Toward Improved Economic Analysis Using Contingent Valuation: Some Methodological Considerations Applied to River Toxics and Dam Removal

Fred Hitzhusen
Professor
hitzhusen.1@osu.edu

Ashraf Abdul-Mohsen
Graduate Research Associate
abdul-mohsen.1@osu.edu

Sarah Kruse
Graduate Research Associate
kruse.22@osu.edu

Dept of Agricultural, Environmental and Development Economics
Ohio State University
2120 Fyffe Rd, Columbus OH, 43210

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, July 1-4, 2004

Abstract

This paper addresses conceptually and empirically some of the biases commonly associated with contingent valuation (CV) elicitation of willingness to pay (WTP) values for non-market goods and services. More specifically, the study focuses on testing for scope, context, and sequence effects in CV mail surveys as well as the assumption of well-defined preferences in mail questionnaires and how this assumption might bias WTP estimates. Our results suggest that the absence of scope effects in the some CV mail surveys might be a result of the complexity or multidimensional aspect of the policy in question (i.e. dredging with and without dam removal) and the assumption that increases in scale of the public good are easier to be comprehended and then translated into dollar values than increases in scope by the average respondent. Moreover, results of the initial mail survey suggest that individuals may not have well defined preferences for goods (i.e. dam removal) with which they are not familiar or experienced. Pre-testing of the structured elicitation groups suggests that this alternative elicitation format based on a philosophy of constructive preferences may lead to more thoughtful and rational WTP values.

Introduction

There is currently much interest in the restoration of polluted rivers and streams in the united states for reasons such as revitalizing the economic conditions in some areas where the rivers/streams are located as well as improving the well-being of their residents. The presence of toxic sediments and obsolete dams is contributing to much of the degradation in ecosystems and the loss of fish habitats in many of these water bodies. Consequently, toxics and/or dam removal are considered by some governmental agencies and environmental groups as vital options for river restoration. But, in order to justify the cost of any restoration project, measuring the benefits of environmental quality improvements is essential for policy purposes. Contingent valuation (CV) is a stated preference method that uses survey techniques to estimate how much people value hypothetical changes in the resource of interest (Smith 1992).There is an ongoing debate regarding the use of contingent valuation to measure the total economic value, including non-use value, for damaged natural resources including rivers and streams.

The validity of stated willingness to pay (WTP) as a measure of value has been examined extensively in the literature [for example, Mitchell and Carson 1989; Smith 1992; Kahneman and Knetsch 1992; Hoehn and Randall 1989]. The validity of a measure, as defined by Mitchell and Carson 1989 (1989), is "*the degree to which it measures the theoretical construct under investigation*". Two types of validity are of more concern to the study at hand ─ theoretical (construct) validity and content validity.

Theoretical validity is concerned with the degree to which the outcomes of a CVM study ─for example, the sign and magnitude of the estimated coefficients in the bid functions and the relative magnitude of willingness to pay under different split-sample conditions ─ are consistent with expectations of economic theory. One way to examine the theoretical validity of a CVM

study is to compare WTP estimates of different scenarios for which theory suggests statistically different or similar values.

For example, economic theory would predict that the greater the amount of the public good being offered, the more an individual would be willing to pay for that good. This has been known in the literature as scope or scale effect or the absence of part-whole bias [whitehead et al. 1998] in CV surveys. Theory would also suggest that the value of an environmental good is independent on the serial position of the good in a sequence of other environmental goods (sequence or order effects) and independent of the context in which the good is presented or what is known as context effects (Carson et al. 2002). The last two effects (order and context) constitute what has been known in the literature as embedding effect: the value of a particular good depends upon whether it is valued alone or as part of a more inclusive agenda or package (Loomis et al. 1993).

Content validity is concerned with how a CVM questionnaire presents the market structure and defines the amenity in a way that accords with well defined preferences. Psychologists define content validity as the extent to which an empirical measurement adequately reflects a specific domain of content [Carmines and Zeller, 1979]. The relevant domain in contingent valuation (CV) studies is the structure of the market and the description of the good being valued. Our examination of content validity focuses on the former, and in particular, how individual preferences translate into the market structure.

A common assumption among many economists is that "each individual has stable and coherent preferences" (Rabin, 1998). It is also accepted that "people know their preferences" (Freeman, 1993), that they have the ability to maximize those preferences, and they will choose

the option that does so (Payne et al., 1999). Under these assumptions, the use of mail surveys to a large sample population may indeed provide accurate estimates of willingness-to-pay (WTP).

But according to Payne et al. (1999), a growing number of researchers believe that the assumptions of well-defined preferences work only when individuals are familiar and experienced with the good being valued. Even if individuals are familiar with the good, it is less likely that they have experience in valuing it, at least in a monetary sense. Under these assumptions, it is perhaps more appropriate to examine a constructive view of preferences.

This paper is divided into two parts. The first part addresses the issues of scope and embedding effects in a CV mail survey. The second part addresses the issue of constructive preferences and how structured small group elicitations could provide a more plausible alternative for eliciting accurate WTP measures than random mail surveys.

