Environ Resource Econ (2008) 39:481–495
DOI 10.1007/s10640-007-9137-0

Cognitive ability and scale bias in the contingent valuation method

An analysis of willingness to pay to reduce mortality risk

Henrik Andersson · Mikael Svensson

Received: 12 July 2006 / Accepted: 12 May 2007 / Published online: 21 June 2007

Abstract This study investigates whether or not the scale bias found in contingent valuation (CVM) studies on mortality risk reductions is a result of cognitive constraints among respondents. Scale bias refers to insensitivity and non-near-proportionality of the respondents' willingness to pay (WTP) to the size of the risk reduction. Two hundred Swedish students participated in an experiment in which their cognitive ability was tested before they took part in a CVM-study asking them about their WTP to reduce bus-mortality risk. The results imply that WTP answers from respondents with a higher cognitive ability are less flawed by scale bias.

Keywords Cognitive ability · Contingent valuation · Mortality risk · Near-proportionality · Scale bias

JEL Codes D80 · I10 · Q51

1 Introduction

When knowledge among analysts about the decision alternatives and the consequences decision makers face is limited, or when market data does not exist for the amenity of interest, stated-preference (SP) methods have an important role to play in evaluating a wide range of non-marketed amenities. One such SP-method is the contingent valuation method (CVM) (Bateman et al. 2002). CVM has, however, been subject to heavy criticism for being inadequate to measure individual preferences (Diamond and Hausman 1994). Much of the

H. Andersson (✉)
Department of Transport Economics, Swedish National Road and Transport Research Institute (VTI), P.O. Box 55685, Stockholm 102 15, Sweden
e-mail: henrik.andersson@vti.se

M. Svensson
Department of Business, Economics, Statistics and Informatics, Örebro University, Örebor, Sweden

criticism, in addition to that against the hypothetical bias (Hanley and Shogren 2005; Harrison 2006), has been levelled at the lack of scale sensitivity found in many CVM-studies (e.g. Hammitt and Zhou 2006). Whereas the advocates of CVM consider the lack of scale sensitivity to be a result of bad survey design (Carson and Mitchell 1995; Carson et al. 2001; Smith 1992), its opponents argue that it is not capable of eliciting individual preferences (Desvousges et al. 1993; Daimond and Hausman 1994; Kahneman et al. 1999).[1]

When individuals' preferences for a reduction of their own mortality risk are elicited, respondents of CVM-studies are usually asked to state their willingness to pay (WTP) for a small reduction in the probability of death. Since individuals will personally benefit from the risk reduction, we expect them to be willing to pay more for a larger risk reduction than for a smaller one. Moreover, a necessary (but not sufficient) condition for WTP answers in CVM-studies to be valid estimates of individuals' preferences for small mortality risk reductions is that WTP is near-proportional (increasing and strictly concave) to the size of the risk reduction (Hammitt 2000), which makes this good especially useful to study scale sensitivity.[2] Hammitt and Graham (1999) analyzed scale sensitivity in 25 CVM-studies and found that WTP increased with the size of the risk reduction in most of them. In none of the studies was the increase in WTP near-proportional to the risk reduction. In addition, they also performed a scale sensitivity test on their own data set in the study, and found that respondents, with higher confidence in their answers, gave answers that were close to theoretical predictions. This is in line with the results in Corso et al. (2001) and Alberini et al. (2004) where proportionality could not be rejected for "better informed" and "more confident about their answers" respondents, respectively.

Psychology literature has shown that we often make decisions based on heuristics (short cuts), since mental short cuts lighten the cognitive burden of decision-making (Kahneman et al. 1982; Kahneman 2003). Eliciting individuals' preferences for reductions in the probability of death seems to be cognitively constraining for the respondents (Carson et al. 2001), and we investigate whether or not cognitive constraints can explain the lack of scale sensitivity found in CVM-studies on mortality risk reductions. It is important to shed light on this issue considering that surveys have to consider a possible tradeoff between conducting statistically efficient surveys and surveys that respondents can understand (Johnson 2006). We assume that individuals with lower cognitive ability have a higher share of decision-making driven by heuristics, which results in less well-considered answers. Our hypothesis is that respondents with a higher cognitive ability are more able to understand the probability changes and, therefore, state a WTP that is more proportional to the magnitude of the risk reduction and thus more in line with the theoretical predictions. The aim here is to test this hypothesis, by letting a sample of Swedish students take part in an experiment where, prior to answering WTP questions, they take part in a test, the result of which is used as a proxy for cognitive ability. In the CVM survey the students are asked about their WTP to reduce the mortality risk they face when riding the local bus, a risk we assume most respondents are familiar with.

The paper is structured as follows. Our experiment is described in Sect. 2. Section 3 then contains the data that was collected and the regression results, which reveal a positive correlation between cognitive ability and WTP answers less flawed by scale bias. This is especially true for respondents with the highest score in the test of cognitive ability and for respondents

[1] Heberlein et al. (2005) provide cautions against scale-sensitivity as the sole validity test of CVM studies. They find that scale insensitivity can be in line with economic theory for non-use goods.

[2] Other validity tests are, e.g., that WTP should be increasing in wealth and baseline risk (Jones-Lee 1974; Weinstein et al. 1980).

with good skills in probability laws, compared with respondents only showing good skills in pure computational exercises. Section 4 offers a discussion and some concluding remarks. Our findings confirm that knowledge and understanding of the good in CVM-studies are important in order to get valid WTP estimates. Part of the scale insensitivity found in CVM-studies may, therefore, be regarded as a methodological problem that can be alleviated with more research.

