

## **Missing Data in Afghanistan: A Latent Class Examination of Item Non-Response**

**Ann Arthur, Jill Heese, Allan McCutcheon**

Gallup Research Center, University of Nebraska-Lincoln, 933 Oldfather Hall, Lincoln,  
NE 68588

### **Abstract**

Previous research found that Afghan women had much higher item non-response on political questions than men (Heese & Arthur, 2012). The goal of this research is to investigate response patterns using latent class analysis. Results indicate that a three-class model fit the data well, yielding a latent class of good responders and two latent classes with higher item non-response. These classes are elucidated with descriptive statistics and results of a multinomial regression model. The research provides supporting evidence for Krosnick's (1991) satisficing theory, which posited two broad groups of respondents ("optimizers" and "satisficers," respectively), and further divided the satisficers into weak and strong satisficers. The classes fit this description, with a large class of optimizers and two classes of satisficers, who are less willing or able to engage in the survey response process.

**Keywords:** Latent class analysis, item non-response, Afghanistan

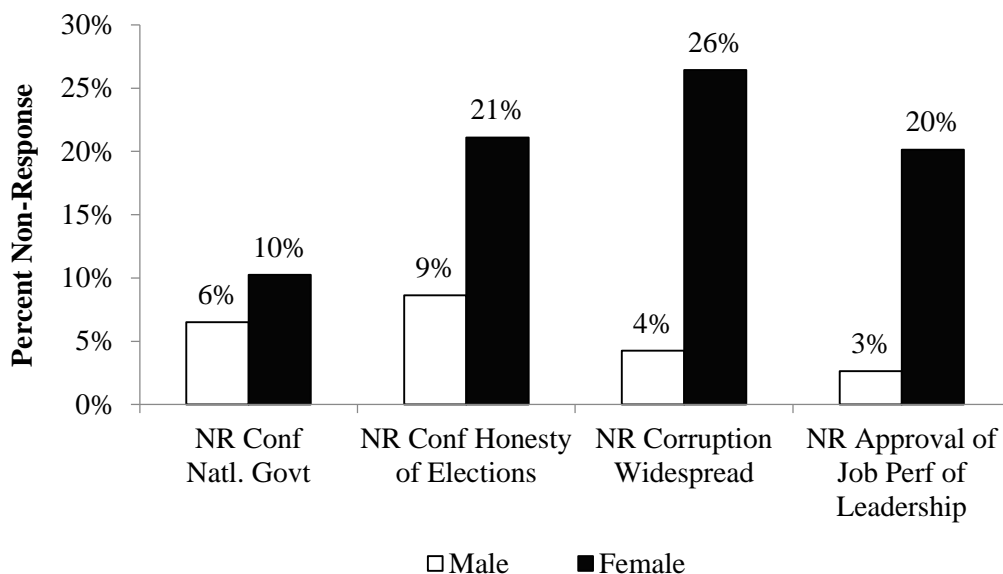
Afghans have lived through decades of instability, civil war, and control by oppressive regimes. In December 1979, the Soviet Union invaded Afghanistan, occupying the country for nearly a decade in what would become known as the Soviet War in Afghanistan. In their wake, they left poverty, civil unrest, and internal war. The Taliban emerged from the warring tribes to fill the power vacuum by means of oppression and violence. The Taliban were very restrictive of the rights of the Afghan citizens, particularly females. Under the Taliban, women were not permitted to work or attend school and could not leave the house without a man. Notably, women were only given limited medical treatment, which resulted in one of the highest infant mortality rates in the world (Central Intelligence Agency, 2001). Following the September 11<sup>th</sup> attacks, the U.S. invaded Afghanistan to begin a "War on Terrorism," and is scheduled to withdraw the remaining combat troops by the end of 2014. Afghanistan held its third presidential election in April 2014, but not everything has gone according to plan. Unfortunately, this election was similar to the presidential election of 2009, in that it was plagued with voter fraud and violence. Thus, Afghanistan is still in the infancy of its democracy.

