parameters vary by experimental measure and by subjects' characteristics. Our data is pooled across elicitation tasks and a dichotomous variable, denoted HL, takes a value of 1 if the choice corresponds to the HL measure and zero otherwise. Variables for subjects' characteristics include Female (equal to 1 for women), Low Income (equal to 1 if the subject falls in the lower third of the income distribution), and Young (equal to 1 if age is below 25). The variable Low Math Score identifies people with low scores on the numeracy measure and is equal to 1 for scores more than one standard deviation below the sample mean score; 16% of the subjects fall into this category. A discrete (rather than continuous) measure was employed for math skills because of our strong prior that a threshold level of math skill was necessary to understand the experimental decisions.[11] Similarly, Young and Low Income are defined as dichotomous variables given our original hypotheses.

3 Methodology

The procedure for obtaining structural estimates for the utility function of the CRRA form is as follows. Subjects faced with several binary choices between two gambles

[11] In Section 4.2 we discuss different "cutoff" scores for classifying low math ability subjects.

are assumed to have a utility for money (*M*) given by $U(M|r)$, where *r* denotes the coefficient of relative risk aversion.[12] For each binary choice between gambles subjects are assumed to conduct an expected utility calculation of the form:

$$EU_i = \sum_k [\, p_k \times U(M_k|r)], \forall k = 1, 2$$

for each gamble *i*, where the probability of occurrence of each amount M_k is denoted by p_k.

Denoting EU_L as the 'left' gamble and EU_R as the 'right' gamble, one can construct a probabilistic choice rule with the ratio:

$$\frac{EU_R^{\frac{1}{\mu}}}{EU_R^{\frac{1}{\mu}} + EU_L^{\frac{1}{\mu}}} \tag{1}$$

The parameter μ allows for deviations from the deterministic choice specified by expected utility theory. As in previous work, μ is interpreted as noise: as $\mu \rightarrow \infty$ the choice becomes a random decision whereas as $\mu \rightarrow 0$ subjects behave exactly as specified by expected utility theory. In this paper, this parameter plays a key role as part of our analysis is based on the estimated noise parameter of each instrument.

The ratio in (1) forms the basis of a logistic conditional logarithmic likelihood function, denoted as $\ell(r, \mu|Y_i)$ that can be maximized with respect to *r* and μ, where the vector Y_i denotes the actual subjects' choices for either the 'Left" or 'Right' gamble. In order to allow for observed heterogeneity, each of the parameters in the vector $[r\ \mu]$ is specified as a function of individual characteristics, X_i, with associated coefficient vector β. The resulting modified likelihood function is written as $\ell(r, \mu, \beta|Y_i, X_i)$.[13]

4 Estimation results

We discuss our results in three parts. First, we present the estimates of the risk and noise parameters and discuss the observed differences across experimental measure, as well as subjects' characteristics (including math skills). The second part discusses in detail the trade-offs that arise when using the two experimental measures considered, with an emphasis on the noise and predictive accuracy of each measure. The third part discusses the time consistency results.

[12] We also employed expo-power and power utility functions in our estimation to verify the extent to which our results (reported below) were sensitive to the assumed functional form for utility. We ran into several problems in such specifications: a) convergence was difficult to achieve, b) estimates produced fitted values that were outside the range of observed data, and c) estimates were at times inconsistent with economic theory. We nevertheless conducted robustness checks with a Fechner error specification (see Hey and Orme 1994) in lieu of the Luce specification in (1) which included linear, logarithmic, power and expo-power functional forms of utility; our main findings were confirmed in such estimations (results can be made available upon request).

[13] We also implemented interval regressions to estimate risk stances with and without inconsistent subjects; however we do not report such results as this specification does not contemplate the 'noise' associated with the data (the focus of our study).

4.1 Risk and noise estimates

Table 4 presents the estimates of the coefficient of relative risk aversion as a function of several variables; the noise parameter μ is presented here as a constant. Results indicate that, *ceteris paribus*, the HL measure generates a significantly higher estimate of the risk aversion parameter, indicating greater aversion to risk, approximately 50% larger than that of the EG measure. In this specification, younger persons are less risk averse, and, notably, subjects with low math skills exhibit significantly more risk aversion.

Table 5 displays estimates when the noise parameter is allowed to vary. Model 1 includes variables for the risk measure, gender, income, age and Model 2 adds the math skills threshold. The estimates of risk aversion differ from Table 4 when noise is allowed to vary systematically; HL now carries a smaller, though still significant, coefficient in r, still indicating a higher level of measured risk aversion than EG. HL also shows a highly significant effect on the noise parameter, indicating a higher level of randomness in HL decisions. Women now appear more risk averse, but also have lower noise. Income has no effect in either parameter. Young subjects are no longer less risk averse, but rather exhibit lower noise. In Model 2, Low Math Score has a significant effect on risk aversion, but also results in marginally higher noise. Including this variable has little effect on the magnitude or significance of the other coefficients.