2 The experiment

The experiment was conducted on 200 undergraduate students at Karlstad University in Sweden in the fall of 2005.[3] The students were informed in class that they would be given the opportunity to take part in an experiment to provide valuable information for government authorities within the transport sector. They were not informed that the aim was to examine how WTP is affected by cognitive skills, since we suspected that it could increase the number of protest answers. The students were informed that the experiment would take place after the next lecture, participation would be voluntary, and they would receive SEK 50 (ca. USD 7) as compensation for participation.[4] Because only one person supervised the whole experiment, we do not expect heterogeneity in the responses due to undue influence exerted by different interviewers.

The experiment consisted of three parts: (i) the test of cognitive ability, (ii) the CVM survey, and (iii) follow-up questions on demographics, socio-economics and some questions related to academic credits and "traffic experience". All of these are described below.[5]

2.1 Cognitive ability

In the cognitive ability test the respondents were given 5 min to answer 17 questions. The test was not in any way a complete cognitive ability test, but rather it focused on the respondents' skills in thinking about probabilities, syllogisms and computation, which have been deemed important in related studies (Benjamin and Shapiro 2005).[6] Recent research has shown, however, that simple cognitive ability tests, with only a few questions, can work as well as longer tests, such as the *Wonderlic* or *SAT* tests (Frederick 2005).

The answers given in our test are assumed to depend on the respondents' underlying cognitive abilities, and factor analysis is employed to explore if our test of cognitive ability measured different aspects of cognitive processes. Since the respondents' cognitive abilities are not directly observable, they are defined as latent variables, and, when there are several indicator variables of the latent variable, factor analysis can be used to study the correlations among a large number of variables (Barthomolew 1987; Lawley and Maxwell 1963). We use

[3] The experiment included students from the following disciplines; business administration, economics, human resources, teaching and political science. Most of these students were business majors (70%). A preliminary analysis showed that the area of study had no significant influence on either the cognitive test or the WTP answers, and the analysis based on area of study is, therefore, omitted.

[4] All prices in the paper are 2005 prices. USD 1 = SEK 7.48 (December 2005, www.riksbanken.se, 01/27/06.)

[5] More detailed information about the experiment, the CVM-survey design, etc., is available upon request from the authors. For a more comprehensive description of the experiment we refer to the working paper version of the study (Andersson and Svensson 2007).

[6] Our test is a crude measure of cognitive ability, and content areas such as verbal reasoning, figural reasoning or short-term memory, which are also thought to be important aspects of mental and cognitive ability, are not tested. The questions used can be found in previous experiments and are similar to those used in intelligence tests (Frederick 2005; Kahneman and Tversky 1972, 1983; Rabin 2002).

this information in Sect. 3, to create variables based on the factors that are significant in the factor analysis.

In the follow-up part we asked the respondents about how many academic credits they had at the time of the survey and how many semesters they had studied. Using this information we can estimate credits per semester, our proxy for ambition and motivation (ambitious students can increase the number of credits taken per semester), which enables us to ascertain whether those with more credits per semester are more eager to "do well" in the experiment, and thus provide us with WTP answers more in line with theoretical predictions.[7]

2.2 Contingent valuation survey

In the second part of the experiment the respondents were asked open-ended questions to show how much they were willing to pay to reduce the risk of being involved in a fatal bus accident.[8] This mortality risk was chosen for two reasons: (i) we assumed it to be a risky activity familiar to the students and one many of them are exposed to,[9] and (ii) it is an activity that the students cannot influence by their own skills (cf. driving).[10] Moreover, since we wanted to avoid a framing effect by letting respondents compare new bus companies with an existing one, all were presented as new companies, but with one as the reference.[11]

Before being given the risk scenario and the WTP questions, the respondents were informed of the annual risk of dying for an individual in his/her 50s and the annual risk of dying in a road-traffic accident.[12] In order to visualize these probabilities, a graph paper containing 10,000 squares was used, where the appropriate number of squares had been blacked out to represent each mortality risk. The respondents were also told that, in a city of the size of Karlstad, eight individuals will on average die annually in road-traffic accidents. The combination of a visual aid and verbal probability analog was used since it has proved to provide answers more consistent with standard economic theory (Corso et al. 2001).

Each respondent was faced with two WTP questions, one for the smaller risk reduction $\Delta p^{\mathrm{S}} = 4 \cdot 10^{-5}$, and one for the larger $\Delta p^{\mathrm{L}} = 6 \cdot 10^{-5}$. Thus, the ratio between the risk reductions was equal to 1.5 for all respondents. To control for order effects (starting point bias), half of the respondents first answered the question on the smaller risk reduction, while the order was reversed for the other half. We also wanted to test if there was any framing

[7] It has been found that student self-discipline better predicts grades than student IQ (Duckworth and Seligman 2005; Wolfe and Johnson 1995).

[8] Bus fatalities constitute ca. one percent of all road accident fatalities (SIKA 2007). We decided to use the open-ended format to avoid anchoring effects (Green et al. 1998). Moreover, in order to examine if scale bias is influenced by preference certainty, each WTP question was followed by a qualitative "certainty question" (see, e.g., Blumenschein et al. 2001). We did not find that scale sensitivity is influenced by preference certainty and the discussion of these results is kept to a minimum.

[9] The campus at Karlstad University is located approximately 9 km (ca. 5.6 miles) from the city center and the students often use the bus in order to get to/from the university and/or city center.

[10] Bus-mortality risk is not a pure private good, and there is a risk that values also represent paternalistic altruism (safety orientated) or that respondents answer strategically when they state their WTP for that kind of risk reduction. Note that, since we conducted internal tests on scale sensitivity, our findings should not be affected by potential strategic bias, altruism, or whether stated WTP also reflects preferences to reduce injury risk (the respondent's answers should be affected to the same relative degree).

