Online Survey Data Quality and its Implication for Willingness-to-Pay: A Cross-Country Comparison

Zhifeng Gao, Assistant Professor (zfgao@ufl.edu)
Lisa House, Professor
Jing, Xie, PhD student
Food and Resource Economics Department
University of Florida

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

1. Introduction

The use of online surveys to elicit consumer preference and estimate welfare measures such as willingness-to-pay (WTP) is growing because of the increasing coverage of internet and several advantages of web-based surveys. One potential advantage of online survey is that it is relatively easy to collect data from multiple countries to compare responses to the same questions and therefore contrast consumer attitude, preference, and WTP across countries (Auger, et al., 2010, Okechuku, 1994). However, using online surveys to collect data may result in problems such as lowering the reliability of the data for analysis, or lower data quality. For instance, with an increasing number of survey companies that recruit consumer panels using reward programs (e.g. www.e-Rewards.com , http://us.toluna.com/ , www.panelbase.net), the chance exists that consumers in the panels are motivated by the monetary reward rather than the motivation to express their true opinions or preferences with regard to an event, policy, program, or product. Additionally, the motivation of taking online surveys may be quite different in different countries, and therefore the data quality may differ significantly across countries. If this is the case, applying the same analysis methods to the data from multiple countries may be inefficient.

Previous studies on multi-country comparison using survey data in general assume that respondents answer surveys truthfully, and that respondents in different countries have the same behavior in answering survey questions. However, most recent research indicates that some respondents may not seriously answer online survey questions, and those respondents demonstrate quite different preferences and WTP compared to the respondents who answer the survey questions seriously (Gao, et al., 2012). This implies that developing instruments to control online survey data quality is an important topic for future research. Gao et al.'s research focuses on the US consumer, and it is unknown whether the data quality problem is common across countries and whether the quality of data collected from different countries differs significantly. If respondents in some countries are more likely to give less reliable answers than respondent in other countries, resulting in poor estimates of consumer preferences, we should take additional actions to improve the estimates of consumer preference. Answers to aforementioned questions will provide insightful information to researchers who heavily rely on survey research data and those who consider conducting the cross-country comparison research with online survey data. It may fundamentally change the way of collecting online survey data by including instruments to control data quality.

The objective of this article is to 1) determine whether an online data quality problem is common across countries; 2) whether a significant difference exists in data quality across countries; and 3) whether the difference, if one exists, significantly affects welfare measures such as consumer WTP, which have important implications for policy development and welfare analysis.

2. Background and Literature Review

A search in the Web of Knowledge database with the key words "online survey" or "web based survey" revealed that the number of article published per year using online survey data have increased from less than 10 in the early 90s to more than 1,000 in 2012 (Web of Knowledge, 2013). The increasing popularity of the online survey attributes to several advantages. Past research has shown that online survey can reach unique population that are difficulty to access by other survey modes (face to face interview, mail or phone survey); it can collect survey data much faster and the cost is relatively low; in addition, online survey makes the incorporation of multimedia contents (e.g. video, music) into surveys much easier and researcher can control the survey flow by using logic, and conditions (Cobanoglu, et al., 2001, Dillman, et al., 2011, Griffis, et al., 2003, Wright, 2005). The most discussed key disadvantages of online survey include low coverage of population, sample self-selection as well as the non-response error because participants are not familiar with the survey formation (Dillman, et al., 2011). However, these disadvantages are becoming less a problem with the development in technology and changing patterns in internet access. For instance, the global internet penetration rate has increased from 0.4% in 1995 to 38.8% in March 2013. In 2012, the internet penetration rate was 78.1% for the United States, 81.3% for Belgium, 79.6% for France, 83.0% for Germany, and 79.5% for Japan. For some countries such as Iceland, Sweden etc. the rate was more than 90% (Internet World Stats, 2013). With increased exposure to the internet, consumers are becoming more familiar with online survey format. The establishment of online consumer panels and the used of the reward programs may significantly reduce the sample self-selection problem and non-response errors in the online survey.

