

**Confidence Elicitation and Anchoring in the Respondent-Generated Intervals (RGI) Protocol**

LiPing Chu<sup>1</sup>, S. James Press<sup>2</sup>, and Judith M. Tanur<sup>3</sup>

**Abstract**

The Respondent-Generated Intervals protocol (RGI) has been used to have respondents recall the answer to a factual question by giving not only a point estimate but also bounds within which they feel it is almost certain that the true value of the quantity being reported upon falls. Here, we elaborate the RGI protocol with the goal of improving the accuracy of the estimators by introducing cueing mechanisms to direct confident (and thus presumably accurate) respondents to give shorter intervals and less confident (and thus presumably less accurate) respondents to give longer ones. Because point estimates of respondents who give short intervals are weighted more heavily in the Bayesian RGI estimator than are point estimates of respondents who give longer intervals, it is advantageous to encourage respondents to respond accordingly. We describe some preliminary results of an experiment embedded in a survey we have designed to test this new thinking. The experimental design varies the instructions about how respondents should construct their intervals. Moreover, the analysis of numerical results uses a new estimation procedure. Preliminary results suggest that the new cueing procedure is effective.

**Key Words:** RGI, Surveys, Bayes, Hierarchical Models, Record-Checks, Anchoring

**1. Introduction**

The Respondent-Generated Intervals protocol (RGI) has been used to have respondents recall the answer to a factual question by giving not only a point estimate but also bounds within which they feel it is almost certain that the true value of the quantity being reported upon falls (Press, 2004). This paper reports on new thinking that aims to elaborate the RGI protocol with the goal of improving the accuracy of the estimators derived from the protocol.

There are two aspects to the new thinking. The first is a new analytical Bayesian procedure for estimating the population mean in an RGI survey; it is derived in the Appendix to Chu, Press, and Tanur (2004). The second is a new “anchoring” type of questioning technique that cues and encourages confident (and presumably accurate) respondents to give short intervals and less confident (and presumably less accurate) respondents to give long intervals. The new analytical procedure is summarized

briefly in Section 2 (and elaborated in the Appendix to Chu, Press, and Tanur, 2004, a published version of the current paper). Section 3 discusses a classroom survey experiment and how it incorporates the new questioning technique. Section 4 provides a discussion of the implications of these innovations.

**2. Vague Prior Bayesian Point Estimator for the Population Mean**

For a sample of  $n$  independent respondents in a survey, let  $y_i, a_i, b_i$  denote the basic usage quantity response, the lower bound response for where the true value to the question lies for that respondent, and the upper bound response for where the true value to the question lies for that respondent, respectively, of respondent  $i, i = 1, \dots, n$ . Suppose that the  $y_i$ 's are all normally distributed. Suppose also that we adopt a vague prior distribution for the population mean,  $\theta_0$ , to represent knowing little, a priori, about the value of the population mean. It is shown in Press, 2004 using a hierarchical Bayesian model, that in such a situation, the posterior distribution of  $\theta_0$  is given by:

$$(\theta_0 | data) \sim N(\tilde{\theta}, \omega^2), \tag{2.1}$$

where the posterior mean,  $\tilde{\theta}$ , is expressible as a weighted average of the  $y_i$ 's, and the weights are dependent upon the intervals defined by the bounds, the smaller the interval the larger the weight. The posterior variance is denoted by  $\omega^2$ . The posterior mean is expressible as:

$$\tilde{\theta} = \sum_1^n \lambda_i y_i, \tag{2.2}$$

where the  $\lambda_i$ 's are non-negative weights that are given approximately by:

$$\lambda_i \doteq \frac{\left( \frac{1}{\frac{(b_i - a_i)^2}{k_1^2} + \frac{(b_0 - a_0)^2}{k_2^2}} \right)}{\sum_{i=1}^n \left( \frac{1}{\frac{(b_i - a_i)^2}{k_1^2} + \frac{(b_0 - a_0)^2}{k_2^2}} \right)}, \quad \sum_1^n \lambda_i = 1, \tag{2.3}$$

<sup>1</sup> Dept. of Statistics, University of California, Riverside  
<sup>2</sup> Dept. of Statistics, University of California, Riverside  
<sup>3</sup> Dept. of Sociology, SUNY, Stony Brook, New York

where:  $a_0 \equiv \min_{1 \leq i \leq n}(a_i)$ ;  $b_0 \equiv \max_{1 \leq i \leq n}(b_i)$ . The interval  $(b_0 - a_0)$  represents the full range of opinions the  $n$  respondents have about the possible true values of their answers to the question, from the smallest lower bound to the largest upper bound. In eqn. (2.3),  $k_1$  and  $k_2$  denote pre-assigned multiples of standard deviations that correspond to how the bounds should be interpreted in terms of standard deviations from the mean. For example, for normally distributed data it is sometimes assumed that such lower and upper bounds can be associated with 2 standard deviations below, and above, the mean, respectively. With this interpretation, we could take  $k_1 = k_2 = 4$  to represent the length of the interval between the largest and smallest values the true value of the answer to the recall question might be for respondent  $i$ . If desired, we might take  $k_1 = k_2 = k$ , and then we would make a choice among reasonable values, such as:  $k=2,4,5,6,7,8$ , and study how the estimate of the population variance varies with  $k$ .