In his political philosophy, *Politics*, Aristotle (2009) argued that citizen participation is a necessary component for establishing a stable democracy. He stated that when all citizens, both rich and poor, participate in the government, the decisions made are of higher quality. More recently, Purcell (2008) elaborates on the viewpoint:

For Aristotle and for participatory democrats, “the best thing” is to be a citizen engaged in political congress with one’s peers. To realize their best self, citizens must develop their civic virtue, which is to say their capacity to discern and act in favor of the common interest....As citizens develop their civic virtue, they become wiser, more just, more capable decisions-makers...The more a policy encourages and enables participation, the argument goes, the wiser the decisions of its citizens will be, to the benefit of the polity as a whole. So participation is not just for the development of each citizen, it is also good for the group. (p. 53)

Thus, it is advantageous to allow people to participate in the democratic process, since better decisions are made on behalf of the common good.

Unfortunately, previous research has indicated that Afghan women may participate in the democratic discourse less than men, in that Afghan women gave fewer substantive responses to political questions than men (Heese & Arthur, 2012). Figure 1 illustrates non-response by gender to four political questions. As can be seen, there is much greater non-response from women. Since the inability of Afghan women to answer questions about political topics may be a potential reason behind the struggle to establish democracy in Afghanistan, the goal of this current research project is to investigate response patterns in Afghan women. Specifically, did the women participants all have similar levels of non-response, or are there multiple groups with different levels of non-response?



**Figure 1.** Non-response on four political questions by gender.

## Methods

### *Participants*

The Gallup World Poll is a probability-based, multinational general population survey conducted several times per year in more than 160 countries. The data for this study consist of a subset of 1,000 Afghan women collected during two time points in 2012 (April and August).

### *Measures*

The Gallup World Poll consists of a core questionnaire, along with additional region- and country-specific questions. The data were collected using face-to-face interviews with gender matching at the final stage (i.e., women interviewing other women). Interviews were conducted in Dari and Pashto.

The four political questions studied are as follows: 1) In this country, do you have confidence in...the national government? 2) In this country, do you have confidence in...the honesty of elections? 3) Is corruption widespread throughout the government in this country or not? 4) Do you approve or disapprove of the job performance of the leadership of this country?

While these questions are not particularly sensitive in nature, it is possible given the unrest in the country that some respondents may have found these questions to be of a delicate nature.

#### *Analysis Plan*

Latent class analysis (McCutcheon, 1988, 2002) is a technique for categorical data that is used to identify unobservable class membership amongst survey participants. Latent class analysis was used to test whether Afghan women all have similar levels of item non-response, and to identify how many latent groups are needed to represent item non-response in Afghan women. Each of the four variables were recoded as 1 = substantive response (e.g., agree/disagree, yes/no) and 0 = unsubstantive response (e.g., don't know, refused). Also peculiar is that there were no refusals on any of the four questions. Latent class analysis was conducted using LEM (Vermunt, 1997). Once the latent class solution was obtained, a regression model was developed to predict membership in the latent classes using Stata (StataCorp, 2011).

It was hypothesized that two latent classes would emerge; however, models with two, three, and four latent classes were tested. To assess model fit, the Pearson  $\chi^2$ , Likelihood-ratio  $G^2$ ,  $-2LL$ ,  $AIC$ , and  $BIC$  were used. Based on these statistics, the three-class solution fit significantly better than the two-class solution.

Table 1 provides the probabilities of giving a substantive response for the unrestricted three-class model. Additionally, it provides the relative frequency, or size, of each latent class in the population. All results have been weighted using Gallup's sampling weights.

**Table 1:** Probability of Giving a Substantive Response and Relative Frequency for the Three-Class Model

Observed Variables	Respondent Type		
	Latent Class 1	Latent Class 2	Latent Class 3
Confidence in National Gov't	1.000	0.631	0.665
Confidence in Honesty of Elections	0.946	0.536	0.253
Corruption in Government	0.905	0.588	0.013
Approval of Country's Leadership	0.966	0.728	0.000
Relative Class Frequency	0.710	0.155	0.135

#### *Describing the Latent Classes*

Latent Class 1 is characterized by high probabilities of providing a substantive response on all four questions and represents of 71.0% of Afghan women. Latent Class 2 has lower response probabilities ranging from .54 to .73, so these respondents provide substantive responses at least half of the time. This class represents 15.5% of Afghan women. The third class is characterized by a mix of response probabilities, including near-zero probabilities on two questions. Latent Class 3 represents 13.5% of Afghan women.