The comparison between online survey and other survey modes demonstrate some inconsistent results. Some studies show that samples from online surveys tend to include more males, respondents with younger ages, higher incomes and education levels (Kwak and Radler, 2002, Lindhjem and Navrud, 2011, Marta-Pedroso, et al., 2007, Olsen, 2009). However, others demonstrate that there are no significant differences in respondents' education level (Kwak and Radler, 2002, Lindhjem and Navrud, 2011, Windle and Rolfe, 2011), income (Nielsen, 2011, Windle and Rolfe, 2011) as well as age and gender (Lindhjem and Navrud, 2011, Nielsen, 2011, Olsen, 2009) between online and other survey methods. Most recent research show that no significant differences exist in the qualitative and quantitative data quality between online and other survey modes. For instance, Lindhjem and Navrud (2011), Nielsen (2011) and Marta-

Pedroso, et al. (2007) show that the hypothesis that the means of the welfare measure estimates from contingent valuation methods (CVM) are the same between the face-to-face interview and online survey cannot be rejected. Olsen (2009) and Windle and Rolfe (2011) show that the WTP estimates from choice experiments (CE) do not differ significantly between the online and mail surveys.

Although online surveys have more powerful tool to control survey flow and thus control the sample profile to improve the data quality. The research by Gao, et al. (2012) is the only one that discuss the used of some qualification rules to detect respondents that may give low quality responses. However, Gao et al. only conduct their research in the United States. This article, use the qualification rules to compares the data quality as well as their impacts on consumer WTP estimates across the United States, Germany, Spain, Belgium, France and Japan. With the increasing collaboration between institutes worldwide and multi-country comparisons regarding consumer opinions, preference, attitude etc., the results from this article may have profound impact on the future application of using online survey for data collection.

3. Methods

3.1 Instruments to Measure Data Quality

The survey data quality is significantly affected by respondents' attitude towards taking the survey. If they read and answer the questions carefully and try to provide truthful information and express their opinions truthfully, they are more likely to provide a complete survey with high data quality. Otherwise, the answers provided are more likely to be unreliable or invalid if they carelessly read or even don't read the survey questions (just bubble in answers). As a result, if we include a question that ask respondents to choose a specific answer, survey data from those who give the right answer are more likely to have a higher quality while the data from those who give the wrong answer are more likely to have a lower quality. Defining this question as the validation question (VQ) and using the conditional rules in online survey, respondents who don't pass the VQ can be instantly direct to the end of the surveys to decrease the possibility of collecting low quality data. Several VQ may be placed into a survey, and different qualification rule can be defined such that a respondent will be labeled unqualified if he/she fails one, two, or several VQs.

3.2 *Factor Affecting Data Quality*

Assuming that the probability that a respondent pass the VQ is

(1) $P_i = X_i\beta + e_i$,

then VQ=1 if $P_i > 0.5$, VQ=0, otherwise. If we assume that e_i follows logistic distribution, then the probability that a respondent passes the VQ can be estimated by logit models such that $P(VQ = 1) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$, where *X* are respondent demographic and country variables. The significance of the difference in data quality across countries can be determined by testing the significance of the coefficient of the country variable. The impact of demographics on the data quality across countries can be tested by adding the interaction term between country and demographics.

3.3 Choice Experiment and Impact of Data Quality on WTP Estimates across Countries

Consumer WTP for product attributes can be elicited using data from CE in which multiple attributes are included. In current study, a CE with three attributes using apple as the subject is designed. The attributes are apple production method (Organic, Tradition and Biotechnology /GM), apple production origin (China, New Zealand and Own Country/ the United States, France, Belgium, Germany, Spain and Japan) and price with five levels. The median prices of apples in the survey of a country are similar to the market prices of apples in that country and are presented in the currency of the country (US dollars, Euro or Japanese Yen) where the survey is implemented. The CE is designed using a factorial design that maximizes the D-efficiency of attribute matrix. In the CE respondents are presented with four apples to choice from, with additional "None" option if they do not want to choice any of the apples (**Error! Reference source not found.**)

Based on the random utility theory, consumer utility function can be defined as

(2) $U_{ij} = \beta' \cdot X_{ij} + u_{ij}$,

where X_{ij} is a vector of attributes of product j *and* β is a vector of parameters. Assuming that u_{ij} independently and identically follows Gumbel distribution, and β contains random parameters that measure consumer heterogeneous preference, the parameters in U_{ij} can be estimated using random parameters logit models (RPL). Particularly,

$U_{ij} = \beta_i' \cdot X_{ij} + u_{ij}$ and $\beta_i = \bar{\beta} + \eta \cdot \varepsilon_i$, where $\bar{\beta}$ measure the mean effect of product attributes, η is triangular matrix that is used to calculate the covariance of random parameters $\sum = \eta \cdot \eta'$ and ε_i is independently identically distributed with certain distribution. The probability that a respondent choose a particular product (e.g. *k*) in the CE is $P_{ik} = \int (\frac{e^{\beta_i' \cdot x_{ik}}}{\sum e^{\beta_i' \cdot x_{ij}}}) f(\varepsilon_i) d\varepsilon_i$ (Greene, 2002, Train, 2003).