Tables 4 and 5 indicate that there is a significant difference between the estimated risk coefficients for subjects conditional on which type of experimental measure they are engaged in. This is a somewhat unexpected result because the measures are similar, in that both consist of choices among relatively simple gambles and both have similar payoff ranges. In principle, we would expect that if both measures are accurately measuring preferences and the degree of heterogeneity, the parameters shown in Table 5 should not vary by experimental measure. It is possible, however, that different experimental measures capture preference and noise heterogeneity differently.

Table 4 Estimates of risk parameter "r" as a function of characteristics

Parameter/Variable	Estimate	Std. Err.[a]	p-value
r			
HL	0.2295	0.0353	0.000
Female	0.0081	0.0255	0.750
Low Income (<$30,000)	0.0065	0.0201	0.746
Young (<25)	−0.0695	0.0290	0.017
Low Math Score (<mean—1 SD)	0.1817	0.0339	0.000
Constant	0.4871	0.0320	0.000
μ	0.0802	0.0070	0.000
Log$L$:	5839.5		
Obs.:	13215		

[a] Clustered by individual

Table 5 Estimates of risk parameter "r" and noise parameter "μ" as functions of characteristics, pooled data regression

Parameter/Variable	Model 1			Model 2		
	Estimate	Std. Err.[a]	p-value	Estimate	Std. Err.[a]	p-value
r						
HL	0.1433	0.0385	0.000	0.1514	0.0339	0.000
Female	0.1289	0.0350	0.000	0.1168	0.0360	0.001
Low Income (<$30,000)	0.0106	0.0345	0.759	−0.0003	0.0293	0.991
Young (<25)	0.0068	0.0390	0.844	−0.0104	0.0305	0.734
Low Math Score (<mean—1 SD)	–	–	–	0.1625	0.0405	0.000
Constant	0.4083	0.0380	0.000	0.4096	0.0369	0.000
μ						
HL	0.0708	0.0076	0.000	0.0605	0.0065	0.000
Female	−0.0210	0.0080	0.009	−0.0212	0.0075	0.005
Low Income (<$30,000)	0.0106	0.0081	0.188	0.0024	0.0066	0.717
Young (<25)	−0.0192	0.0075	0.010	−0.0149	0.0062	0.016
Low Math Score (<mean—1 SD)	–	–	–	0.0290	0.0178	0.103
Constant	0.0596	0.0073	0.000	0.0571	0.0062	0.000
Log$L$:	5781.4			5695.3		
Obs.:	13215			13215		

[a] Clustered by individual

In the field, some subjects appeared to have greater difficulty understanding the task in the HL measure, requesting more clarification from the session staff, whereas the one-choice format of the EG measure appeared to be understood more quickly and easily. Thus, we hypothesize that while decisions appear to be noisier with the HL measure, this increased noise is attributable in part to individuals with limited mathematical skills. Adding interaction terms of the HL variable with heterogeneity covariates would allow us to investigate this hypothesis, but we encountered poor convergence in such a model. Alternatively, we analyze two separate specifications, one for the HL measure and one for the EG measure; Table 6 contains these results.

In the risk parameter, we find that the coefficients are fairly consistent across the two models. The constant is substantially higher in HL, reflecting the previously observed pattern. In both cases, women are more risk averse than men, though this effect is twice as large in the EG regression. Income and Young are insignificantly related to risk aversion for both measures. Low Math Score carries an insignificant coefficient in the HL regression; this is also true for the EG regression, though the coefficient is large. In the noise parameter, the constant is again much higher in the HL regression, about three times the magnitude of the constant in the EG regression. Gender is weakly significant for EG and strongly significant for HL, indicating lower noise for women in that measure. Young carries a negative coefficient in both, but is only significant in the HL regression. Notably, Low Math Score individuals exhibit significantly higher noise in HL but not in EG; moreover the coefficient on this variable is an order of magnitude larger than other coefficients (except for the

Table 6 Estimates of risk parameter "r" and noise parameter "μ" as functions of characteristics, separate specifications for EG and HL instruments

Parameter/Variable	EG Instrument			HL Instrument		
	Estimate	Std. Err.[a]	*p*-value	Estimate	Std. Err.[a]	*p*-value
r						
Female	0.2584	0.0658	0.000	0.1275	0.0386	0.001
Low Income	0.0202	0.0511	0.693	−0.0271	0.0388	0.485
Young	−0.0190	0.0552	0.731	0.0415	0.0384	0.280
Low Math Score	0.2344	0.1570	0.136	0.0218	0.0921	0.813
Constant	0.3260	0.0382	0.000	0.5479	0.0356	0.000
μ						
Female	−0.0134	0.0080	0.092	−0.0337	0.0128	0.009
Low Income	0.0002	0.0063	0.979	0.0113	0.0117	0.332
Young	−0.0074	0.0069	0.283	−0.0390	0.0115	0.001
Low Math Score	0.0123	0.0227	0.587	0.1023	0.0466	0.028
Constant	0.0462	0.0045	0.000	0.1302	0.0117	0.000
Log*L*:	2490.3			3172.2		
Obs.:	4405			8810		

[a] Clustered by individual

constant) indicating that limited math skill is an important source of noisy behavior in HL. We next explore further the importance of mathematical skills in the estimation of risk preferences.