I- Scope and Embedding Effects in Contingent Valuation Mail Surveys

We start off by examining the issue of scope or scale effects and then we turn to the issue of embedding. Carson and Mitchell (1995) developed a 'component sensitivity test', which rejected the hypothesis that respondents are insensitive to the scope of the good being valued. That is, respondents in this study were able to perceive different levels of provision of the environmental good and took this difference into account when asked to state a value for that good. Also, whitehead et al. (1998) found that WTP estimates, including non-use values, were sensitive to the scope of the policy (improving water quality in two recreational sounds in North Carolina), and that the use of inexpensive survey methods such as telephone and mail surveys may not be the cause of the presence of part-whole bias in some recent CVM studies. On the other hand, some critics of the contingent valuation technique (Kahneman and Knetsch 1992)

found that WTP did not change significantly between a less inclusive good and more inclusive goods and that the measures of value obtained using CVM might be consequently arbitrary. Loomis et al. (1993) found mixed evidence of scope effects in their valuation of forest protection in Australia.

While Smith (1992) and Carson and Mitchell (1993) have argued that the absence of scope effects in some studies might be the result of survey design and administration problems such as poor description and framing of the commodity to be valued and bad implementation of the CVM survey, few studies[♣] (to our knowledge) have looked exclusively at the distinction between increases of the same good and adding more goods to the good being valued and how this is related to sensitivity to scope in CV surveys. In other words, the literature on contingent valuation does not seem to distinguish between scale or size effects and scope effects. In addition the studies that value more (quantitative or geographic) of the same good (for example, Carson and Mitchell 1995; Whitehead et al. 1998) are more likely to find size or scale effects than studies that value additions of other goods (scope) to the good in question (for example, Kahneman and Knetsch 1992). In this study we test for scope effects using a mail survey that elicits both use and non-use values.

Defining the Good and Designing the Survey

The study case is restoring the Lower Mahoning River in northeast Ohio to its pre-industrialization conditions. This segment of the River is filled with contaminated sediments that have polluted the river and degraded its ecosystem for almost a century. Contaminated sediments are from waste disposal of the early steel mills that once developed along the river banks, and from the disposal of adjacent communities into the river. In addition, there are 10 low-head dams

[♣] For example, Carson and Mitchell's (1995) classification of scope into quantitative nesting and categorical nesting

that impede the continuous flow of the river in this stretch and as a result prohibit the spawning of some fish species. Two programs have been proposed by the Army Corps of Engineers to restore the biotic and aquatic integrity of the river to pre-contamination conditions. These programs are: (1) dredging of the contaminated sediments (D) without any dam removal and (2) dredging and removal of some of the low-head dams (D-DR), with the second program providing more and longer-lasting benefits than the first.

Dredging only would result in partial restoration of the river since sediments will eventually build behind the dams and reduce the benefits of restoration over time. Dredging with dam removal would allow the above benefits to be observed for a longer period of time. Additionally, the partial removal of the dams would allow fish to migrate freely in the river and thus make the improvement in fish habitat more sustainable. Having said that, the first program (D) is nested within the second program (D-DR) and that allows testing for scope effects.

Survey Version	**Program/Agenda**
1	(D)
2	(D-DR)
3	(D) → (D-DR)
4	(D-DR) → (D)

Figure 1: Mahoning River Survey Versions

The data for this study are from a 2003 mail survey of a randomly selected sample of respondents in northeastern Ohio. To test for scope and other effects four versions of the CVM survey were sent to seven counties within the Mahoning River Watershed (MRW). The four versions are listed above. Version one contained a contingent market for the dredging only (D) program. Version two contained a contingent market for the (D-DR) program. The difference

between these versions is that in version two we added "with partial low-head removal" after "dredging" and then explained the additional benefits of dam removal. Version three contained a contingent market for an agenda in which (D) is valued first then (D-DR) is valued second. And version four contained a contingent market for an agenda in which (D-DR) is offered first then (D) is offered second. Versions three and four are identical in every way except for the order of the valuation questions.

The demographic profile of the sample is similar to that of Northeast Ohio except for median age. The median age of the sample (56 years) is higher than that of the population (36 years) but this is typical of mail surveys where older people tend to respond more than the young. Using simple T tests for the difference in means, none of the demographic or behavioral characteristics is significantly different at the 0.05 level among the four versions of the survey.

Table 1. Data Summary

Variable	version 1 N=113	version 2 N=84	version 3 N=85	version 4 N=78
PROTECT	4.12	4.1	4.09	3.97
OWNBOAT	0.19	0.23	0.18	0.08
LAKES	4.62	4.54	5.22	3.17
YEARSIN	22	24	20	23
AGE (years)	58	55	57	55
PERSONS (avg.)	2.25	2.29	2.55	2.64
MALE	0.65	0.59	0.72	0.61
WHITE	0.97	0.95	0.88	0.96
EDU	3.14	2.51	2.98	2.99
INCOME (2000 $)	51649	45365	47833	42279

The first part of the survey was about eliciting information regarding respondents' attitudes (on a scale of 1 to 5) toward environmental awareness and importance (PROTECT), participation in environmental entities, knowledge about river contamination and participation in recreational activities on the river and on lakes in the area (LAKES). The second part of the

survey contained the policy scenario which compared the status quo of the river with the alternative situation in which a restoration program/agenda is proposed. Then, the contingent market was established and a dichotomous choice WTP question followed by an open-ended question was asked to elicit a respondent's maximum WTP for the proposed program.