Since these latent classes represent the response patterns, it is helpful to refer to each class/group with descriptive names. The first latent class represents classic good respondents and will be referred to as the “Engaged” class since they seemed to engage in the response process. The second latent class seemed to be more reluctant to give substantive answers; therefore, they will be referred to as the “Reluctants.” The third latent class is named the “Traditionalists,” as this group appears to represent women from very traditionalist portions of society, where women do not typically express much of an opinion on anything outside the home.

#### *The Restricted Model*

Three restrictions were then imposed to values that were near or at the boundary values of 0 and 1. Fixing these values at 0 or 1 frees up the degrees of freedom used to estimate those parameters. Model fit for the restricted three-class model fit was not significantly worse than the unrestricted three-class model ( $\Delta\chi^2(3) = 0.012, p = .999$ ). The final model, seen in Table 2, has very similar probabilities and relative frequencies as the unrestricted model, with the general pattern being the same. With the restrictions in place, the Engaged class now represents 70.8% of the population, the Reluctants are 16.1%, and the Traditionalists are 13.2%.

**Table 2:** Probability of Giving a Substantive Response and Relative Frequency for the Restricted Three-Class Model

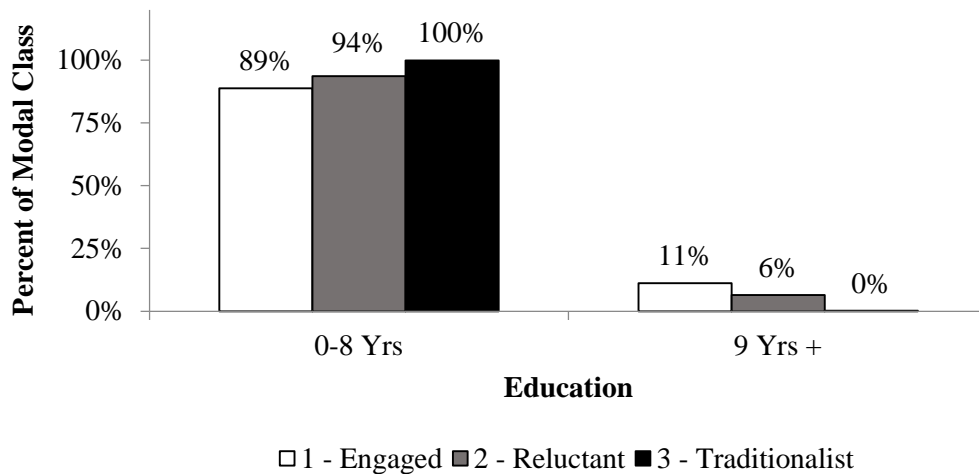
Observed Variables	Respondent Type		
	Latent Class 1	Latent Class 2	Latent Class 3
Confidence in National Gov't	1.000 <sup>a</sup>	0.639	0.664
Confidence in Honesty of Elections	0.947	0.534	0.251
Corruption in Government	0.906	0.591	0.000 <sup>a</sup>
Approval of Country's Leadership	0.967	0.714	0.000 <sup>a</sup>
Relative Class Frequency	0.708	0.161	0.132

<sup>a</sup> Restricted values.