4.2 Consistency, noise and predictive accuracy

In this section we investigate three themes. First, we look at the sensitivity of our results to a subset of the population that excludes inconsistent subjects. Then, we take a closer look at the differential noise estimates across instruments, with a focus on low math skills subjects. Finally, we analyze the differential predictive accuracy across instruments, again with an eye on low math skills subjects.

One reason that the estimates for HL and EG in Table 6 might differ is because of the way choices are made and coded. Recall that for EG, one gamble is chosen from a set. To model HL and EG decisions in parallel ways, we have inferred the hypothetical binary choices that result from the chosen gamble in EG. Therefore a subject cannot make inconsistent choices in EG. However, in HL, a subject can exhibit inconsistent choices, by switching between the less and more risky options more than once, or can make "backwards" choices by beginning with B and switching to A choices. In Table 7 we identify and remove the 75 inconsistent subjects (8.5% of the sample, including 30 of the 141 low math subjects), and re-estimate both equations using only consistent subjects.[14]

[14] Four additional subjects chose the safe outcome for all ten HL decisions. Excluding them does not change the results.

Table 7 Estimates of risk parameter "r" and noise parameter "μ" as functions of characteristics, separate specifications for EG and HL instruments, Consistent subjects only[a]

Parameter/Variable	EG instrument			HL instrument		
	Estimate	Std. Err.[b]	p-value	Estimate	Std. Err.[b]	p-value
r						
Female	0.2629	0.0721	0.000	0.1268	0.0388	0.001
Low Income	0.0108	0.0549	0.684	−0.0164	0.0420	0.696
Young	0.0071	0.0580	0.903	0.0480	0.0384	0.211
Low Math Score	0.2549	0.1159	0.028	0.0673	0.0910	0.459
Constant	0.3245	0.0378	0.000	0.5536	0.0379	0.000
μ						
Female	−0.0143	0.0097	0.140	−0.0331	0.0123	0.007
Low Income	0.0011	0.0073	0.880	0.0092	0.0113	0.415
Young	−0.0085	0.0090	0.342	−0.0360	0.0109	0.001
Low Math Score	0.0098	0.0210	0.642	0.0357	0.0370	0.335
Constant	0.0455	0.0051	0.000	0.1192	0.0114	0.000
Log$L$:	2260.5			2636.9		
Obs.:	4030			8060		

[a] An inconsistent subject is one that, as he/she moves down the decisions on the HL task, switches back to the safe gamble (A) after having chosen a risky gamble (B)

[b] Clustered by individual

The pattern of coefficient estimates is very similar to those in Table 6, with a couple of exceptions. In EG, the slightly higher coefficient on Low Math Score combined with a substantially lower standard error means that it now carries a significant coefficient in r, indicating this group is more risk averse. However, for HL, the coefficient on Low Math Score in r remains insignificant. As expected, for the noise parameter estimates, the coefficient on Low Math Score for HL drops in magnitude and is no longer significant, in line with the higher frequency of inconsistent choices among low math score subjects displayed in Fig. 1. In EG there are no longer any significant determinants of noise: noise does not vary by any of the characteristics. However, in HL, men and older persons continue to have higher noise parameters.

Our interpretation of these results is that switching (or making reversed choices) in HL is an indicator that the subject may not understand the task, and so HL generates a less reliable measure of risk aversion for this population. Inconsistent behavior is a much bigger problem for low math skill subjects, as shown in Fig. 1 above. Critical for purposes of measurement is being able to identify whether low math skill individuals are in fact more risk averse, or whether higher risk aversion might be an artifact of the measurement process; our analysis suggests that the former may be more plausible. This possible correlation between math skills and risk attitudes is discussed further below.

To visualize the differences in the estimated noise parameters across instruments, we calculate the predicted noise parameter for each subject under both tasks. Figure 2, Panel A presents the cumulative distribution of the predicted noise parameter by experimental measure (HL and EG), and by math skills level (for the HL measure only) based on the estimates in Table 6. There are striking differences between the two measures' estimated noise. The range of the noise parameter for the EG measure is between 0.025 and 0.059 and for the HL measure between 0.058 and 0.244: there is virtually no overlap between these two distributions. For HL, the range of the noise parameter shifts considerably for people with low math skills: 0.160 to 0.244.[15]

Panel B of Fig. 2 graphs the same relationships using only subjects who made consistent choices in the HL measure based on Table 7 estimates. Here the range for EG is between 0.023 and 0.059; for HL it is between 0.050 and 0.164 for all subjects, and between 0.086 and 0.244 for low math skills subjects. Compared to Panel A, the HL distributions are shifted to the left, and are much closer to EG, with a larger overlap in the distributions. Using the distributions in Panel B, one can compute a statistical test of the difference in the *median* noise: at any confidence level, the median noise parameter for low math skills subjects under the HL measure is larger than the median noise parameter under the EG measure.