Each respondent was asked to vote on a one-time payment to a multi-county special district fund that could only be used for restoration efforts of the river. We avoided using increases in taxes or utility prices as payment vehicles to reduce protest responses while keeping the payment mechanism as mandatory as possible; the special district fund is not voluntary and usually used by counties in Ohio to collect money for local issues. The price amounts were randomly selected from four WTP values: 50, 100, 200, and 400[1]. The last part of the survey contained questions about various demographic characteristics of the respondents to be used in the bid functions to explain variations in WTP amounts and to generalize WTP sample estimates to the population.

Scope Results

Table 2 shows frequencies of the no responses for dredging only (D) and dredging with dam removal (D-DR). These proportions are then used to calculate the Turnbull distribution-free lower bound estimate of mean WTP (Haab and McConnell 2002). The underlying notion of the distribution free estimator is that when a respondent i answered no to the offered bid t_j we have learned that his willingness to pay is less than t_j. Then, the probability that a randomly chosen respondent having WTP less than t_j is

$$\Pr(WTP_i < tj) = F_W(tj) = F_j$$

[1] Choice of the bids was based in part on the pre-test survey results and bid amounts differ with the scope of the good: 50, 100, and 200 for (D), and 100, 200, and 400 for (D-DR).

Where F_j is the cumulative distribution function of WTP and should be monotonically increasing in the bid amount. This is because we would expect a higher proportion of respondents to answer no at a higher price. The Turnbull estimator guarantees this monotonicity by pooling responses to prices that fail to meet this condition. The lower bound estimate of WTP is obtained by multiplying each offered price by the probability that WTP falls between that price and the next highest price:

$$WTP_{LB} = \sum_{j=1}^{M+1} c_{j-1} f_j$$

Where $f_j = F_j - F_{j-1}$, M+1 is the upper bound on the range of WTP, and c is the price.

Table 2. Frequency of the "NO" response

(D)				(D-DR)		
Bid		No	Total	Bid	No	Total
50		17	42	100	12	23
100		24	41	200	25	31
200		18	30	400	22	30
WTP_{LB}[a]	\$90.49 (10.45)[b]			\$70.78 (12.87)		
Δ WTP			\$19.71			
T stat			1.19[c]			

[a] Turnbull lower-bound mean WTP
[b] Standard error of WTP
[c] Not significant at the 0.05 level

From the split-sample data, the difference in minimum mean WTP between the D and D-DR samples is \$19.71, which is insignificantly different from zero at the 0.05 level using a one-tailed test (t = 1.19). Also, the lower bound on the range of median WTP, the price for which a probability of no response equals 0.5, is \$50-\$100 for D and \$0-\$100 for D-DR assuming a non-negative WTP. These estimates for the lower-bound mean and range of median WTP show

that no scope effects are present in the split-sample data; i.e., respondents offered one program or the other are not sensitive to scope of the good being valued.

Regression analysis is used to relate WTP to demographic and behavioral characteristics of the respondents. First, we test the sensitivity to scope using a dummy variable for the more inclusive good (D-DR) in the probit equation while holding constant other potentially influential variables. Second we test the equality of coefficients in the bid functions for the two scopes of the good using a likelihood ratio test. Our regression model assumes an exponential WTP function, which bounds WTP from below to be non-negative and does not bound it from above. The functional form is:

$$WTP_j = e^{\beta X_j + \varepsilon_j}$$

Where X is the vector of independent variables, β is the corresponding vector of parameters, and ε is a normally distributed random error term with mean zero and constant variance σ^2. Mean and median WTP from the probit model are calculated using the following equations:

$$E_{\varepsilon}(WTP \setminus X_j, \beta) = e^{\beta X_j + 0.5\sigma^2}$$

$$MD_{\varepsilon}(WTP \setminus X_j, \beta) = e^{\beta X_j}$$

Since expected WTP for the exponential model is increasing in σ^2, the difference between mean and median WTP will be bigger for higher values of σ^2 (Haab and McConnell 2002). For the current data, $\sigma^2 = 1.03$ so the difference between the mean and the median is relatively large (of a degree of magnitude). As such, the more conservative estimate of WTP (the median) is used to test for between and within sample scope effects.