[11] The survey asks about the city council selecting from one of three bus companies that are distinguished by the levels of safety, when contracting out the service to a private firm. This corresponds to the actual procedure in the city of Karlstad as well as most other Swedish cities.

[12] We decided to use a baseline risk of an age group other than the respondents' own. The risk of a 50-year-old was used by both Persson et al. (2001) and Carlsson et al. (2004), which is why we decided to use the risk of that particular age group.

effect from the initial bus fare, and half of the sample were, therefore, told that the annual bus fare of the reference bus company was SEK 3,000 and the other half that it was SEK 4,200.[13] Thus, the sample consisted of four subgroups divided according to the order of the WTP question and the amount of the annual bus fare.

3 Results

The first section contains the descriptive statistics and variable names of the data collected in the experiment, as well as the results of the factor analysis of the cognitive ability test. The section thereafter shows the regression results of WTP and the analysis of scale sensitivity. Scale sensitivity can be divided into two validity tests, “weak” and “strong” (Corso et al. 2001), where weak sensitivity is fulfilled if WTP increases with the size of the risk reductions, and strong sensitivity when WTP is near-proportional to the risk reduction. Scale sensitivity can be tested on group (external test) and individual level (internal test). In the following sections we will perform both tests, but mainly focus on the internal test.

3.1 Descriptive statistics

Variables used in the analysis and descriptive statistics are found in Table 1. The variables *Weak sensitivity*, *Proportion* and *Absolute deviation* indicate how well, according to our hypothesis, respondents answered. *Weak sensitivity* is a dummy variable which takes the value one if respondents state a higher WTP for the larger risk reduction than for the smaller one. *Proportion* is a measure of the ratio between the WTP for the larger and smaller risk reductions, and is equal to 1.5 for 14% of the sample and has a mean value of 1.58, which is close to proportionality. *Absolute deviation* is estimated as the absolute difference between the WTP ratio and risk-reduction ratio. For instance, since the ratio between the risk reductions is 1.5, a WTP ratio of 1.75 or 1.25 will mean an absolute deviation of 0.25.

The factor analysis indicates three factors, from which we create three variables: (i) *Probabilities*, from four questions, three of them primarily measuring processes of understanding the logic of probabilities (Bayes’ theorem and law of large numbers) and the fourth question measuring general logic,[14] (ii) *Computational*, based on two questions measuring pure computational ability, or algebra, and (iii) *Intuition* which is a single question and included as a binary variable.

In Table 2 the mean estimates of the WTP answers to the first question show that the respondents had a higher WTP for the larger risk reduction than for the smaller one.[15] Moreover, the results reveal no statistically significant framing effect of the different bus fares on mean WTP. However, the results indicate a starting point bias. Respondents who first stated their WTP for the larger risk reduction (*Large Risk*), stated a higher WTP for the smaller risk reduction than those respondents who were first asked about their WTP for the smaller risk reduction, and vice versa.[16] The differences are not statistically significant, though.

[13] The annual bus fare at the time of the survey in Karlstad was SEK 3,690.

[14] The qualitative results of scale sensitivity in Sect. 3.2 also hold if *Probabilities* only includes the first two and/or three of the questions instead of all four.

[15] Fourteen of the respondents stated a zero WTP for both risk reductions. The follow-up questions for the zero-answers indicated that 12 zero-answers were “true” (“risk reduction too small” or “cannot afford to pay”) and that two were protest answers (“question unclear” or “WTP should not be used”).

[16] This is contrary to what has been found for environmental public goods, where the WTP for the smaller quantity was larger if stated first (for references, see Clark and Friesen 2006).

Table 1 Description of dependent and explanatory variables

Variable name	Description	Mean (Std. Dev.)	*N*
Cognitive	Score on the cognitive test with a max of 17[a]	9.83 (2.12)	200
Proportion	Ratio between respondent's WTP for the large and the small risk reduction	1.58 (0.93)	186
Absolute deviation	Deviation in absolute terms from linear proportionality	0.49 (0.79)	186
Weak sensitivity	Dummy variable coded as one if respondent shows scale sensitivity and zero otherwise	0.66 (0.48)	200
Risk help	Dummy variable coded as one if respondent stated that risk illustration was helpful and zero otherwise	0.24 (0.43)	200
Bus	Dummy equal to one if respondent frequently travels by bus and zero otherwise	0.55 (0.50)	199
Bus accident	Dummy equal to one if respondent has personal experience of a bus traffic accident and zero otherwise	0.12 (0.32)	199
Age	Age of the respondents	23.01 (3.94)	199
Female	Dummy equal to one if respondent is female and zero otherwise	0.54 (0.50)	199
Credits/semester	Earned academic credits per semester[b]	18.87 (4.72)	164
Income	Income based on medians from 8 different income categories. (2005 prices, USD 1 = SEK 7.48)	6,836 (2,887)	198
Sure	Dummy equal to one if respondent stated definitely or probably sure in the WTP certainty question	0.76 (0.43)	198
Factor analysis[c]			
Probabilities	Questions related to logic and probability laws	1.15 (0.85)	4
Computational	Questions related to algebra	1.52 (0.68)	2
Intuition	Question intended to measure a process driven by intuition and instant reaction	0.10 (0.30)	1

[a] The lowest and highest score was 4 and 16

[b] An academic semester corresponds to 20 academic credits

[c] *N* for factor analysis refers to number of questions

The relationships between *Cognitive* and *Weak sensitivity* on the one hand, and *Absolute deviation* on the other, are examined in Table 3.[17] The highest scoring group has the highest share of respondents showing weak scale sensitivity and the lowest absolute deviation from linear proportion. For instance, in the specification with two groups, 77% and 59% show weak scale and the deviation is 0.38 and 0.57 for the high (11–16) and low (4–10) score groups, respectively, differences that are statistically significant. Further, the share of respondents showing weak scale decreases and the deviation increases in size for groups with lower cognitive scores. A comparison of geometric means also shows that the results are not sensitive to outliers. Moreover, there is no clear pattern between *Cognitive* and *Credits/semester*. Partial correlation analysis between *Cognitive* and *Credits/semester* shows no statistically significant correlation either.