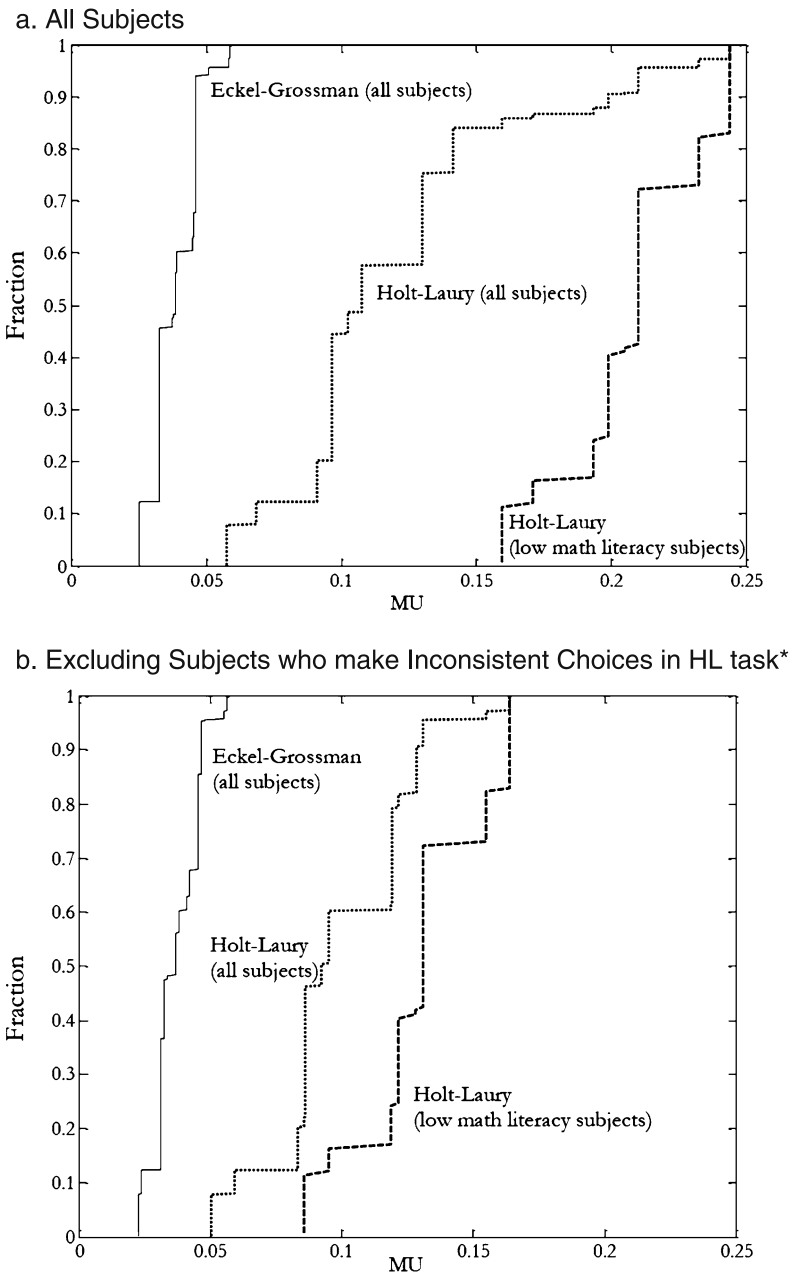
All else equal, if we interpret the noise parameter as mistakes or (more formally) as deviations from expected utility theory towards random behavior, a risk elicitation method that generates a smaller noise parameter should be preferred to any other method. In part because of its "coarseness" (only 5 categories as compared with 10 for HL) the price of less noise in the simpler EG measure may come in the form of a reduced accuracy in its ability to predict subjects' actual choices.

We employ a commonly used measure of *predictive accuracy*: the fraction of choices that are correctly predicted with the estimated parameters. To illustrate, the risky choice (i.e. choosing the right gamble) is predicted when the estimated probability in Eq. (1) is greater than 0.5; the less-risky choice (left gamble) is predicted otherwise. Computing the fraction of actual choices that are correctly predicted generates a measure that is bounded between zero and one, with a larger number representing a better fit (or predictive accuracy). Using the estimates from Table 6, the HL measure has a predictive accuracy of 0.84 for all individuals, 0.76 for low math skills subjects and 0.85 for high (i.e. non-low) math skills subjects.[16] The EG measure has a predictive accuracy of 0.72 for all individuals (predictive accuracy does not vary importantly by math skills).