Table 3 shows results of the probit regression on data from versions one and two only. The coefficient on the (D-DR) dummy variable is insignificant at the .05 level, meaning that

Table 3. Probit Regression Results[a]		
Variable	Coefficient	t-Value
Constant	-.96	-.94
Protect	.55***	3.93
Lakes	.03**	2.07
Income/1000	.01***	3.63
D-DR	-.2	-.75
Log (Bid)	-.47**	-2.36
Δ WTP	$10.74	
Log Likelihood Function	-91.82	

[a] sample size = 167, $\sigma^2 = 1.03$
** Significant at the .05 level
*** Significant at the .01 level

respondents are not willing to pay more for the dredging with dam removal project relative to the dredging only project. The importance of protecting the environment and income coefficients are significantly different from zero at the .01 level while the lakes and bid price coefficients are significant at the .05 level. All coefficients except for the D-DR coefficient have the expected signs. That is, WTP is increasing in household median income, whether the respondents sees that protecting the environment is an important national goal, and whether the respondent recreates on lakes in the Mahoning River Valley. Consistent with economic theory, WTP is decreasing in the logarithm of the bid price. On the other hand, WTP is decreasing in the scope of the good but the relationship is highly insignificant as indicated by the difference in median willingness to pay (Δ WTP) between the two scopes. Δ WTP is $10.74 and is insignificantly different from zero at the .05 level based on a comparison of the 95% confidence intervals for the two goods─ confidence intervals are [$37-$234] for the dredging only policy and [$2-$158] for the dredging plus dam removal policy while median WTP is $93.23 and $82.49 for D and D-DR, respectively.

Table (4) shows probit equations estimated for each scope of the good. Regression results show that all variables maintain their respective signs across the two versions but the magnitude and significance of the coefficients does change from version to version. The bid coefficient becomes more influential and significant as the scope of the good gets smaller. Then, using the likelihood ratio test we accept the assumption of equal coefficients across levels of aggregation at the .05 level (χ^2 = 0.89). In other words, there was no statistically different WTP behavior being exhibited in responses to questions about dredging only and dredging with dam removal. The end result is that there were no across-sample scope effects. This result is consistent with previous findings of CVM studies where respondents are asked to value programs or policies that are not necessarily providing more of the same good in terms of quantity, geographic scale, or temporal scale. In these cases, respondents may not be able to perceive the scope of the good in the same way they would perceive consuming more apples or bananas, and that it is the difference between measuring scope and scale effects.

Table 4. Likelihood Ratio Test

Variable	D	D-DR
Constant	-0.7	-1.89
	(-0.56)[a]	(-0.94)
Protect	0.53***	0.61***
	(2.99)	(2.57)
Lakes	0.02	0.03*
	(1.20)	(1.77)
Income/1000	0.01***	0.02**
	(2.8)	(2.33)
Log (Bid)	-0.51*	-0.39
	(-1.92)	(-1.26)
Log Likelihood	-55.51	-36.15
LR (d.f.)[b]	0.89 (5)	

* Significant at the .10 level
** Significant at the .05 level
*** Significant at the .01 level

[a] t-Value
b χ^2 (tabulated) = 11.07

In the Mahoning River case, respondents were clearly informed that the more inclusive project (D-DR) would provide the same benefits as the other project (D) for a longer period of time in addition to providing other benefits more specific to dam removal such as improving fish habitat and allowing for continuous navigation in the river. However, these amenities cannot be considered as involving more quantity of the same good. Moreover, we did not explore the possibility that some residents might be opposing dam removal because they like the dams or benefit from their presence; for example, a steel mill that still use water behind some of the dams for cooling purposes. We could not examine these possible negative effects of dam removal because our survey targeted the general public not specifically stakeholders in the steel industry (owners, workers, and other related industries in the region). However the contingent valuation method should in principal be able to account for these effects because respondents are expected to incorporate all components of value including negative values when responding to a WTP question.

The issue of Embedding

The problem of embedding has been defined and tested for differently, and sometimes confused with the problem of scope in contingent valuation studies. Kahmeman and Knetsch (1992) define perfect embedding as the equality of directly elicited WTP values regardless of the degree of scale the good is defined over. Carson and Mitchell (1995) argue that CV responses are not context free (due to substitution effects and budget constraints), and that the size and nature of the choice set are important determinants of how an individual values a particular good. This is consistent with the notion by Hoehn and Randall (1989) that "…*as the number of policy proposals becomes large, conventional benefit cost procedures* [in which proposals are evaluated independently of each other] *are certain to overstate a valid measure of net benefits*."

A more general definition of the embedding effect is that the value of a particular good depends upon whether it is valued alone or as part of a larger package or agenda (Loomis et al. 1993). This notion is more general because it does not prescribe scale or nesting as part of the embedding problem. Our examination is based on this definition because it assumes that testing for embedding can be accomplished only by comparing a scenario in which the resource is valued on its own to other scenarios in which the resource is valued within an agenda regardless of whether the agenda offers more of the same resource or other resources for the respondent to value. We prefer to call this 'context effect' because it looks at how CV responses are affected by the context in which the valuation question is asked and also because it is less confusing than the term 'embedding'.

The presence of context effects in CV surveys has been attributed to substitution or complementarity effects and budget constraints (Hoehn 1991; Hoehn and Randall 1989): as the number of policy proposals increases, there is more likelihood of substitution or complementarity effects between programs within the larger agenda. In addition, as the dollar value of total WTP approaches a maximum limit in the respondent's expenditure function, budget constraint becomes a restraining factor on WTP for each program.