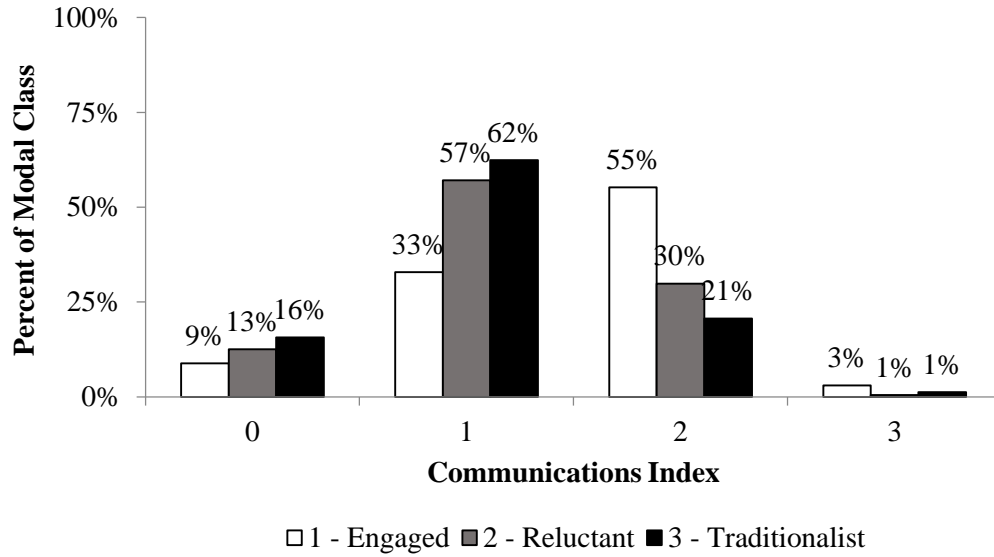
#### *Respondents in These Classes*

Once the latent class solution was obtained, attention was turned to the individuals within each latent class. For this, the modal assignment was used. Since each individual is given a probability of being in each of the three latent classes, the modal assignment classifies each individual to the class that has the highest probability (McCutcheon, 1987). The modal assignment contains some classification error, although there is no way to distinguish which individuals should be in a latent class and which should not be. However, it is commonly used to extract individual-level classifications (McCutcheon, 1987, 2002) as an indicator for membership in the latent classes.

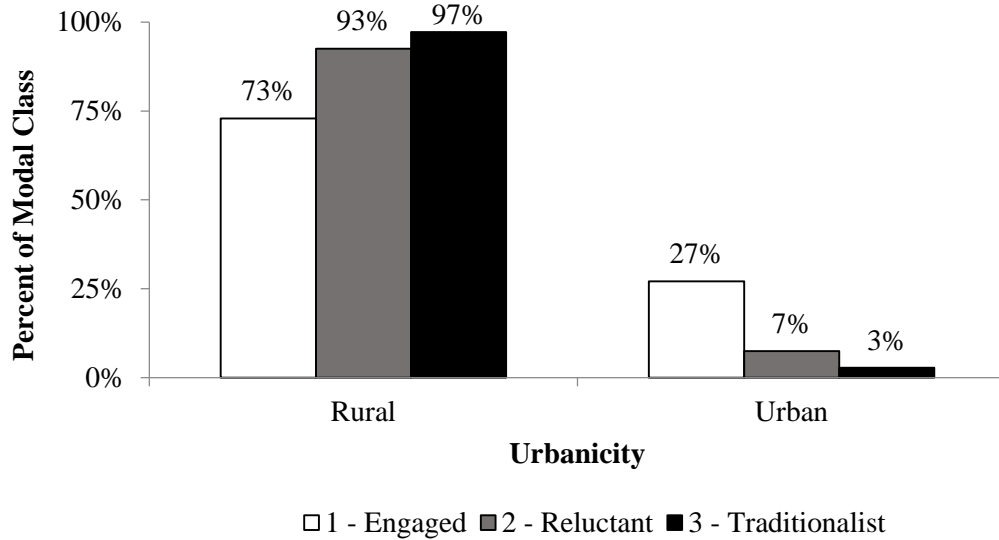
To give a better sense of the women in each latent class, they were examined using several demographic variables (see Figures 2-7). The majority of women in Afghanistan have eight or less years of education, but of the women who have additional education, the Engaged group is largest (see Figure 2). The Communications Index ranges from 0-3 and is a summary score of: access to a television, access to a phone (either cell phone or landline), and access to the internet in the home. As shown in Figure 3, very few women have access to all three forms of communication. However, the Engaged group is most likely to have access to more forms of communication and thus is more connected to current events. Figure 4 reflects urbanicity and rurality. Much of the country is still rural, but of women who live in urban areas, most are Engaged. Figure 5 is “Do you feel safe walking alone at night in the city or area where you live?” Again, the majority of women across all classes do not feel safe. However, the Traditionalists group was the least likely to feel safe walking alone at night. Figure 6 shows the number of children under 15 that are now living in the household. While large families (e.g., 10 or more children) are not uncommon, the Engaged group tends to have fewer larger families and the Traditionalist group tends to have larger families. The entire country is Muslim, and the majority religion is Sunni Islam. However, the Traditionalist group is most likely to follow Sunni Islam (Figure 7). Not all respondents specified Sunni or Shiite, 14.9% said “Muslim” but did not go on to specify. It is possible that religion was a sensitive question for who did not specify, especially for those in the minority, Shiite Islam.