At first, these values would suggest that the HL measure always generates a superior predictive accuracy than the EG measure. However, as pointed out previously, the two measures are different on a few dimensions. Importantly for the predictive accuracy measure, the larger number of choices and the different range of the implied CRRA in the HL task may give it an advantage. To see this, consider the summary statistics presented in Table 3. It is clear that very few subjects (43 or 5.33%) choose to switch to the risky gamble before Decision 5 (i.e. 4 safe choices). One obvious reason for this is the negative implied CRRA for these choices; EG implied CRRA is almost always strictly positive. Also, only very few subjects (4 subjects, or 0.5%) choose ten safe choices (not shown in Table 3). Intuitively, the

[15] For medium and high literacy people in HL, the range of predicted noise is closer to that of EG but still with cumulative distributions to the right of that of EG (not shown).

[16] These predictive accuracy values are nearly identical when inconsistent subjects are removed (i.e. using Table 7 estimates).



* An inconsistent subject is one that, as he/she moves down the decisions on the HL task, switches back to the safe gamble (A) after having chosen a risky gamble (B).

Fig. 2 Cumulative Distribution of the Noise Parameter by Instrument and by Math Skills almost deterministic behavior in these two ends of the distribution makes prediction 'easy' for these choices, and so improves the apparent predictive accuracy of HL over the full range of choices (subjects' choice in the EG measure is much less deterministic for any given gamble, see Table 2).

In absence of the ideal experimental data (an HL measure with 5 decisions with similar implied CRRA ranges to those of EG) that allows a direct comparison of predictive accuracy between the two measures, we "collapse" the ten HL decisions into five, effectively making the two sets of decisions more comparable. This can be easily done by excluding 5 decisions from the estimation; this procedure exactly transforms the 10 observed choices into the 5 choices that would have been observed if only these 5 decisions had been shown to subjects. Of course, this means that the implied CRRA ranges are modified into (typically) larger ranges. We carried out several estimations of (1), eliminating a different set of 5 decisions each time; the sets we eliminated tended to include the initial and final decisions in the HL task. Figure 3a and b display the predictive accuracy by decision when we exclude Decisions 1, 2, 4, 9 and 10 to estimate (1) with HL; Figure 4a and b depict the predictive accuracy with the 5 (implied) decisions in EG.[17] Clearly, the corresponding CRRA ranges for the 5 collapsed HL decisions are much closer to those corresponding to the 5 (implied) decisions in EG thus allowing a fairer comparison.

To maximize comparability across the two measures, the following predictive accuracy analysis excludes inconsistent subjects. The more comparable 5-decision HL measure now produces a predictive accuracy value of 0.76 for all subjects and 0.68 for low math skills subjects.[18] These values appear to suggest that the EG measure may be unambiguously preferred for low math skills subjects as it generates lower noise and a slightly better predictive accuracy than the HL measure, over comparable decisions. However, we interpret this result with caution as we cannot conduct formal statistical tests on predictive accuracy differences. Despite this limitation, it is interesting to investigate what the math score used as the cut-off for the low math score indicator variable would need to be to generate an equal predictive accuracy for both EG-all-subjects and HL-low-math-skills-subjects. For our sample, this occurs at the 25th percentile in the math literacy test. For lower scores, EG does better.

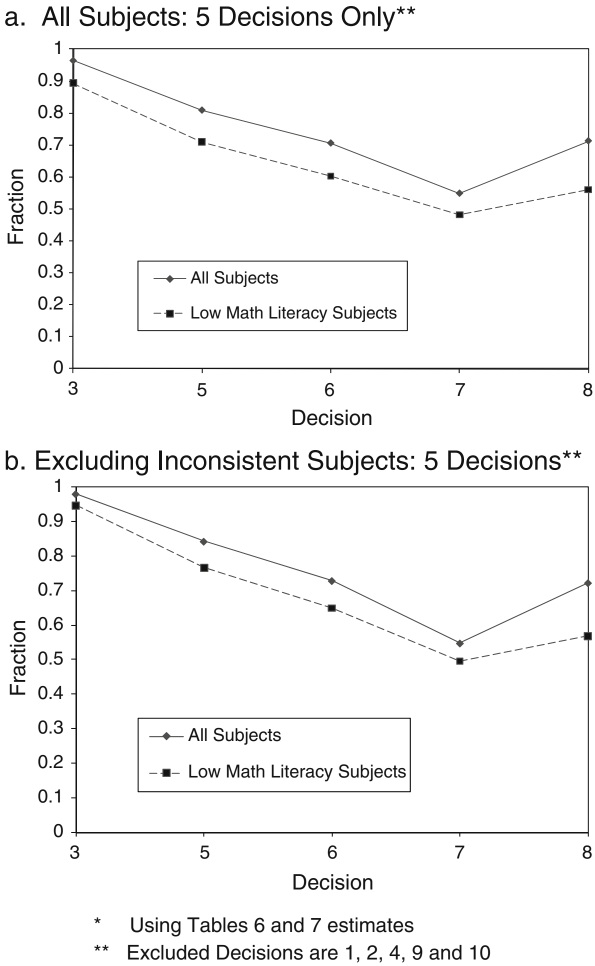
5 Conclusion and discussion

In this paper we set out to examine the trade-offs that arise when choosing among methods used to measure risk aversion when the measures may require different degrees of mathematical skills to be properly understood. We make use of a unique field experiment that included a mathematical skills assessment to address this central question. We find that a measurement method that is more difficult to understand is sometimes superior: the finer, more complex measurement method

[17] If, instead, Decisions 1, 2, 3, 4 and 10 are excluded, a similar shape is obtained.