Another problem that CV elicited values may suffer from is an order or sequence effect: the value of a particular resource depends on the order in which it is valued within an agenda. Payne et al (2000) found a strong sequence effect when respondents were asked to value five environmental goods using WTP and other evaluative attitude ratings. They also found that the total WTP for the bundle of the five environmental goods depended on the value of the good evaluated first. This result contradicts with the theoretical prediction by Carson and Mitchell (1995) that the sum of WTP for a bundle of goods should be invariant to the serial positions of

these goods in the sequence. In the current study, we explicitly test for context and sequence effects using the Mahoning River survey data. First, we test for context effects by comparing scenarios one and two, in which D and D-DR are each valued alone, to scenarios three and four, in which D and D-DR are valued first in an agenda, respectively. Next, we test for sequencing by comparing scenarios three and four, in which both goods are valued within an agenda but in a different sequence.

Table 5. Context Effects

Variable	Coefficient	t-Value	Coefficient	t-Value
	D		**D-DR**	
Constant	2.402***	2.61	1.696	1.41
Parks	0.077***	2.53	0.093**	2.70
Own Boat	0.102	0.41	0.655***	2.12
Age	-0.015***	-2.39	-0.005	-0.74
Ver 3	***-0.013***	***-0.07***		
Ver 4			***0.380***	***1.67***
Log (Bid)	-0.387**	-2.23	-0.441**	-2.05
Log Likelihood	-121.82		-88.22	
Sample Size	190		152	

** Significant at the .05 level
*** Significant at the .01 level

Table 5 shows multivariate analyses of the context effects, in which a dummy variable for the multiple-program scenario is included as a covariate. In the second column, the coefficient on the scenario three dummy variable is insignificant at the 0.05 level, indicating that there are no context effects in the valuation of the dredging only project. Respondents are willing to pay $95.13 for D when it is valued on its own, and $91.87 when it is valued first in an agenda offering D-DR after D. similarly, the coefficient on scenario four in the fourth column of Table 5 is insignificant at the 0.05 level, indicating no context effects in the valuation of the dredging with dam removal project. Although WTP for D-DR is $72.14 when the good is valued alone

(scenario 2) and $129.75 when it is valued first in an agenda (scenario 4), the difference in WTP is insignificant at the .05 level.

Table 6. Order Effects

Variable	Coefficient	t-Value	Coefficient	t-Value
	D		**D-DR**	
Constant	1.570	1.53	2.343**	1.99
Parks	0.069**	2.131	0.080**	2.39
YEARS IN	-0.007	-1.167	-0.006	-0.87
EDU	0.183**	1.983	0.279***	2.90
VER 4	***-0.468****	***-2.111***	***-0.096***	***-0.43***
Log (Bid)	-0.467**	-2.272	-0.633***	-2.97
Log Likelihood	-90.34		-87.33	
Sample Size	150		148	

** Significant at the .05 level
*** Significant at the .01 level

Sequence results are shown in Table 6 for both D and D-DR. The coefficient on scenario four in the second column has a negative sign and is significant at the 0.05 level, indicating that D is valued less when it is offered second than when offered first within an agenda. This is reflected in WTP values for dredging only, $104.73 in scenario three and $40.79 in scenario four. The presence of an order effect in valuing D is consistent with previous findings in the literature (Payne et al. 2000), which indicates that a particular resource is valued more when it is presented first than later in a sequence. The coefficient on scenario four in the fourth column of Table 6 is negative however insignificant, showing no order effect in the case of the more inclusive good (D-DR) as reflected in WTP amounts for that good─ $183.16 in scenario three and $157.46 in scenario four─ with the difference in WTP being insignificant at the .05 level. In fact, the negative sign on scenario four in the later case indicates that D-DR might be valued more when it is offered after D as in scenario three. This could be attributed to a possible

composite scope-sequence effect in valuing multiple-program agendas containing nested public goods: a more inclusive good is more likely to be valued more when presented after a less inclusive good (scenario 3) than when presented before (scenario 4). The opposite is also correct and conforms to the negative sequence result in the valuation of D. That is, D (the smaller good) is valued less when it is presented after D-DR (the more inclusive good) than when presented before D-DR in a sequence.

II- Constructive Preferences

The two major tenets of constructive preferences are (1) expressions of preference are generally constructed at the time the valuation question is asked and (2) the construction process is shaped by the properties of the decision task and the ability of the respondent to process the information [Payne et al., 1992; Slovic, 1995]. In other words, values given by respondents are not based on well-defined preferences, but rather, on information stored in memory and information gained at the time the valuation question is asked.

Due to the relatively complex non-market, and often, non-use values of many of the goods for which contingent valuation attempts to assess WTP values, we believe that the constructive preference philosophy is perhaps a more plausible argument. As Fischhoff (1997) points out, even if respondents have thought about an environmental good, the proposed changes are new and individuals cannot instantly convert these new details into valuations. For some, it may take time to conceptualize environmental resources as "goods" that have a monetary value (Fischhoff, 1997). Therefore, the ability of the researcher to provide more a constructive approach to elicitation would be imperative to obtaining more accurate estimates of respondent's WTP values. Under this assumption, the use of a random mail survey may fail to provide accurate estimates of individual's WTP.