**Figure 2.** Modal class by education.



**Figure 3.** Modal class by communications index. Respondents were given one point on the Communications Index for each of the following: access to a telephone (either landline or cellular), access to the internet, and access to television.



**Figure 4.** Modal class by urbanicity.

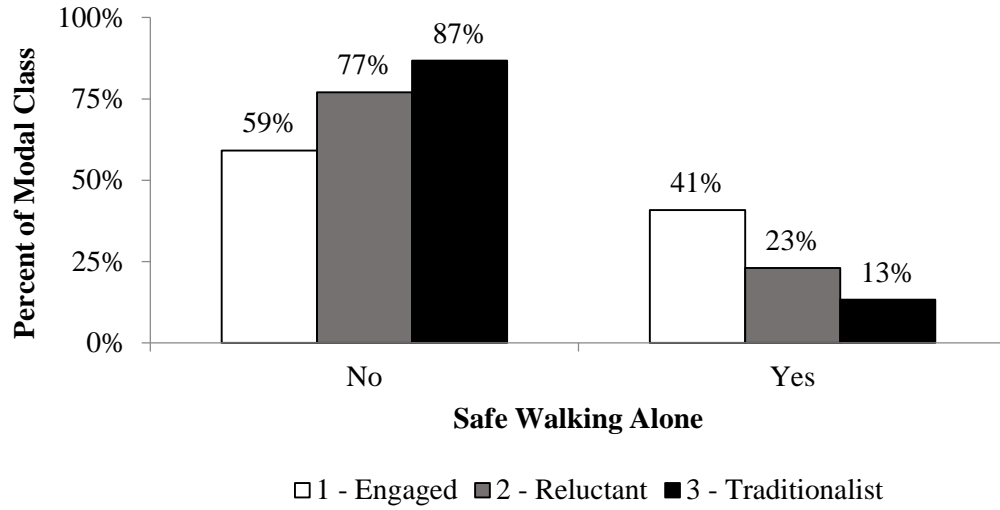


Figure 5. Modal class by safe walking alone.

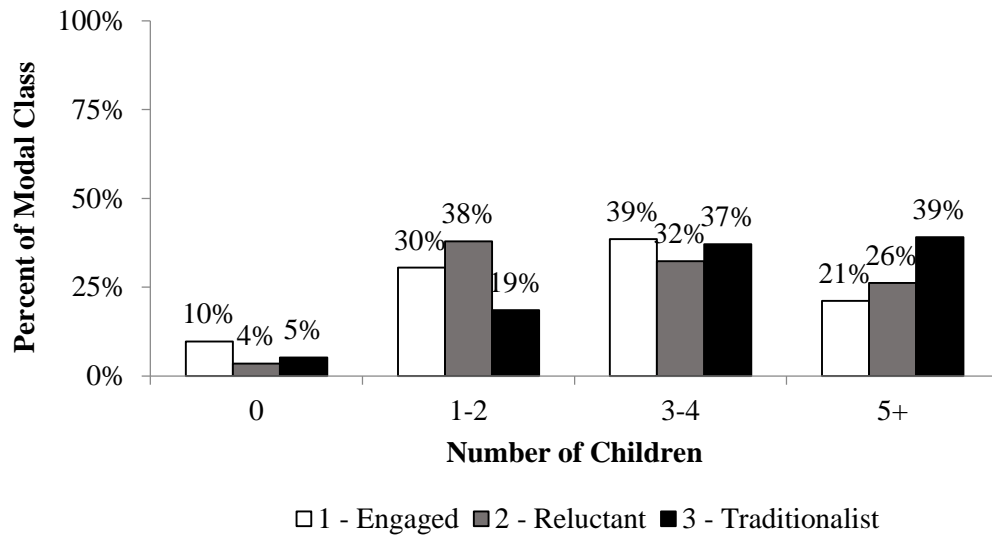
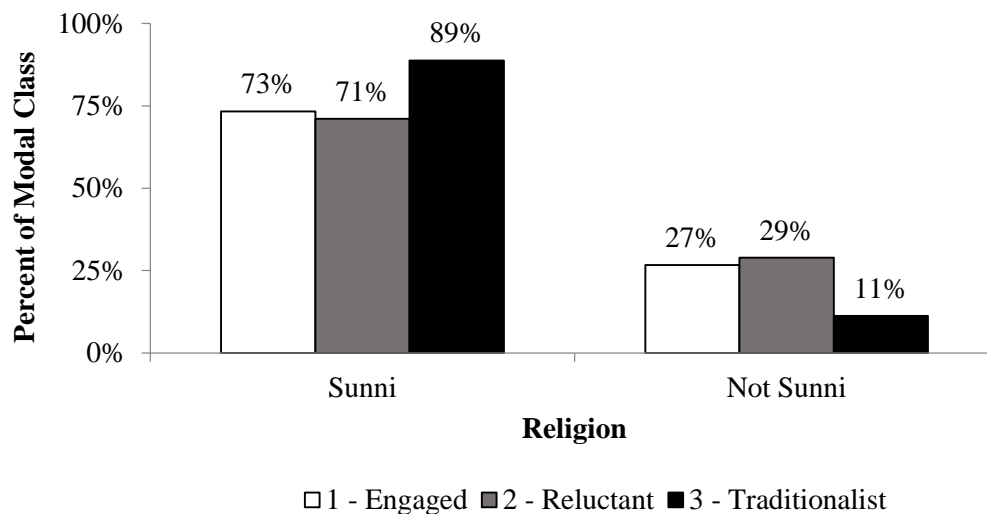