[18] Predictive accuracy is inferior if Decisions 1–4 and 10 are excluded: 0.74 for all individuals and 0.63 for low math ability people. The reason for this is that there is a higher predictive accuracy for Decision 3 than for Decision 9. Thus, excluding Decisions 1, 2, 4, 9 and 10 is a relatively conservative way to collapse HL into 5 decisions.

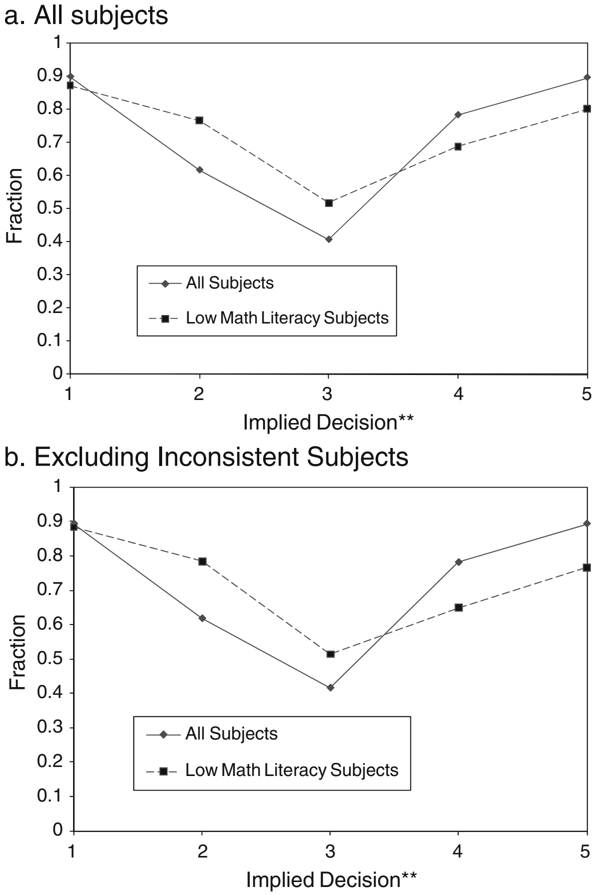
Fig. 3 Fraction of Decisions that are Correctly Predicted in HL Task*



developed by HL appears to be better suited for subjects with higher levels of mathematical skill. However, this finer measure may sacrifice too much in the way of comprehensibility (as measured by a smaller noise in decisions) and predictive accuracy if it is used to estimate risk preferences for subjects with low math skills. Thus, for less able subjects, care must be taken to design experiments that are easily accessed and comprehended. Further, we find some evidence that the simpler task may generate risk preference estimates that are more stable across time. An overall conclusion is that economists should be concerned with the elicitation method when investigating risk preferences, and realize that the 'ruler' used to measure risk preferences can have an important impact on the elicited preferences.

Our analysis of the trade-offs between predictive accuracy and noise is less than ideal. However, we believe that our results (strengthened by our anecdotal experience in the field) can guide researchers when choosing a risk instrument. For people with high math skills, our analysis suggests that the preferred instrument should be HL as it has better predictive accuracy than EG and there is no statistically

Fig. 4 Fraction of (implied) Decisions that are Correctly Predicted in EG Task*



* Using Table 6 estimates

** See Section 2.4 for details on how implied decisions are obtained

significant difference in the median noise. On the other hand, the EG measure appears to be an unambiguously better instrument (in terms of smaller noise and better fit) for those individuals in our sample who have low mathematical skills. Of course, the choice of instrument will ultimately depend on whether the researcher's premium is on higher predictive power or less noisy choices.

Our results also support the mounting evidence that there is an important degree of preference heterogeneity in the population that needs to be incorporated into how economists think about risk preferences. However, results suggest that different elicitation methods may produce different results on preference heterogeneity. This is a result that has not been discussed so far as it was not part of our original question. However, it is important that we spend some time discussing it as it raises important issues.