[17] The intention was to divide the sample into equal-sized groups based on the test result. However, the test score is made up of integers, which is why the group sizes differ.

Table 2 Willingness to pay

	WTP Q1		WTP Q2	
	Δp_1^S	Δp_2^L	Δp_1^L	Δp_2^S
High fare	492* (674)	976 (1,642)	709*** (943)	754*** (1,571)
Low fare	539 (750)	669 (815)	678*** (821)	578 (901)
Mean	515** (709)	822 (1,299)	666*** (1,277)	693*** (880)

WTP Q1 and WTP Q2 refer to the first and second WTP questions

Superscripts S and L refer to small and large risk reductions, and subscripts 1 and 2 refer to sample group

Standard deviations in parentheses

$H_0 : \Delta p^S = \Delta p^L$; **significant at 0.05 level, *significant at 0.10 level

$H_0 : \text{WTP}_i\ \text{Q1} = \text{WTP}_i\ \text{Q2}$; ***significant at 0.01 level, $i \in \{1, 2\}$

Framing effect of different bus fares not statistically significant

WTP Q1 and WTP Q2 for same Δp, and given the same bus fare, not statistically significantly different

Wilcoxon rank-sum employed for mean comparisons

Table 3 Cognitive score, weak sensitivity, deviation from proportionality and credits per semester

Cognitive score	Weak sensitivity		Absolute deviation		Credits/semester	
	Mean	*N*	Mean	*N*	Mean	*N*
12–16	0.81[b]	43	0.37[e]	42	19.55	34
11	0.70	30	0.40	30	18.26	22
9–10	0.62	77	0.55	68	18.47	65
4–8	0.54	50	0.58	46	19.24	43
9–11[a]	0.64[c]	107	0.51[f]	98	18.42	87
11–16	0.77[d]	73	0.38[g]	72	19.04	56
4–10	0.59	127	0.57	114	18.78	108
Full sample	0.66	200	0.49	186	18.87	164

Values for Weak sensitivity; share of respondents showing weak scale sensitivity

[a] Cognitive score groups (CSG) 9–10 and 11 grouped together

[b] Statistically significantly different from CSG 4–8 and 12–16 at 0.01 and 0.05 level, respectively

[c] Statistically significantly different from CSG 4–8 and 12–16 at 0.05 and 0.01 level, respectively

[d] Statistically significantly different from CSG 4–10 at 0.05 level

[e] Statistically significantly different from CSG 4–8 at 0.05 level

[f] Statistically significantly different from CSG 4–8 and 12–16 at 0.1 and 0.05 level, respectively

[g] Statistically significantly different from CSG 4–10 at 0.1 level

Cognitive score and Credits/Semester not statistically significantly correlated

Wilcoxon rank-sum employed for mean comparisons

Respondents who stated WTP= 0 for both Δp excluded when examining Absolute deviation

Table 4 Regression results WTP to reduce bus-mortality risk

Variable	All respondents		Cognitive score			
			4–8	9–10	11	12–16
ln(Cognitive)	0.22 (0.34)	–	–	–	–	–
Probability	–	0.07 (0.10)	–	–	–	–
Computational	–	0.09 (0.12)	–	–	–	–
Intuition	–	−0.18 (0.27)	–	–	–	–
Large risk	0.42*** (0.16)	0.41** (0.16)	−0.01 (0.42)	0.50* (0.29)	0.33 (0.48)	0.69** (0.29)
Low fare	−0.21 (0.16)	−0.20 (0.16)	−0.13 (0.43)	−0.34 (0.29)	−0.19 (0.42)	0.08 (0.31)
Female	−0.08 (0.16)	−0.07 (0.17)	0.12 (0.41)	0.01 (0.29)	0.38 (0.62)	−0.27 (0.31)
ln(Age)	0.07 (0.31)	0.06 (0.31)	−1.01 (1.67)	0.11 (0.38)	−1.01 (1.58)	−0.35 (1.31)
ln(Income)	−0.17 (0.21)	−0.17 (0.21)	−0.45 (0.50)	0.20 (0.37)	0.01 (0.65)	−0.03 (0.46)
Intercept	6.56*** (2.05)	6.84*** (1.95)	12.84** (6.00)	3.69 (3.25)	7.76 (6.37)	7.38* (3.81)
N	185	185	45	68	30	42
R^2	0.08	0.09	0.10	0.11	0.06	0.19

Dependent variable, the natural logarithm of the WTP answer to the first WTP question

Bus, Bus accident, Risk help and Sure included as covariates in the regressions, but omitted from table. Besides Bus accident and Risk help, positive and negative in Cognitive score 12–16, coefficient estimates stat. insignificant (0.1 level)

Two-tailed test: ***significant at 0.01 level, **at 0.05 level, *at 0.1 level

Standard errors in brackets

Large risk and *Low fare* are dummies equal to one if the respondent in the first WTP question stated his/her WTP for the larger risk reduction and was faced with the lower bus fare

3.2 Regression analysis

This section analyzes, using multivariate OLS, how the respondents' first WTP answers are affected by different attributes. Multivariate probit and OLS regressions, with three different specifications of cognitive ability, are then run to test for weak and strong scale sensitivity: (i) *Cognitive*, (ii) the variables created from the factor analysis, *Probability*, *Computational* and *Intuition*, and (iii) to allow for non-linearity, *Cognitive* divided into four different categories based on test scores from Table 3, with the group with the lowest score as the reference group. Functional forms of the regressions were determined by goodness-of-fit and diagnostic tests.