Figure 6. Modal class by number of children less than 15 years of age.



**Figure 7.** Modal class by religion.

#### *Predicting Modal Assignment*

As a final step, a regression model was developed to predict modal assignment. All variables in Figures 2-7 were included, as well as additional variables that are pertinent to women, such as household income, employment by an employer, “Were you treated with respect all day yesterday?” and “Are you satisfied or dissatisfied with your freedom to choose what you want to do with your life?” Initially, two-way interactions with age and education were included in the model, but implausibly large standard errors resulted. This problem is often associated with sparseness from severely imbalanced variables. Most Afghan women are very young, reflecting a low life expectancy. Also, as noted earlier, the vast majority of Afghan women have less than eight years of education. To resolve these issues, a simpler model without interactions was pursued.

For the regression model, two types were considered: adjacent categories and multinomial. The classes seem to be ordered in terms of good, bad, and worse respondents, so an adjacent categories model would have reflected this ordering. Ultimately, a multinomial model was selected (see Table 3), with the Engaged group as the referent, for two reasons. First, the Engaged group is the largest, making it also suitable as a referent. Second, the substantive interpretation it afforded was more important. This model allows for the examination of the groups of bad and worse respondents to attempt to determine why they are not good respondents.

Compared to the Engaged, the Reluctant class has significantly lower communications index scores and live in significantly less urban areas. The women in the Traditionalist group, compared to the Engaged, not only have significantly lower communications index scores and live in less urban areas, but also feel less safe, have more children, have lower education, and are more likely to be Sunni Islam.



**Table 3.** Multinomial Regression Predicting Modal Assignment with Engaged as Referent

Variable	Reluctant				Traditionalist			
	B	SE	RRR	95% CI	B	SE	RRR	95% CI
Lower Comm_Index	0.61	0.21	1.84**	[1.22, 2.79]	0.58**	0.16	1.79**	[1.29, 2.47]
Less Safe	0.54	0.35	1.72	[0.86, 3.41]	1.07**	0.32	2.92**	[1.57, 5.43]
More Children	-0.01	0.07	0.99	[0.87, 1.14]	0.12**	0.05	1.13**	[1.03, 1.24]
Less Education	-0.18	0.48	0.83	[0.33, 2.13]	3.17**	1.02	23.69**	[3.18, 176.33]
Less Urban	1.30*	0.60	3.66*	[1.13, 11.80]	1.80**	0.70	6.02**	[1.54, 23.63]
Sunni	-0.03	0.36	0.97	[0.48, 1.98]	0.95*	0.37	2.58*	[1.24, 5.34]
(constant)	1.22	0.40	3.39	[1.55, 7.42]	0.64*	0.29	1.90*	[1.07, 3.37]

Note. RRR = relative risk ratio. 95% CI is for RRR.