The parameter of price is specified as non-random and the parameters of other attributes are specified as random parameters following normal distributions. Specifying price coefficient as non-random avoids the possible positive values that are not consistent with economics theory (decreasing demand function with respect to price). In addition, RPL models with all random parameters are barely identified, and randomness of the price coefficient would makes the distribution of the WTPs hard to evaluate (Ruud, 1996, Train, 2003). Additional benefit of not specifying price coeffcient as radom is that the WTP estimate will have the same disribution as the parameters of other attribute. Because we specified all other random parameters to follow normal distribution, the WTP estimates of the non-price attributes will also have normal distributions.

Separate models are estimated for respondents who pass and fail the VQ in each country. This is done because the error u_{ij} in the utility function of the two groups of respondents (in the same country) may be different, and a scale parameter must be estimated in order to pool the data from the two group of respondents (Train, 2003). In addition, comparing the impact of data quality on consumer preferences becomes complicated if all the data are pooled to estimate a single model because more than ten scale parameters need to be estimated. Instead, we focus on the WTP estimates for respondents passing and failing the VQ in different countries. Because when calculating the WTP the scale parameters will cancel out, this enables us to compare the WTP estimates across different groups by using parameters from individual models (Gao and Schroeder, 2009). Base on the estimates of distributions of the preference parameters in consumer utility functions, bootstrap method is used to generate 1,000 values for each parameter (Krinsky and Robb, 1986). Simulated WTP is calculated as the ratio between the parameters of price and attribute, such that $WTP_n = \frac{\beta_k}{\beta_p}$ for respondents who pass the VQ, and $WTP_n{}' = \frac{\beta_k}{\beta_p}$ for respondents who fail the VQ. The simulated WTP estimates can be compared across countries to determine the divergence in WTP after controlling the data quality. The difference in WTP between groups $WTP_n - WTP_n'$ can also be compared across countries to determine whether the data quality has the same influences the WTP estimates in different countries. The analysis is conducted using standard t-test because WTP estimates have normal distributions.

4. Data Collection

In June of 2012, a survey company was hired to distribute online surveys to its national representative consumer panels in the United States, France, Belgium, Germany, Spain and Japan. The survey participants must be the primary grocery shopper of the household, and aged 18 years or older. The same questionnaire was used but was translated into the official language of the corresponding country where the survey was conducted. A VQ was placed in the middle of questionnaire by asking respondents to select a specific answer to a question, such as: “As a validation check, please answer *Strongly Agree* for this question.” If the respondents did not select the answer as guided, they were still allowed to continue the survey to be compared with respondents who passed the VQ. The CE for apples was included in the survey with the purpose to study consumer preference of genetically modified (GM) food and country of origin (COO). Consumer demographics such as age, gender, education etc. were collected at the end of the survey.

5. **Results**

A total of 2,147 survey completes were collected, including 532 completes for the United States, 303 for France, 264 for Belgium, 270 for Germany, 274 for Spain and 504 for Japan. Table 1 reported the statistics of respondents’ demographics and the percentage of respondents that passed the VQ. Results showed that the passing rate of the VQ was the lowest in France (70%), followed by that in Spain (75%), the United States (75%), Germany (76%), Japan (80%) and Belgium (80%). Statistical tests demonstrated that there was no significant difference in gender of respondents across countries, while other demographics of respondents across counties were statistically significantly different at 5% significance level. In general, the US sample had older and young respondents; France and Japan samples had more respondents who had more than 16 years of education; Japan sample had more respondents with annual household income higher than $100,000, while France and Spain samples had more respondents with income less than $30,000 (Table 1).