The results of the WTP regressions are found in Table 4. In the two regressions on the whole sample, WTP for the larger risk reduction (*Large risk*) is statistically significantly higher than for the smaller reduction. The null hypothesis of proportionality means that the coefficient estimate for *Large risk* is not different from 0.405. The coefficient estimates are 0.42 and 0.41, which implies a 52% and 51% higher WTP for the larger risk reduction. We cannot reject the hypothesis that WTP is proportional to the size of the risk reduction in either of the regressions. All the other coefficient estimates of the covariates are statistically insignificant. This implies, among other things, that WTP is not related to cognitive ability, irrespectively of whether measured by *Cognitive* or by the variables *Probability*, *Computational* and *Intuition*.[18]

[18] For good measure, in this regression we also control for age, gender, and income, but age and income are not significant determinant of the WTP for the stated risk reductions. This is probably because of insufficient variation in age in our sample and mismeasured income.

Table 5 Regression results: Weak and strong scale sensitivity

Variable	Weak 1	Weak 2	Weak 3	Weak 4	Strong 1	Strong 2	Strong 3	Strong 4
Cognitive[a]	0.15*** (0.05)	–	–	–	−0.14** (0.05)	–	–	–
Probability	–	0.28** (0.13)	–	–	–	−0.06*** (0.02)	–	–
Computational	–	0.30** (0.15)	–	–	–	1.45 (0.03)	–	–
Intuition	–	0.35 (0.36)	–	–	–	0.08 (0.06)	–	–
Score 9–10	–	–	0.26 (0.24)	–	–	–	−0.04 (0.05)	–
Score 11	–	–	0.35 (0.32)	–	–	–	−0.09* (0.05)	–
Score 12–16	–	–	0.82*** (0.29)	–	–	–	−0.10** (0.04)	–
Credits/semester[a]	–	–	–	0.04* (0.02)	–	–	–	0.01 (0.04)
Large risk[b]	0.10 (0.20)	0.12 (0.20)	0.11 (0.19)	0.21 (0.21)	0.05 (0.04)	0.05 (0.03)	0.05 (0.03)	0.04 (0.04)
Low fare[b]	−0.10 (0.19)	−0.05 (0.20)	−0.05 (0.19)	−0.12 (0.21)	0.02 (0.04)	0.02 (0.04)	0.02 (0.03)	0.01 (0.05)
Female	0.13 (0.20)	0.17 (0.20)	0.14 (0.20)	0.18 (0.21)	0.06** (0.03)	0.07* (0.04)	0.07** (0.03)	0.06 (0.04)
Age[a]	−0.05 (0.03)	−0.05 (0.03)	−0.04 (0.03)	−0.07** (0.03)	0.02 (0.02)	0.03 (0.07)	0.03 (0.03)	0.01 (0.04)
Income[a]	$5 \cdot 10^{-5}$ $(4 \cdot 10^{-5})$	$4 \cdot 10^{-5}$ $(4 \cdot 10^{-5})$	$4 \cdot 10^{-5}$ $(4 \cdot 10^{-5})$	$4 \cdot 10^{-5}$ $(4 \cdot 10^{-5})$	0.04 (0.03)	0.04 (0.05)	0.04 (0.03)	0.02 (0.04)
Intercept	−0.51 (0.84)	0.28 (0.72)	0.66 (0.68)	0.50 (0.81)	0.47 (0.30)	0.18 (0.42)	−0.20 (0.29)	0.46 (0.35)
N	198	198	198	162	184	184	184	149
R^2 (Pseudo-R^2 in Weak)	0.08	0.09	0.07	0.05	0.06	0.09	0.06	0.03

Weak: Probit, dependent variable *Weak sensitivity*

Strong: OLS, dependent variable ln(1.5+*Absolute deviation*)

Two-tailed test: *** significant at 0.01 level, ** at 0.05 level, *at 0.1 level

Standard errors (robust in "Strong") in parentheses

[a] Natural logarithm of variables used in regressions on strong sensitivity

[b] Dummy equal to one if respondent in the first WTP question stated his/her WTP for the larger risk reduction and was faced with the lower bus fare

Bus, Bus accident, Risk help and Sure included as covariates in the regressions, but omitted from table. Besides Risk help, positive in Weak 1–3, coefficient estimates stat. insignificant (0.1 level)

To further examine the relationship between WTP and cognitive ability the sample is divided into four subgroups (the same groups as in Table 3) based on the respondents' test score. We find a statistically significant higher WTP for the larger risk reduction only in the groups with test scores 9–10 and 12–16. The coefficient estimates are 0.50 and 0.69, which imply that the WTP for the larger risk reduction is 65% and 100% higher in the 9–10 and 12–16 group, respectively. We cannot reject, however, that WTP is proportional to the size of the risk reduction. Thus, both groups pass the weak and strong scale sensitivity tests.