We have attempted to develop a WTP elicitation format that accounts for, and helps respondents construct more rational and thoughtful preferences; structured small group elicitation. We suggest that if an individual's preferences are well-defined, then a slight variation in the structure, but not content, of the elicitation format should not significantly alter their stated WTP. On the other hand, if individuals are constructing preferences at the time the question is being asked, there may be a significant difference between the WTP values obtained from the two response formats, and/or their perceptions of the process and their stated WTP value. In order to test the hypothesis two treatments using the same survey will be conducted; a random mail survey and a structured elicitation group (SEG) format with individually administered surveys at the end of each session.

Because we cannot validate the WTP values obtained from either elicitation format, it is difficult to determine which values are "better". Our argument is that if the elicitation process successfully encourages individuals to give more thoughtful and rational preferences, then their WTP values will reflect that. In other words, the process should justify the results. In order to measure the quality, or content validity, of both elicitation formats, we included within-group and across-group treatments.

Defining the Good and Designing the Survey

The study case for the survey sampling and treatments will be the potential removal of the Ballville Dam, located on the Sandusky River in Northwestern Ohio. The sample population chosen for this part of the study was a randomly selected subset of (1) individuals residing in Sandusky County and (2) individuals living within a 30-mile radius of the dam, but not residing in Sandusky County. The main considerations with respect to the removal of the dam are (1) the water supply for the City of Fremont, (2) safety and (3) the potential for restoring river quality

and increasing fish spawning habitats. Overall, Sandusky County was chosen based on the belief that these individuals will be more familiar and knowledgeable on these issues. The second sample was smaller than the one drawn for Sandusky County based of the assumption that individuals living farther from the dam would be less familiar with the good being valued, and also less likely to be affected positively, or negatively, by the removal of the dam.

A total of 974 surveys were mailed on March 1, 2004 to a randomly selected population of 724 Sandusky residents and 250 individuals that lived within a 30-mile radius of the Ballville Dam. Two different types of surveys were distributed randomly to the sample population. They were identical in all respects except for the willingness-to-pay question; one subset used a dichotomous choice (DC) format and the rest used an open-ended (OE) format. Five different bid values were used for the DC format and these bids were randomly distributed. Procedures attempted to follow as closely as possible those developed by Dillman (1978) in his total design method. Each survey included a personalized cover letter and return envelope that included postage. Approximately three weeks after the first mailing, March 23, 2004, a second mailing of the survey was sent out with a second cover letter.

The total response rate across all three surveys was 30%, with higher response rates within Sandusky County (35% for the DC survey). This result is consistent with a study by Heberlein and Baumgartner (1978) that found higher response rates when "respondents perceive the CV's purpose be *directly connected* with their interests" (Bishop, Heberlein, and Kealy, 1983).

Demographics

The demographic profile of survey respondents is similar to that of the general population of Sandusky County and of the state of Ohio. As is common in many CV surveys the percent of male respondents (70.4%) was higher than the female response rate (29.6%). In the case of this particular study, the survey was mailed to the individual listed as the head of household, which is commonly a male.

Eighteen percent of respondents have lived at their current residence for less than five years, while 50% have lived in the same place for 15 years or more. This high percentage of long-time residents suggests that perceptions of environmental and recreational quality are based on long-term interaction and observance of the Sandusky River and the Ballville Dam. Similarly, 52% of respondents live 5 miles or less from the river, suggesting that they would have more information than the average individual on river quality, etc.

The first part of the survey asked questions regarding respondents' awareness of the Ballville Dam (AWARE), knowledge of dam removal, perceptions of the Sandusky River in terms of environmental and recreational quality, participation levels in recreational activities on the river, and their desire to see the recreational and environmental quality of the river improve (CONCREC and CONCEV). The second part of the survey included a description of the potential advantages and disadvantages of removing the Ballville Dam. Next, the contingent market was created with each respondent being asked to vote on (for the dichotomous choice question) or state their maximum WTP (for the open-ended question).

In order to avoid increasing taxes, the payment vehicle was a one-time payment to a trust fund set up specifically for the project. Five bid values were chosen based on pre-testing of the contingent market; 10, 20, 50, 75, 100. Finally, respondents were asked general demographic

questions, and completed a brief survey evaluation. Those results will later be compared to the responses of individuals participating in the structured elicitation groups.

Table 7. Data Summary

Variable	Sandusky DC	Sandusky OE	30-Mile DC
	N=144	N=72	N=40
AWARE (percent)	0.92	0.94	0.33
EORGS	0.11	0.13	0.10
MAUMEE	0.06	0.14	0.18
AGE (years)	53	52	55
HOUSEHOLD (avg.)	2.58	2.61	2.33
MALE	0.70	0.74	0.67
WHITE	0.97	0.97	0.92
EDU	4.20	4.19	4.23
INCOME (2003 $)	51,880	53,280	54,580

Estimation Results

A linear random utility model was used to analyze the responses for the dichotomous choice survey, in which WTP is defined as

$$WTP_j = \alpha z_j / \beta + \varepsilon_j / \beta .$$

Table 8 shows the results of the probit analysis on the data from Sandusky County only. Another probit regression was run which included both Sandusky and 30-mile DC respondents and used a dummy variable (SANDUSKY), but this variable was not significant. The importance of improving recreational quality on the Sandusky River was significantly different from zero at the 0.01 level, while the bid value, participation in environmental organizations, and knowledge of the Ballville Dam are significant at the 0.05 level. Also, a dummy variable for respondents who stated that they might have answered survey questions differently with more information

(CHANGE) is significant at the 0.01 level. All coefficients had the expected sign, with WTP decreasing in bid value. Mean WTP for the DC survey was $42.93 if non-responses were considered a "no", and $50.91 if non-responses were omitted from the analysis.