\* $p < .05$ . \*\* $p < .01$ .

### Discussion

Three latent classes were detected, representing three groups of respondents. The classic good respondents, the Engaged, represent 71% of respondents, which is reassuring for those conducting survey research in Afghanistan. On the other hand, this means that nearly 30% of respondents are grouped into the other two classes and represent less than desirable respondents. Examination of the modal assignments facilitates understanding as to why these two other classes provide fewer substantive responses.

The regression results fit nicely with existing theory on survey response. Krosnick's (1991) satisficing theory posited two broad groups of responders (which he calls "optimizers" and "satisficers"), and further divided the satisficers into weak and strong satisficers. The current research reflects this layout, with a large class of optimizers, the Engaged, and two classes of satisficers, who are less willing or able to engage in the survey response process for one reason or another. Here, analyses show that the Reluctant class had lower access to communications and lower urbanicity (more rural domiciles) than the Engaged group, while the Traditionalist class had several additional factors that interfere with the response process. For instance, young children in the household could be a distraction for the respondent, affecting the quality of responses. This current research therefore contributes supporting evidence for Krosnick's satisficing theory.

What this research was unable to answer was the Reluctants' and Traditionalists' *motivations* for answering "Don't know." However, there are several possibilities for this. First, low cognitive ability could be the primary culprit, whether due to low education, a lack of direct experience with the question content (e.g., corruption), low exposure to information about the topic from others or the media, or a combination of these. Another distinct possibility is social desirability. It could be a cultural norm in Afghanistan to say "Don't know" rather than to say something ill of your country. Finally, this could reflect a "fear effect." Some of the most rural areas are the most contested areas in Afghanistan, so it is possible that the "Don't know" responses were a deflection to avoid giving answers that might put respondents at risk. Additional research could examine the effects on the estimates for these questions by either imputing a substantive response or examining the bias by excluding them. Finally, third-person effects may have played a role in the response process. Anecdotally, there were reports of male relatives attending these interviews and possibly influencing the women's responses: after the interviewer asked a question, the male relative would whisper into the respondent's ear, and then the respondent would give her answer. Thus it is likely that the male relative influenced the responses, although there is no way of knowing what the male relatives said. However, Gallup does code for whether

a third person was present during an interview, so future research may be able to further investigate these third-person effects.

Many Afghan women came of age during a long period of political instability. Severe oppression by the Taliban produced a country where women were not allowed to participate in education or the workforce, and the consequences of this oppression are still felt in Afghanistan today. While many factors may contribute to item non-response, this research provides insight into the driving factors behind the Afghan women's non-response to political questions. This research indicates that as Afghanistan struggles to establish democracy, it must encourage women, in particular, to engage in the national political discourse.

## References

- Aristotle (2009). *Politics: A treatise on government*. Retrieved from <http://www.gutenberg.org/files/6762/6762-h/6762-h.htm> (Original work published 350 B.C.E.)
- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Central Intelligence Agency (CIA). (2001). *The world factbook: Afghanistan*. Retrieved November 1, 2011, from <https://www.cia.gov/library/publications/the-world-factbook/index.html>
- Goodman, L. (2002). Latent class analysis. In J. A. Hagenaars and A. L. McCutcheon (Eds.) *Applied latent class analysis*, (pp. 3-55). New York: Cambridge University Press.
- Heese, J. M. & Arthur, A. M. (2012). Effects of Afghanistan's 2009 presidential election on public opinion. Paper presented at the annual meeting of the American Association of Public Opinion Research (AAPOR), Orlando, FL.
- Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- McCutcheon, A. L. (1987). *Latent class analysis*. Thousand Oaks, CA: Sage.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. In J. A. Hagenaars and A. L. McCutcheon (Eds.) *Applied latent class analysis*, (pp. 3-55). New York: Cambridge University Press.
- Purcell, M. (2008). *Recapturing democracy: Neoliberalization and the struggle for alternative urban futures*. New York: Routledge.
- StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.
- Vermunt, J.K. (1997). LEM: A general program for the analysis of categorical data. Department of Methodology and Statistics, Tilburg University.