While economists' conception of risk attitudes suggests that any experimental measure designed to gauge the curvature of a utility function should produce the same result, differences in cognitive skills may hamper subjects' ability to reveal their true preferences via an experimental task. In fact, we find that the two tasks yield different risk preference estimates, but we are cautious in interpreting this difference. First, we do not know the true underlying risk preferences, so it is not possible to know which measure is more accurate. Second, the difference in the estimated risk parameter may be due to other, as yet unidentified, aspects of the tasks. While we leave a definitive answer of this question for future research, it is important to note that several researchers have found a relationship between cognitive skills and measured risk aversion. Dohmen et al. (2010) measure risk and time preferences using a representative sample of 1,000 German adults, and find a correlation with the score on a widely-used IQ test. Burks et al. (2009) note a strong relationship between cognitive skills and risk and time preferences using a sample of subjects in a trucking firm, whereas Huck and Weizsacker (1999) note a positive relationship between lottery complexity and risk aversion.[19] Our point is that a real correlation between numerical proficiency and risk aversion can be obscured by a task that produces different measures for low-numerate subjects, as well as the reverse: low numeracy can produce an effect that looks like risk aversion.

The results suggest that a simpler, more intuitive measure may provide better accessibility, and so more accurate measures of risk aversion, for subjects with low levels of analytic proficiency. However, we think it is important to distinguish between general intelligence and math skills per se. Peters et al. (2006) examine the specific effect of math skills (as distinct from IQ) and find that more highly skilled subjects are less subject to framing effects. Higher math skills should produce more consistent results across measurement methods. Methods that are designed so that low math skills persons can understand and complete them are more likely to allow researchers to find any real correlation between ability and risk attitudes.

Experimental research is ideal for sorting out these issues. One strategy, for example, is to adapt the HL measure to make it simpler and easier for low math subjects to process. Johnson et al. (2007) report results of an experiment using a new interface that presents the HL choices to subjects in a visual, one-at-a-time format. Subjects choose between two circles representing the two gambles, with shaded pie-slice areas representing probabilities and images of stacks of money for payoffs. They report a substantial increase in consistency. Recent refinements of EG also include visual displays of the gambles and images of money. When these visual representations were used with high school students, we found no relationship between math skills and risk attitudes (Eckel et al. 2007). These results further illustrate the importance of the task interface for collecting accurate preference information.

Acknowledgements We thank Glenn Harrison who generously provided his advice and code; Scott Murray and Fernando Cartwright provided valuable assistance. Comments from participants in workshops and conferences (ASSA, ESA, Dallas Federal Reserve Board, Georgia State University, Texas A&M University, and University of Arizona) substantially improved the paper. The experiments were funded by Canada Student Loans Program Directorate and the Applied Research Branch of Human Resources Development Canada (HRDC). Statistics Canada provided the numeracy assessment. Support for the research was provided by the John D. and Catherine T. MacArthur Foundation, Network on Preferences and Norms. Eckel was supported by the National Science Foundation (SES-0443708, SES-0094800).

[19] See also Eckel 1999 who shows that grade point average is strongly related to deviations from risk neutrality, providing early evidence that cognitive skills may affect elicited risk preferences.

Appendix: Decision forms for risk choices

Eckel-Grossman Risk Task

For **Decision 41** you will select from among six different gambles the one gamble you would like to play. The six different gambles are listed below.

- You must select one and only one of these gambles.
- To select a gamble place an **X** in the appropriate box.

Each gamble has two possible outcomes (ROLL LOW or ROLL HIGH) with the indicated probabilities of occurring. Your compensation for this part of the study will be determined by:

- which of the six gambles you select; and
- which of the two possible payoffs occur.

For example, if you select Gamble 4 and ROLL HIGH occurs, you will be paid $52. If ROLL LOW occurs, you will paid $16.

For every gamble, each ROLL has a 50% chance of occurring.

At the end of the study, if **Decision 41** is randomly selected, you will roll a ten-sided die to determine which event will occur. If you roll a 1, 2, 3, 4 or 5, ROLL LOW will occur. If you roll a 6, 7, 8, 9 or 0, ROLL HIGH will occur.

Decision 41

Mark your gamble selection with an **X** in the last box across from your preferred gamble.

	ROLL	Payoff	Chances	Your Selection **Mark only one**
Gamble 1	LOW	$28	50%	
	HIGH	$28	50%	
Gamble 2	LOW	$24	50%	
	HIGH	$36	50%	
Gamble 3	LOW	$20	50%	
	HIGH	$44	50%	
Gamble 4	LOW	$16	50%	
	HIGH	$52	50%	
Gamble 5	LOW	$12	50%	
	HIGH	$60	50%	
Gamble 6	LOW	$2	50%	
	HIGH	$70	50%	

Holt-Laury Risk Task

In this next set of 10 decisions, you are given a chance to earn a cash prize today. For each decision, you will choose between playing two Gambles, A and B. Here is an example:

Gamble A		Your Choice A or B	Gamble B	
3/10 of $40 (roll: 1,2,3)	7/10 of $32 (roll:4,5,6,7,8,9,0)		3/10 of $77 (roll: 1,2,3)	7/10 of $2 (roll:4,5,6,7,8,9,0)

Each Gamble is composed of two outcomes. Which one occurs depends on the roll of a ten-sided die. For instance, let's look at Gamble A. You have 3 out of 10 chances to win $40 and 7 out of 10 chances to win $32. If you roll a 1, 2 or 3, (3 chances out of 10) then you win $40. If you roll a 4,5,6,7,8,9,0, (7 chances out of 10) then you win $32.