5.1 Impact of Demographics and Country on Data Quality

The impact of demographics and country on the probability that respondent passing the VQ was determined by specifying the equation (1) as follows:

(3) $P_i = \beta_0 + \sum_{i=1}^{5} \beta_i \cdot X_i + \sum_{j=1}^{7} \alpha_j \cdot Y_j + \sum_{i=1}^{5} \sum_{j=1}^{7} X_i \cdot Y_j + e_i,$

Where X_i are dummy variables of countries such as Belgium, Germany, Japan, Spain, the United State, with the dummy variable of France being removed to avoid dummy trap; Y_j are demographics variables such as age, gender, education, employment status, annual household income and the number of kids in the household (Table 1). One dummy variable of gender was created with male as the benchmark; two dummy variables, Full Time and Part Time were created for employment status, with other employment status as the benchmark. Other demographic variables were treated as continuous variables because these variables were in ordinal scales and there were many categories for each variable. The model was first estimated with all the interactions between country and demographic variables as that specified in equation (3). Results of Model 1 in table 1 indicated that 30 of the interaction variables are not significant. Log likelihood ratio test showed that that hypothesis that all the 30 variables are jointly equal to zero couldn't be rejected. Therefore, we estimated another model only keeping five interaction variables that were at least significant at 10% level in Model 1. Results of Model 2 indicated that of the five variables that were significant in Model 1, only the coefficient of Education*Spain was significant at 10% level. To determine the best model, we further estimated Model 3, in which one interaction variable Education*Spain was included, and the log likelihood ratio test indicated that there was no statistical difference between the estimates of Model 2 and Model 3, which indicated that Model 3 might be more efficient. We estimated another model with all the interaction variables excluded. However, the test indicated that the estimates between Model 3 and Model 4 were significantly different. As results, Model 3 should be considered as the best model to determine the impact of country and demographics on data quality. Results of Model 3 in Table 1 demonstrated that after controlling the effect of demographic variables, respondents in Belgium, Japan and Spain were more likely to pass the VQ, or provided high quality data than French respondents in online survey. Across countries, respondents that were older and those with higher income were more likely to pass the VQ. These results were consistent with some of the results reported by Gao et al. (2012), which showed that income had a positive impact on respondent probability of passing the VQ. Respondents with full time job were less likely to pass the VQ, which was inconsistent with Gao et al.'s study of the US consumers that employment status did not have significant impact. The negative impact of full time employment status on the probability of passing the VQ may be that full time employed respondents had less time to spend in answering surveys, thus might be more careless in reading survey questions.

5.2 Results of Mixed Logit Models across Countries

Table 3 reported the results of mixed logit models for respondents who passed and failed the VQ across countries. The average log likelihood and Akaike information criterion (AIC) values per observations of models for respondents who passed the VQ were smaller than those for respondents who failed the VQ, which implied that the models for respondents who passed the

VQ were more efficient than models for respondents who failed the VQ. This makes sense because respondents failing the VQ might read questions more careless thus made more inconsistent choices to increase the noise in the correspondent models. The results showed that Most of the coefficients were statistically significant at 5% significance levels. For both groups that passed and failed the VQ, the price coefficients were significant and negative. The coefficients of organic and conventional in most models were significant, indicating that consumers in all the six countries were willing to pay a premium for organically or traditionally produced apples to the GMO apples. The significant negative signs of China and New Zealand indicated that consumers were willing to pay a premium for apples that were produced in their own countries. The standard errors of random parameters in all the models were significant, which implied that heterogeneous preferences existed in all the respondents, no matter they passed or failed the VQ and in whichever countries. Most of the elements in the covariances of random parameters were statistically significant, implying possible significant correlations in the random parameters. Comparing the models for respondents who passed and failed the VQ in each country, it could be seen that the signs of parameters between models were in most cases the same, but with different scales. However, because of the impacts of scale parameters as discussed in the method section, we could not draw any conclusion regarding the difference in consumer preferences between respondent groups. The next section compared the WTP between respondents groups to draw conclusion regarding impact of data quality on the estimates of consumer preferences.

5.3 Data Quality and Willingness to Pay

Table 4 reported the statistics of WTP estimates of apple attributes and the differences in WTP estimates for respondents passing and failing the VQ. First, all the WTP for organic and traditional attributes were statistically significantly positive and the WTP for China and New Zealand attributes were statistically significantly negative. This implied that consumers in all the six countries had negative attitudes toward GMO apples and they all preferred domestically produced apple to imported apples. In addition, consumers in all countries had more negative attitudes toward apples from China than those from New Zealand.