Table 8. Probit Regression Results for Sandusky DC[a]

Variable	Coefficient	t-Value
Constant	-0.87	-1.36
Aware	-1.10**	-2.11
Concrec	0.65***	4.42
Change	-0.77***	-2.58
Eorgs	0.90**	2.13
Useh2o	0.48*	1.73
Bid	-0.01**	-2.34
MWTP		$42.93
Log Likelihood Function	-70.08	

[a] Sample size = 134
* Significant at the .10 level
** Significant at the .05 level
*** Significant at the .01 level

Results of the linear regression for the open-ended responses are shown in Table 9 below. Concern for improving the recreational quality of the Sandusky was significant at the 0.01 level, while education, recreating on the Maumee and boating were all significant at the 0.05 level. All variables had the expected sign, except for CONCREC which was positive and significant in the DC model, but is negative and significant in this model. Mean WTP was $47.59.

Nineteen percent of OE respondents did not answer the WTP question, while only 4% of DC respondents left the WTP question blank. Our hypothesis is that individuals have a more difficult time coming up with a WTP value versus checking yes/no to a given bid. This is one of the issues that we believe will be addressed by the SEG.

Another finding was that 28% of all respondents to the mail survey stated that they might have answered survey questions differently if they had been provided with more detailed information on the issue, and approximately 20% of these individuals had given a zero WTP

value (OE) or said no to their assigned bid value (DC). It is our hypothesis that the use of structured elicitation groups will reduce this degree of uncertainty with respect to WTP values.

Table 9. Linear Regression Results for Sandusky OE[a]

Variable	Coefficient	t-Value
Constant	86.85	0.75
Concev	42.13*	1.75
Concrec	-78.81***	-2.74
Edu	27.40**	2.41
Useh2o	80.69*	2.16
Maumee	150.46**	2.49
Mile	-29.43	-1.60
Boat	14.71**	2.16
MWTP		$47.59

[a] Sample size = 57
* Significant at the .10 level
** Significant at the .05 level
*** Significant at the .01 level

Structured elicitation groups

One requirement of a good contingent valuation survey is that individuals' valuation responses accurately represent their true willingness to pay for the good or service the researcher is attempting to value. If the problem statement is not clearly defined, it is possible that the individual respondent may provide a WTP value for something different than what the researcher is asking. This may be due to confusing presentation, lack of knowledge, or other factors (Fischhoff et al., 1999). In any case, it is extremely difficult for the researcher to know exactly what a particular respondent understood and exactly what question they are answering when giving their WTP value. The use of small structured elicitation groups may help minimize the risk of scenario misspecification by allowing for question clarification.

It has been found that mail survey respondents are often unwilling to make tradeoffs. This avoidance of tradeoffs is reflected in behaviors such as selection of the status-quo, the use

of a "protest" zero vote, or delaying choice (Luce, 1998). The potential for avoiding tradeoffs could be minimized by using structured small groups to provide time and tools to analyze the objectives, scenarios, and attributes of the problem. In the case of environmental goods, it is likely that even if an individual has a strong value for the resource they may have difficulty expressing that value in monetary terms. If this translation is biased, then the response value will not provide an accurate description of the individual's true value for the good.

The method proposed uses multiple elicitation sessions, with each session using a randomly selected sample of 7-10 people from the affected population. Participants are recruited through local churches and sessions will be held at the participating churches. This recruitment format was suggested by Joe Arvai, a natural resource decision policymaker, who has found this style to be effective for several reasons including; the head of the church recruits the members which in effect gives his/her approval for the sessions, the individuals meet in the church ---a familiar setting, and participate in sessions with other members of their church.

We are currently in the process of speaking to several ministerial commissions in an effort to select churches that are most representative of the general population. In order to make sure that the SEG are representative of the general population, we intend to monitor the demographic status of each session and if necessary, recruit selectively for the comparative purposes of this study.

The sessions include the use of individual workbooks, group questions, and a facilitator, which are meant to reduce the cognitive demands for individuals participating in the sessions by making the required tasks more manageable. The handouts used in the group sessions include two tasks where individuals work alone to rank alternatives as they relate to the effects of dam removal. These tasks are meant to facilitate the decision process by encouraging individuals

understand the alternatives and consequences of their decision. In order to ensure no overlap between the SEG and the mail survey, individuals will be asked if they filled out a mail survey on the proposed dam removal and will not be included in the sessions if they have.