WTP has so far been subjected to external tests, we now move on to internal tests of scale sensitivity. The results of the regressions on weak and strong scale are shown in Table 5. A positive coefficient in the regression on *Weak sensitivity* (Weak 1–4) indicates a higher probability of showing weak scale sensitivity, i.e. stating a higher WTP for the larger risk reduction. The results in the first column show that a higher score on the cognitive test is associated with a higher probability of showing weak scale sensitivity. In the second specification, with cognitive ability divided into three variables, a higher score on the probability- as well as computational-related variable is positively related to showing weak scale sensitivity. In the third specification, when *Cognitive* is recorded into binary variables, we find that coefficient estimates are increasing with cognitive score, but only statistically significantly for the group with the highest test score. In other words, only the group with the best test score is statistically significantly more likely than the group with the lowest test score to show weak scale sensitivity.[19] In Weak 4 we see that the coefficient sign for *Credits/semester* is also positive and statistically significant.

Regarding strong scale, the results in Table 3 indicate a positive relationship between strong scale sensitivity and cognitive ability. The test for strong scale sensitivity is continued by examining the correlation between *Absolute deviation* and different explanatory variables. As mentioned in Sect. 1, WTP should be increasing and concave, i.e. *Proportion* $\in (1, 1.5)$, and near-proportional to the size of the risk reduction, i.e. close to 1.5. Restricting the sample to respondents with *Proportion* $\in (1, 1.5)$ only gives us 21 observations, however. Thus, very few respondents' WTP is strictly increasing and concave, and since there are no statistically significant relationships for this small group we analyze the deviation from proportionality for the whole sample. The dependent variable in the test for strong scale sensitivity is $\ln(1.5 + \textit{Absolute deviation})$, and the constant $(1.5 = \text{perfect proportion})$ is needed since we take logarithms. In the analysis, negative coefficient estimates indicate a more near-proportional WTP.

The same pattern as in the test for weak scale sensitivity appears in Strong 1–3. A higher cognitive score implies less deviation from proportionality. In the second specification we see that a higher score on *Probability* is significantly associated with a more near-proportional WTP. No significant association can be found for either *Computational* or *Intuition*. In the third specification we also see that groups with higher cognitive scores show lower deviation from proportionality, a deviation statistically significant for the two highest scoring groups. In the fourth specification, the coefficient sign for *Credits/semester* is not statistically significant. Moreover, the coefficient for *Female* is positive and statistically significant in three of the regressions, implying that male respondents stated a more near-proportional WTP.

[19] In Weak 1–3, respondents who stated that the risk illustration in the questionnaire was helpful when they answered the WTP questions (*Risk help*), are more likely to show weak scale sensitivity. An analysis of *Risk help* did not provide us with any clear answers to why the risk illustration was considered helpful. Among other things, we did not find that the risk illustration was helpful for those with better understanding of probabilities.

4 Discussion

The results of this study support our hypothesis, that more cognitively skilled respondents will give answers less flawed by scale bias. When analyzing both the descriptive statistics and the regression results, we find that respondents scoring higher in the cognitive ability test gave answers less flawed by scale bias. However, in the test for strong scale sensitivity only 21 out of 200 respondents (10%) stated a WTP that was increasing and strictly concave with with the size of the risk reduction, and we, therefore, treat positive as well as negative deviations from proportionality as equals. Accordingly, our strong scale sensitivity test should be seen as a test of deviation from proportionality, where we allow for positive "errors" in both directions, i.e. WTP ratios outside the 1–1.5 range.

An alternative hypothesis to the one tested in this study, i.e. that scale sensitivity is related to cognitive ability, is that our findings are a result of more motivated respondents being more eager to do well in the cognitive test and thinking over the WTP questions more carefully. However, on the basis of our analysis, we find indications that respondents' test scores on the questions demanding skills in probability laws, rather than e.g. pure computational ability, are most importantly related to scale sensitivity. That certain parts of the cognitive test, rather than the full test, are significantly associated with scale sensitivity, indicates that motivation might not explain our results. This would mean that exogenous differences in cognitive ability among respondents might be a significant factor for our results, and thus respondents with lower cognitive ability might have a harder time grasping the scenario and the concept of WTP for risk reductions. Considering that the anomaly of scale bias is not random, but systematically varies with cognitive ability, makes it important to design surveys so this systematic pattern will have as little effect as possible (Sugden 2005). Note that our results do not tell us how people with lower cognitive ability behave in real market situations, only that they might find it harder to grasp this kind of hypothetical scenario.

Standard scale tests in the literature generally focus on the comparison of different subsamples that have stated WTP for different risk reductions (external tests). We primarily focus on scale tests on individual level, i.e. an internal scale test. A critique against internal tests is that they are irrelevant and meaningless since it is hard to imagine that an internal scale test will not pass a validity test of at least weak scale sensitivity (NOAA 1993; Hammer and Johansson-Stenman 2004). This is also supported by findings in Frederick and Fischhoff (1998), one of few studies that have examined internal and external tests, where they show that internal tests lead to more scale sensitivity (Frederick and Fischhoff 1998). However, 36.5% of the respondents in our study showed no scale sensitivity at all and only 10% stated a WTP that was increasing and concave with the size of the risk reduction, i.e. as predicted by economic theory. Our data clearly demonstrate that the assumption that internal tests necessarily pass a scale test is an all too optimistic assumption. The literature review conducted in Hammitt and Graham (1999) also points out the fact that plenty of respondents state the same WTP, even if they get to answer different risk reduction questions after one another. Moreover, Bateman and Brouwer (2006) asked respondents to state their WTP to reduce skin-cancer risk for themselves and their household and tested for weak scale sensitivity. They used a split sample design, and "[i]n summary" (p. 210) found weak scale sensitivity in a subsample that received dichotomous choice questions, but in contrast to our findings, no weak scale sensitivity in a subsample that answered open-ended questions.