Except for the WTP for traditional attribute in France, the mean WTP estimates of all attributes between respondents who passed and failed VQ were significantly different at 5% significance levels in all six countries. The means of WTP estimates of China or New Zealand attributes for respondents who failed the VQ were consistently smaller than those for respondents who passed the VQ. However, the WTP estimates of organic and traditional attributes for respondents failing the VQ were significantly larger than these for respondents passing the VQ in the United States, Belgium and Spain, while the WTP estimates for respondents failing the VQ were significantly smaller than these for respondents passing the VQ in the other three countries. These results imply that the impact of data quality on the means of WTP estimates depended on the attributes and countries.

Most interestingly, among the 24 comparison of WTP estimates between respondents passing and failing the VQ, 22 variances of WTP estimates for respondents who failed the VQ were significantly larger than those for respondents who passed the VQ. This indicates that WTP estimates for respondents who passed the VQ were more efficient than these for respondents who failed the VQ. Considering that respondents who failed the VQ were more likely to read and answer survey questions carelessly, the less precise estimates for these respondents were not surprising because they were more likely to make inconsistent and random choices in CE.

6. Conclusion

With the increasing application of online surveys to study consumer preferences, attitude, and behavior, as well as increased multi-country studies that may be driven by the globalization of collaboration and economy, research on the online survey data quality is crucial. Using validation quality as an instrument to measure data quality, our results show that online data quality problem is common among countries and the quality of the data collected from different countries may vary significantly. The proportions of respondents that failed the VQ range from 31% to 20%. However, after controlling the country effect on the probability of passing the VQ, our results demonstrate that people who are older and with higher income are more likely to provide higher quality online survey data. This implies that when distrusting online survey, more respondents that are young and with lower income should be included to make the respondents who pass the VQ more close to reprehensive samples.

The respondent groups identified by the VQ reveal significant different preferences by the WTP estimates of apple attributes. In general the models for respondents who pass the VQ are more efficient and the WTP estimates from those respondents are more accurate, which indicate that using VQ as a standard practice in online survey can result in more efficient estimates of consumer preferences. The impact of data quality on WTP estimates differ significantly across countries, and sometime have opposite impacts which implies that if data quality is not considered as a factor that affects the estimates of consumer preferences in multicounty studies, misleading conclusions may be drawn by attributing difference to consumer preferences, where it may be a result of the data quality.

More research on developing other similar instruments to improve online data quality is needed with the prevailing applications of online survey. This includes using multiple VQ in surveys and to develop optimal rules to effectively detect respondents who are more likely to provide low quality data as well as develop optimal sampling strategies to recruit respondents who are more likely provide high quality data but at the same time obtaining representative samples.

Reference

Auger, P., T.M. Devinney, J.J. Louviere, and P.F. Burke. 2010. "The importance of social product attributes in consumer purchasing decisions: A multi-country comparative study." *International Business Review* 19(2):140-159.

Cobanoglu, C., B. Warde, and P.J. Moreo. 2001. "A comparison of mail, fax and web-based survey methods." *International journal of market research* 43(4):441-452.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2011. *Mail and Internet surveys: The tailored design method--2007 Update with new Internet, visual, and mixed-mode guide*: Wiley.

Gao, Z., L. House, and X. Bi (2012) Finding True Consumer Attitudes: Do Validation Questions Help?, Agricultural and Applied Economics Association.

Gao, Z., and T.C. Schroeder. 2009. "Effects of label information on consumer willingness-to-pay for food attributes." *American Journal of Agricultural Economics* 91(3):795-809.

Greene, W.H. 2002. *NLOGIT: Version 3.0: Reference Guide*: Econometric Software.

Griffis, S.E., T.J. Goldsby, and M. Cooper. 2003. "WEB-BASED AND MAIL SURVEYS: A COMPARISON OF RESPONSE, DATA, AND COST." *Journal of Business Logistics* 24(2):237-258.

Internet World Stats (2013) Internet usage statistics. The Internet big picture: World Internet users and populations stats, vol. 2013.

Krinsky, I., and A.L. Robb. 1986. "On approximating the statistical properties of elasticities." *The Review of Economics and Statistics*:715-719.

Kwak, N., and B. Radler. 2002. "A comparison between mail and web surveys: Response pattern, respondent profile, and data quality." *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-* 18(2):257-274.

Lindhjem, H., and S. Navrud. 2011. "Are Internet surveys an alternative to face-to-face interviews in contingent valuation?" *Ecological economics* 70(9):1628-1637.

Marta-Pedroso, C., H. Freitas, and T. Domingos. 2007. "Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews." *Ecological economics* 62(3):388-398.

Nielsen, J.S. 2011. "Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews." *Resource and energy economics* 33(1):119-129.