	OPTION A		Your Choice, **A or B**	**OPTION B**	
Decision 1	1/10 of $40	9/10 of $32		1/10 of $77	9/10 of $2
Decision 2	2/10 of $40	8/10 of $32		2/10 of $77	8/10 of $2
Decision 3	3/10 of $40	7/10 of $32		3/10 of $77	7/10 of $2
Decision 4	4/10 of $40	6/10 of $32		4/10 of $77	6/10 of $2
Decision 5	5/10 of $40	5/10 of $32		5/10 of $77	5/10 of $2
Decision 6	6/10 of $40	4/10 of $32		6/10 of $77	4/10 of $2
Decision 7	7/10 of $40	3/10 of $32		7/10 of $77	3/10 of $2
Decision 8	8/10 of $40	2/10 of $32		8/10 of $77	2/10 of $2
Decision 9	9/10 of $40	1/10 of $32		9/10 of $77	1/10 of $2
Decision 10	10/10 of $40	0/10 of $32		10/10 of $77	0/10 of $2

References

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583–618.

Barr, A., & Genicot, G. (2008). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association, 6*(6), 1151–1185.

Becker, G., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*, 226–232.

Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academies of Science, 102*, 4209–4214.

Binswanger, H. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics, 62*, 395–407.

Binswanger, H. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural India. *Economic Journal, 91*, 867–890.

Bohm, P., Linden, J., & Sonnegard, J. (1997). Eliciting reservation prices: Becker-DeGroot-Marschak mechanisms vs. markets. *Economic Journal, 107*, 1079–1089.

Burks, S., Carpenter, J., Götte, L., & Rustichini, A. (2009). Cognitive skills explain economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences, 106*(19), 7745–7750.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260.

Eckel, C. (1999). Commentary on “The effects of financial incentives in experiments: A review and capital-labor production framework.” *Journal of Risk and Uncertainty, 19*(1–3), 47–48.

Eckel, C. (2007). Experiments on gender differences. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). Basingstoke: Palgrave Macmillan.

Eckel, C., & Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*(4), 281–295.

Eckel, C., & Grossman, P. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization, 68*(1), 1–17.

Eckel, C., Grossman, P., Johnson, C., Milano, A., Rojas, C., and Wilson, R. (2007). Explaining risk preferences in high school students: A preliminary look. CBEES Working paper, University of Texas at Dallas.

Eisenberger, R., & Weber, M. (1995). Willingness-to-pay and willingness-to-accept for risky and ambiguous lotteries. *Journal of Risk and Uncertainty, 10*, 223–33.

Engle-Warnick, J., Escobal, J., and Laszlo, S. (2005). The effect of an additional alternative on measured risk preferences in a field experiment. Unpublished manuscript. Experimental 0511003, EconWPA.

Harbaugh, W., Krause, K., & Vesterlund, Lise. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics, 5*(1), 53–84.

Harbaugh, W. T., Krause, K., & Vesterlund, L. (2010). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal, 120*(545), 595–611.

Harrison, G. (1986). An experimental test for risk aversion. *Economics Letters, 21*, 7–11.

Harrison, G. (1992). Theory and misbehavior of first-price auctions: Reply. *American Economic Review, 82*, 1426–1443.

Hey, J., & Orme, C. (1994). Investigating generalizations of expected utility using experimental data. *Econometrica, 62*(6), 1291–1326.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655.

Horowitz, J. (2006). The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters, 93*(1), 6–11.

Huck, S., & Weizsacker, G. (1999). Risk, complexity, and deviations from expected value maximization. *Journal of Economic Psychology, 20*, 699–715.

Isaac, M., & James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty, 20*(2), 177–187.

Johnson, C., Montmarquette, C., and Eckel, C. (2003). Fostering adult education: A laboratory experiment on the efficient use of loans, grants, and saving incentives. Social Research and Demonstration Corporation.

Johnson, C., Eckel, C., and Engle-Warnick, J. (2007). An adaptive instrument to elicit risk preferences. Working paper, University of Montreal, 2007.

Kachelmeier, S., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review, 82*, 1120–1141.

Karni, E., & Safra, Z. (1987). Preference reversal and the observability of preferences by experimental methods. *Econometrica, 55*(3), 675–685.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C., Mazzocco, K., & Dicker, S. (2006). Numeracy and Decision Making. *Psychological Science, 17*(5), 407–413.

Slovic, P. (1962). Convergent Validation of Risk Taking Measures. *Journal of Abnormal and Social Psychology, 65*(1), 68–71.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332–382.

Statistics Canada and OECD. (2003). Learning a Living: First Results of the Adult Literacy and Life Skills Survey. http://www.statcan.ca/english/freepub/89-603-XIE/2005001/pdf.htm