External tests are, however, important when VSL (or any other non-marketed amenity) is estimated for policy purposes. A weakness of the standard external tests is that they focus on average values, which do not necessarily reveal anything about the consistency and internal validity of each respondent's WTP. Large positive and negative deviations might result in a near-proportionality of the mean estimate.[20] We could not reject proportionality in the WTP regression, and thus near-proportionality could not be rejected using external tests. Internal tests measure the validity of each individual's stated WTP to a larger degree, and we, therefore, argue that internal and external tests are not substitutes but rather complements, and that internal scale tests can be important as validity tests of respondents' preferences. However, it is important in internal validity tests to be aware of the probability that respondents had the opportunity to give answers they believed were "correct" but did not necessarily reflect their preferences.

Appendix: Cognitive ability test

SURVEY PART 1[21]

To answer this part of the survey you have 5 min at your disposal. The person in charge of the survey will signal when 1 min remains of your time, and when the time is up. When the time is up, please put this part of the survey in your designated envelope.

Question 1. Which number in the following group of numbers represents the smallest amount?
☐ 7 ☐ 0.8 ☐ 31 ☐ 0.33 ☐ 2
Question 2. The ninth month of the year is?
☐ October ☐ January ☐ June ☐ September ☐ May
Question 3. Complete the following series: 100% 0.75 1/2 ?
☐ One ☐ An eighth ☐ A fourth ☐ Zero
Question 4. Look at the row of numbers below. What number should come next?
8 4 2 1 1/2 1/4 ?
Answer:…………
Question 5. Assume that the first two statements are true: "1: The boy plays football." and "2: All football players wear hats." Now evaluate the statement: "The boy wears a hat."
☐ True ☐ False
Question 6. Paper sells for 21 cents per pad. What will 4 pads cost?
Answer:…………
Question 7. How many of the five pairs of items listed below are exact duplicates?

Nieman, K.M.	Neiman, K.M.
Thomas, G.K.	Thomas, C.K.
Hoff, J.P.	Hoff, J.P.
Pino, L.R.	Pina, L.R.
Warner, T.S.	Wanner, T.S.

Answer:…………
Question 8. A train travels 20 feet in 1/5 second. At this same speed, how many feet will it travel in three seconds?
Answer:…………

[20] A result we also got in this study when group comparisons, based on the groups in Table 3, of *Proportionality* were conducted.

[21] The test of cognitive ability was originally in Swedish and has been freely translated from Swedish to English by the authors. To layout has also been slightly changed to save space.

Question 9. When rope is selling at 10 kronor per meter, how many meters can you buy for sixty kronor?

Answer:…………

Question 10. Assume that I spin a coin (one krona) six times. Which of the following combinations are most likely to occur?

Alternative A: Heads Heads Heads Tails Tails Tails

Alternative B: Heads Tails Heads Tails Tails Heads

Most likely: ☐ Alternative A ☐ Alternative B ☐ They are equally likely to occur

Question 11. Three individuals form a partnership and agree to divide the profits equally. X invests 18 000 kronor, Y invests 14 000 kronor, Z invests 8 000 kronor. If the profits are 9 600 kronor, how much less does X receive if the profits are divided in proportion to the amount invested?

Answer:…………

Question 12. Assume the first two statements are true. Is the final one: 1. Tom greeted Lisa 2. Lisa greeted Anna 3. Tom did not greet Anna

☐ True ☐ False ☐ Not certain

Question 13. A boy is 17 years old and his sister is twice as old. When the boy is 23 years old, what will be the age of his sister?

Answer:…………

Question 14. A baseball bat and a ball together cost $11. The baseball bat costs $10 more than the ball. How much is the ball?

Answer:…………

Question 15. You are at a party of truth-tellers and liars. The truth-tellers always tell the truth and the liars always lie. You meet someone new. He tells you that he has just heard a conversation in which a girl said she was a liar. Is the person you met a liar or a truth-teller?

☐ Liar ☐ Truth-teller ☐ Not certain

Question 16. A certain city is served by two hospitals. At the larger hospital 45 babies are on average born each day, and at the smaller hospital 15 babies are on average born each day. As you know, around 50 percent of all newborns are boys, but the actual percentage shifts from day to day. Sometimes the share of boys is higher than 50 percent, and sometimes the share is lower.

During a period of one year the two hospitals registered each day when the share of boys among newborns was equal to or above 60 percent. Which hospital do you think recorded more of such days?

☐ The big hospital ☐ The small hospital

☐ More or less the same (difference only by up to five percent)

Question 17. Linda is 31 years old, single, loud and very intelligent. She has a university diploma from a prestigious university and is interested in discrimination issues and social justice. Based on this, which of the alternatives below do you consider most likely describes Linda?

☐ Linda is an elementary school teacher

☐ Linda works in a bank

☐ Linda is an elementary school teacher and is active in a feminist organization

☐ Linda works for an insurance company

Acknowledgements The authors would like to thank Lars Hultkrantz, Gunnar Isacsson, Gunnar Lindberg, Fridtjof Thomas, our anonymous reviewers, seminar participants at VTI, Karlstad University, Örebro University and Gothenburg University, and conference participants at the "27th Nordic Health Economists' Study Group Meeting" for valuable comments and support. Financial support from the Swedish National Road

Administration is gratefully acknowledged. The authors are solely responsible for the results and views expressed in this paper.