Okechuku, C. 1994. "The Importance of Product Country of Origin:: A Conjoint Analysis of the United States, Canada, Germany and The Netherlands." *European Journal of Marketing* 28(4):5-19.

Olsen, S.B. 2009. "Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods." *Environmental and Resource Economics* 44(4):591-610.

Ruud, P. (1996) Approximation and simulation of the multinomial probit model: an analysis of covariance matrix estimation. Work. Paper, Dept. of Econ., University of California, Berkeley.

Train, K. 2003. *Discrete choice methods with simulation*: Cambridge University Press.

Web of Knowledge (2013), Thomson Reuters.

Windle, J., and J. Rolfe. 2011. "Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys." *Economic Analysis and Policy* 41(1):83-97.

Wright, K.B. 2005. "Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services." *Journal of Computer-Mediated Communication* 10(3):00-00.

Table 1 Statistics of Respondent Demographics across Country

	Country					
	Belgium	Franc	Germany	Japan	Spain	US
Age	%	%	%	%	%	%
<=24	4.55	4.29	4.44	2.78	5.84	6.58
25-34	21.59	24.42	26.30	20.83	22.63	25.56
35-44	25.38	31.68	34.07	36.31	36.86	21.99
45-54	23.48	22.11	21.11	29.37	25.18	15.60
55-64	19.32	12.21	12.22	9.52	8.76	18.98
>=65	5.68	5.28	1.85	1.19	0.73	11.28
Gender						
Male	50.38	48.18	47.41	47.22	51.46	46.99
	49.62	51.82	52.59	52.78	48.54	53.01
Years of School						
1-6 or 7-12 years	16.29	9.90	14.44	3.41	6.20	16.95
13-14 years	18.18	24.75	47.04	30.52	17.88	29.57
15-16 years	38.64	27.72	27.78	51.81	37.59	32.39
>=16 years	26.89	37.62	10.74	14.26	38.32	21.09
Employment						
Full Time	68.94	72.28	75.93	64.88	73.72	55.83
Part Time	7.2	6.27	12.22	10.52	7.66	13.53
Others	23.86	21.45	11.84	24.61	18.6	30.64
Annual Income						
<=$24,999	4.55	8.25	4.44	0.99	9.85	3.20
$25,000 - 29,999	9.47	12.21	4.81	0.00	9.12	3.20
$30,000 - 34,999	12.12	8.58	5.93	0.00	11.31	4.70
$35,000 - 39,999	12.50	14.19	10.37	0.79	18.61	6.95
$40,000 - 49,999	10.98	14.85	12.96	1.39	13.87	12.41
$50,000 - 59,999	11.74	17.16	11.48	2.78	12.41	12.78
$60,000 - 74,999	10.61	6.27	11.85	17.26	8.03	17.86
$75,000 - 99,999	11.36	9.57	22.59	31.15	9.49	19.92
$100,000 - 149,999	9.47	3.63	9.63	30.36	4.01	10.90
$150,000 - 199,999	0.00	0.66	2.22	7.74	0.36	3.38
>=$200,000	7.20	4.62	3.70	7.54	2.92	4.70
# of Kids						
0	55.73	45.87	54.48	49.30	40.15	60.42
1	22.14	22.77	25.00	23.86	37.96	20.08
2	14.89	22.11	17.16	22.07	17.88	13.07
>=3	7.25	9.24	3.36	4.77	4.01	6.44
Passing VQ						
	80.30	68.98	75.56	79.96	74.82	75.38

In the US, the number of years of school is equivalent to 1-6 years- Primary school; 7-12 years-Secondary /High School; 13-14 years-technical, associate, or equivalent; 15-16 years - university graduate; More than 16 years-Post university/Masters, Ph.D.

Table 2 Results of Logit model for the Impact of Factors Affecting the Probability of Passing VQ