Conclusions

We find that CV-based estimates of WTP are not sensitive to the scope of the policy being proposed for Mahoning River restoration─ Respondents are not willing to pay more for the dredging with dam removal policy than for the dredging only policy. These results suggest that the absence of scope effects in some CV studies might be a result of the complexity or multidimensional aspect of the good in question and the assumption that scale effects are easier to be comprehended and then translated into dollar values than scope effects by the average respondent. Further research should be directed to testing for scope and scale effects within the same survey to control for the possible effects of respondent's characteristics or constraints on WTP estimates.

We find also no context effects in WTP values across survey scenarios, indicating that WTP responses are not sensitive to whether the policy is valued on its own or within an agenda. Based on the general definition of embedding, this finding contradicts with what previous studies (Kaneman and Knetch 1992; Loomis et al. 1993) have found: CV studies that fail the cross-sample scope test ought to be suffering from an embedding problem. Our results does not suggest that passing the scope test is not important in assessing the validity of CV value measures but that more research should be focused on developing a clearer definition of the embedding effect and on how testing for different types of effects (scope, context, and order) could be combined in one matrix or protocol for assessing the conformity of stated WTP estimates with economic value theory.

With regard to constructive preferences, results of the initial mail survey suggest that individuals may not have well defined preferences for goods (i.e. dam removal) with which they are not familiar or experienced. Pre-testing of the structured elicitation groups suggests that this alternative elicitation format based on a philosophy of constructive preferences may lead to more thoughtful and rational WTP values, which we believe will be indicated not only by participants' perceptions of the process, but will also be reflected in their WTP values. Our expectations for the SEG are (1) individuals will find the process of translating preferences to dollar values more tractable and we will see fewer non-responses to the OE WTP question and (2) we will see a lower percentage of participants who feel uncertain about their WTP values.

References

Bishop, Richard C., Thomas A. Heberlein, and Mary Jo Kealy. (1983). Contingent valuation of environmental assets: Comparisons with a simulated market. *Natural Resources Journal*, 23, 619-633.

Carmines, Edward G., and Richard A. Zeller. (1979). *Reliability and Validity Assessment*, Sage: Beverly Hills, CA.

Carson, Richard D. and Robert C. Mitchell. (1995). Sequencing and nesting in contingent valuation surveys. *Journal of Environmental Economics and Management*, 28, 155-173.

Carson, Richard T. and Robert C. Mitchell. (1993). The issue of scope in contingent valuation studies. *American Journal of Agricultural Economics*, 75, 1263-1267.

Dillman, D.A. (1978). *Mail and Telephone Surveys,* John Wiley: New York.

Fischhoff, Baruch, Ned Welch, and Shane Frederick. (1999). Construal processes in preference assessment, *Journal of Risk and Uncertainty,* 19, 139-176.

Freeman, A. Myrick, III, *The Measurement of Environmental and Resource Values*, Washington D.C.: Resources for the Future, 1993.

Haab, Timothy C. and Kenneth E. McConnell, *Valuing Environmental and Natural Resources: the Econometrics of Non-Market Valuation*, Northampton, MA: Edward Elgar, 2002.

Heberlein, Thomas A. and Robert Baumgartner. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature, *American Sociological Review*, 43(4), 447-462.

Hoehn, John P. (1991). Valuing multidimensional impacts of environmental policy: theory and methods. *American Journal of Agricultural Economics*, 73(2), 289-299.

Hohen, John P. and Allan Randall. (1989). Too many proposals pass the benefit cost test. *American Economic Review*, 79(3), 544-551.

Kahneman, Daniel and Jack L. Netsch. (1992). Contingent valuation and the value of public goods: reply. *Journal of Environmental Economics and Management*, 22, 90-94.

Kahneman, Daniel and Jack L. Netsch. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22, 57-70.

Loomis, John, Michael Lockwood, and Terry DeLacy. (1993). Some empirical evidence on embedding effects in contingent valuation of forest protection. *Journal of Environmental Economics and Management*, 24, 45-55.

Luce, Mary Frances. (1998). Choosing to avoid: Coping with negatively emotion-laden consumer decisions. *Journal of Consumer Research,* 24, 409-433.

Mitchell, Robert C. and Richard T. Carson, *Using Surveys to Value Public Goods: the Contingent Valuation Method*, Washington D.C.: Resources for the Future, 1989.

Payne, John W., David A. Schkade, and William H. Desvousges. (2000). Valuation of multiple environmental programs. *Journal of Risk and Uncertanity*, 21(1), 95-115.

Payne, John W., James R. Bettiman, David A. Schkade. (1999). Measuring constructed preferences: towards a building code, *Journal of Risk and Uncertainty*, 19(1-3), 243-270.

Payne, John W., James R. Bettman, and Eric J. Johnson. (1993). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, 43, 87-131.

Rabin, Matthew. (1998). Psychology and economics. *Journal of Economic Literature*, 36, 11-46.

Slovic, Paul. (1995). The construction of preference. *American Psychologist,* 50, 364-371.

Smith, V. Kerry. (1992). Comment: Arbitrary values, good causes, and premature verdicts. *Journal of Environmental Economics and Management*, 22, 71-89.

Whitehead, John C., Timothy C. Haab, and Ju-Chin Huang. (1998). Part-whole bias in contingent valuation: will scope effects be detected with inexpensive survey methods. *Southern Economics Journal*, 65(1), 160-168.