References

Alberini A, Cropper M, Krupnick A, Simon NB (2004) Does the value of a statistical life vary with the age and health status? Evidence from the USA and Canada. J Environ Econ Manage 48(1):769–792
Andersson H, Svensson M (2007) Cognitive ability and scale bias in the contingent valuation method. Working paper 2007:1, VTI, Department of Transport Economics, Stockholm, Sweden
Barthomolew DJ (1987) Latent variable models and factor analysis. Oxford University Press, New York, NY, USA
Bateman IJ, Brouwer R (2006) Consistency and construction in stated WTP for health risk reductions: a novel scope-sensitivity test. Resour Energy Econ 28(3):199–214
Bateman IJ, Carson RT, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Özdemiroḡlu E, Pearce DW, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar, Cheltenham, UK
Benjamin DJ, Shapiro JM (2005) Does cognitive ability reduce psychological bias? Harvard University, Mimeo
Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR (2001) Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. J Health Econ 20(3):441–457
Carlsson F, Johansson-Stenman O, Martinsson P (2004) Is transport safety more valuable in the air? J Risk Uncertainty 28(2):147–163
Carson RT, Mitchell RC (1995) Sequencing and nesting in contingent valuation surveys J Environ Econ Manage 28(2):155–173
Carson RT, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence Environ Resour Econ 19(2):173–210
Clark J, Friesen L (2006) The causes of order effects in contingent valuation surveys: an experimental investigation. University of Canterbury, New Zealand, Mimeo
Corso PS, Hammitt JK, Graham JD (2001) Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation J Risk Uncertainty 23(2):165–184
Desvousges WH, Reed JF, Dunford RW, Boyle KJ, Hudson SP, Wilson KN (1993) In: Contingent valuation: a critical assessment. Hausman JA (ed) Measuring Natural resource damage with contingent valuation: a test of validity and reliability. The Netherlands: Norht-Holland, Amsterdam, pp 91–159
Diamond PA, Hausman JA (1994) Contingent valuation: is some number better than no number? J Econ Perspect 8(4):45–64
Duckworth AL, Seligman ME (2005) Self-discioline outdoes IQ in predicting academic performance of adolescents Psychol Sci 16(12):939–944
Frederick S (2005) Cognitive reflection and decision making J Econ Perspect 19(4):25–42
Frederick S, Fischhoff B (1998) Scope (In)sensitivity in elicited valuations J Risk Decision Policy 3(2):109–123
Green D, Jacowitz KE, Kahneman D, McFadden D (1998) Referendum contingent valuation anchoring and willingness to pay for public goods Resour Energy Econ 20(2):85–116
Hammitt JK (2000) Evaluating contingent valuation of environmental health risks: the proportionality test AERE (Association of Environmental and Resource Economics) Newslett 20(1):14–19
Hammitt JK, Graham JD (1999) Willingness to pay for health protection: inadequate sensitivity to probability J Risk Uncertainty 18(1):33–62
Hammar H, Johansson-Stenman O (2004) The value of risk-free cigarettes – do smokers underestimate the risk? Health Econ 13(1):59–71
Hammitt JK, Zhou Y (2006) The economic value of air-pollution-related health risks in China: a contingent valuation study Environ Resour Econ 33(3):399–423
Hanley N, Shogren J (2005) Is cost-benefit analysis anomaly proof? Environ Resour Econ 32(1):13–34
Harrison GW (2006) Experimental evidence on alternative environmental valuation methods Environ Resour Econ 34(1):125–186
Heberlein TA, Wilson MA, Bishop RC, Schaeffer NC (2005) Rethinking the scope test as a criterion for validity in contingent valuation J Environ Econ Manage 50(1):1–22
Johnson R (2006) Comment on "Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment" Environ Resour Econ 34(1):45–50

Jones-Lee MW (1974) The value of changes in the probability of death or injury J Political Econ 82(4):835–849

Kahneman D (2003) Maps of bounded rationality: psychology for behavioral economics Am Econ Rev 93(5):1449–1475

Kahneman D, Tversky A (1972) Subjective probability: a judgement of representativeness Cogn Psychol 3(3):430–454

Kahneman D, Tversky A (1983) Can irrationality be intelligently discussed? Behav Brain Sci 6:509–510

Kahneman D, Ritov I, Schkade D (1999) Economic preferences or attitude expressions? An analysis of dollar responses to public issues J Risk Uncertainty 19(1–3):203–235

Kahneman D, Slovic P, Tversky A (1982) Judgement under uncertainty: heuristics and biases. Cambridge University Press, New York, NY, USA

Lawley DN, Maxwell AE (1963) Factor analysis as a statistical method. Butterworths, London, UK

NOAA (1993) Report of the NOAA panel on contingent valuation. Federal Register 58, Arrow, K, Solow, R, Portney, P, Leamer, E, Radner, R, and Schuman, H

Persson U, Norinder A, Hjalte K, Gralén K (2001) The value of a statistical life in transport: findings from a new contingent valuation study in Sweden J Risk Uncertainty 23(2):121–134

Rabin M (2002) Inference by believers in the law of small numbers Quart J Econ 117(3):775–816

SIKA (2007) Vägtrafikskador 2002–2005 (Road traffic injuries 2002–2005). Internet, www.sika-institute.se, 4/26/07

Smith VK (1992) Arbitrary values, good causes, and premature verdicts J Environ Econ Manage 22(1):71–89

Sugden R (2005) Anomalies and stated preference techniques: a framework for a discussion of coping strategies Environ Resour Econ 32(1):1–12

Weinstein MC, Shepard DS, Pliskin JS (1980) The economic value of changing mortality probabilities: a decision-theoretic approach Quart J Econ 94(2):373–396

Wolfe R, Johnson S (1995) Personality as a predictor of college performance Educ Psychol Measure 55:177–185