	Model 1[a]		Model 2		Model 3		Model 4	
Variable	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Intercept	-0.02	0.98	-0.35	0.38	-0.39	0.30	-0.25	0.50
Belgium	0.11	0.94	0.27	0.64	0.55***	0.01	0.55***	0.01
Germany	0.00	1.00	-0.25	0.67	0.31	0.12	0.30	0.13
Japan	0.53	0.70	1.24*	0.07	0.36**	0.05	0.37**	0.04
Spain	0.43	0.74	1.11	0.19	1.48**	0.05	0.35*	0.07
USA	-0.52	0.62	0.23	0.19	0.17	0.32	0.16	0.33
Age	0.21*	0.07	0.21***	0.00	0.20***	0.00	0.20***	0.00
Female	0.10	0.71	0.13	0.24	0.13	0.22	0.12	0.27
Education	0.12	0.36	0.06	0.38	0.02	0.74	-0.01	0.91
Employment								
Full Time	-0.48	0.20	-0.38***	0.01	-0.39***	0.01	-0.40***	0.01
Part Time	0.13	0.84	-0.16	0.43	-0.17	0.41	-0.18	0.39
Income	-0.03	0.51	0.03	0.25	0.06***	0.01	0.06***	0.01
# of Kids	0.01	0.90	0.01	0.78	0.02	0.76	0.02	0.76
Education*Spain	-0.44**	0.05	-0.35*	0.06	-0.27	0.12		
Education*Japan	-0.38*	0.07	-0.20	0.23				
Income*Belgium	0.13*	0.10	0.04	0.58				
Income*Germany	0.14*	0.06	0.07	0.30				
Income*Spain	0.16**	0.04	0.09	0.19				
-2 Log L	2244.346		2272.267		2276.151		2284.958	
# Variables	48[b]		18		14		13	
P-Value		0.574[c]		0.422		0.003		

Notes: *** indicates statistically significant at 1% significance level; ** indicates statistically significant at 5% significance level; * indicates statistically significant at 10% significance level; a- the estimates of all there 30 interaction variables that are not statistically significant at 10% significance level are not reported to save space; b- Number of variables in the model; c-The probability that the χ^2 statistics is bigger than the critical value for the hypothesis that there is no significant difference between two models (e.g. Model 1 vs. Model 2, Model 2 vs. Model 3, and Model 3 vs. Model 4).

Table 3 Results of Mixed Logit Models for Respondents Who Passed and Failed VQ across Countries

Countries	US		France		Belgium		Germany		Spain		Japan	
Variables	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass
Random Parameters in Utility Functions												
Organic	3.21***	1.97***	0.47**	2.86***	1.31***	0.71***	0.31	2.19***	1.06***	1.07***	0.39	2.87***
Traditional	2.46***	2.03***	0.72***	2.77***	1.80***	2.00***	1.00***	3.35***	1.25***	2.64***	0.94***	2.85***
China	-3.22***	-4.42***	-4.68***	-4.91***	-4.45***	-5.91***	-4.25***	-3.98***	-3.99***	-3.57***	-6.49***	-7.20***
New Zealand	-1.52***	-1.73***	-1.88***	-1.68***	-2.20***	-2.85***	-2.30***	-1.57***	-1.96***	-1.56***	-1.88***	-1.93***
Non Random Parameters in Utility Functions												
Price	-1.58***	-2.50***	-0.60***	-2.28***	-1.21***	-1.93***	-0.59***	-1.64***	-0.97***	-2.43***	-0.005***	-0.01***
Constant	-3.68***	-5.06***	-2.03***	-4.69***	-2.12***	-4.11***	-1.16***	-2.53***	-2.21***	-5.05***	-2.52***	-5.08***
Diagonal Values in Cholesky Matrix												
Organic	4.80***	3.86***	2.11***	3.11***	2.26***	3.68***	2.43***	2.90***	2.08***	3.49***	2.92***	2.77***
Traditional	1.34***	1.40***	1.03***	1.36***	1.53***	1.72***	1.31***	2.15***	2.74***	1.45***	0.61***	0.40***
China	4.00***	3.39***	3.10***	4.87***	4.99***	3.69***	4.27***	4.01***	5.25***	3.23***	5.22***	1.25***
New Zealand	0.81***	1.31***	1.83***	1.16***	1.50***	1.39***	0.93***	0.88***	1.52***	1.08***	1.00***	1.56***
Covariances of Random Parameters												
Traditional: Organic	-9.11***	7.64***	-2.52***	5.86***	1.65***	6.00***	3.53***	-4.95***	-1.33***	-7.92***	4.74***	-5.89***
China: Organic	4.35**	-7.97***	1.47***	-3.06***	0.68	-8.34***	0.96	3.97***	-2.87***	6.20***	5.24***	3.27***
China: Traditional	0.25	-4.60***	-1.46***	1.83**	0.93	-7.87***	0.51	-1.11	-3.22**	-4.73***	0.79	-1.34**
New Zealand: Organic	1.67	-4.18***	0.22	-2.43***	0.41	-5.29***	2.33***	2.12***	0.17	2.83***	3.08***	2.28***
New Zealand: Traditional	0.44	-2.74***	0.02	-0.81*	-0.09	-4.82***	1.43**	-1.04**	-1.33	-2.23***	1.43***	-1.37***
New Zealand: China	12.08***	8.59***	-4.68***	14.18***	6.15***	-2.71*	12.73***	10.04***	-14.30***	-4.63***	-8.34***	2.46***
Standard Deviations of Random Parameter Distributions												
Organic	4.80***	3.86***	2.11***	3.11***	2.26***	3.68***	2.43***	2.90***	2.08***	3.49***	2.92***	2.77***
Traditional	2.32***	2.42***	1.58***	2.32***	1.69***	2.37***	1.96***	2.75***	2.81***	2.69***	1.73***	2.17***
China	4.36***	3.98***	3.23***	5.65***	5.02***	4.97***	4.29***	4.28***	5.63***	3.72***	6.54***	3.43***
New Zealand	2.90***	2.53***	2.38***	2.89***	1.95***	3.57***	3.19***	2.52***	3.25***	2.36***	2.73***	2.30***
# Observations (N)	1908	6174	1332	3258	720	3168	972	2880	1044	3132	1476	6066
Log Likelihood/N	-0.92	-0.84	-0.95	-0.80	-0.91	-0.74	-0.88	-0.82	-0.84	-0.79	-0.90	-0.78
AIC/N	1.85	1.68	1.93	1.61	1.87	1.50	1.79	1.57	1.71	1.58	1.82	1.56

Notes: *** indicates statistically significant at 1% significance level; ** indicates statistically significant at 5% significance level; * indicates statistically significant at 10% significance level.

Table 4 Statistics of WTP Estimates of Respondents Who Passed and Failed VQ across Countries

WTP for	Organic			Traditional			China			New Zealand		
Country	Fail	Pass	Fail - Pass	Fail	Pass	Fail - Pass	Fail	Pass	Fail - Pass	Fail	Pass	Fail - Pass
US	2.02 (2.98)	0.78 (1.51)	1.24[a] [1.47]	1.55 (1.48)	0.81 (0.95)	0.75 [0.52]	-2.05 (2.78)	-1.76 (1.57)	-0.28 [1.22]	-0.95 (1.85)	-0.68 (1.01)	-0.28 [0.84]
France	0.77 (3.44)	1.25 (1.34)	-0.48 [2.10]	1.19 (2.64)	1.21 (1.00)	**-0.02** [1.64]	-7.77 (5.50)	-2.16 (2.47)	-5.61 [3.02]	-3.06 (4.03)	-0.73 (1.27)	-2.33 [2.76]
Belgium	1.08 (1.84)	0.36 (1.87)	0.72 **[-0.04]**	1.48 (1.41)	1.03 (1.22)	0.45 [0.20]	-3.70 (4.21)	-3.05 (2.57)	-0.65 [1.65]	-1.80 (1.66)	-1.45 (1.88)	-0.35 [-0.22]
Germany	0.51 (4.04)	1.33 (1.73)	-0.82 [2.31]	1.68 (3.28)	2.04 (1.70)	-0.36 [1.58]	-7.22 (7.38)	-2.43 (2.65)	-4.78 [4.73]	-3.87 (5.52)	-0.95 (1.56)	-2.92 [3.95]
Japan	82.06 (615.60)	264.60 (250.90)	-182.60 [364.70]	199.90 (364.60)	263.90 (197.40)	-64.02 [167.20]	-1391.60 (1454.00)	-666.70 (318.80)	-724.90 [1135.20]	-399.00 (580.00)	-175.70 (214.90)	-223.30 [365.10]
Spain	1.09 (2.12)	0.44 (1.41)	0.66 [0.71]	1.28 (2.98)	1.09 (1.11)	0.19 [1.87]	-4.12 (5.85)	-1.47 (1.57)	-2.65 [4.29]	-1.99 (3.40)	-0.63 (0.96)	-1.35 [2.44]

Notes: a - difference in WTP estimates for respondents failing and passing the VQ; WTP for respondents in the United States, EU and Japan are in dollars, euros, and Japanese Yen, respectively; all the values are statistically significant at 5% significance level, except for the values that are in bold; numbers in the parentheses are standard deviations of WTP estimates; numbers in square brackets are difference in the standard deviations of WTP estimates.