The new estimating procedure used here substitutes for  $(b_0 - a_0)$ :

$$4\tau \doteq (b_0 - a_0) + \frac{2}{\sqrt{n}}(s_a + s_b)$$

to form what we will refer to as the “extended range estimator”, and

$$4\tau \doteq (\bar{b} - \bar{a}) + \frac{2}{\sqrt{n}}(s_a + s_b)$$

to form what we will refer to as the “extended average estimator” (see Appendix of Chu, Press and Tanur, 2004)).

Here  $\bar{b}$  and  $\bar{a}$  are the means of the upper bounds and of the lower bounds given by the respondents, respectively; and  $s_a$  and  $s_b$  are the sample standard deviations of the lower bounds and upper bounds, respectively.

### 3. The Classroom Survey

#### 3.1 Confidence and Question Wording

Because point estimates of respondents who give short intervals are weighted more heavily in the Bayesian RGI estimator than are point estimates of respondents who give longer intervals (see eqn. (2.3)), it is advantageous to encourage respondents who are more accurate to give shorter intervals and respondents who are less accurate to give longer ones. We know from earlier uses of the RGI procedure that, among respondents who do not receive any special guidance about the length of their intervals, there is a substantial correlation between interval length and accuracy (with less accurate respondents giving longer

intervals; Press and Tanur, 2003). There is also a correlation between confidence and interval length (with less confident respondents giving longer intervals; Press and Tanur, 2002). Our aim is to increase the correlation between accuracy and interval length, by working through respondents’ confidence and cueing them appropriately. We have developed a questioning protocol that aims to increase that correlation.

First, the respondent is requested to give his/her “best guess” about the quantity being investigated, and then is asked how confident s/he is of that answer on a scale from 0 (least confident) to 10 (most confident). Figure 1 shows the form of this confidence scale for a question used in our experiment involving recall of the respondent’s grade on a classroom exam. Respondents who represent themselves as highly confident (confidence ratings 7.5 or 10) are directed to a question that encourages them to give a narrow bounding interval. Less confident respondents (confidence ratings of 5 or less) are directed to a question encouraging a wide bounding interval.

The design for this experimental application of the new protocol used three versions of the bounding questions (and each version was completed by a different group of respondents). Version 1, which we refer to as “unanchored,” simply asks the respondent to give a narrow, or a broad, interval; this version was administered to Group 1. See Figure 2 for the wording of Version 1 for the question about the classroom exam. Version 2, administered to Group 2, which we refer to as the “narrow-wide anchored condition,” not only encourages respondents to give narrow or wide intervals, but it also tells them that the narrow interval should be no more than a specified width and that the wide interval should be at least a specified width. See Figure 3 for the wording of Version 2 as used for the question about the classroom exam. Version 3 (referred to as the “wide-wide anchored condition” and administered to Group 3) is the same as Version 2, except that the suggested width of the wide interval was considerably wider (see Figure 4).

#### 3.2 Ratings of Memory

Respondents were asked to evaluate their memory on the scale shown in Figure 5. If respondents are good judges of their own memory, then perhaps rather than asking confidence questions for each survey item we can use a procedure that simply classifies respondents into “good memory” and “poor memory” groups and encourage good memory respondents to give short intervals and poor memory respondents to give long ones. Such a procedure would impose considerably less respondent burden than does asking for confidence for each question.

#### 3.3 The Survey

In the spring of 2003 we ran a small experimental record-check survey in an undergraduate, lower division,

statistics class at the University of California at Riverside. In a randomized design three groups of students were each given a different version of the questionnaire and students were asked to recall their midterm exam score, their score on their second homework assignment, and the amount they had paid at the beginning of the quarter as a registration fee. Because there were three versions of the questionnaire, and because participation was voluntary, sample sizes in the three groups were rather small, but sufficiently large for us to derive some preliminary results. (A similar experiment from a larger class was run several months later in the fall of 2003 – results will be available shortly.) With the students’ permission we were able to compare their reported grades with those recorded in the professor’s grade book; the registration fee was fixed by the university for all full-time students at \$239.

**3.4 Preliminary Findings**

(1) Our first finding was that the manipulation worked. Table 1 shows that the mean length of intervals generated by respondents who were asked to give a wide interval were always wider than those from respondents asked to give a narrow interval. In every case in which a t-test was possible (that is, whenever both group sizes were greater than 1) this finding reached at least marginal statistical significance, in spite of the small sample sizes.

**Table 1: Average Lengths of Intervals for Wide and Narrow Anchors**

	Narrow Interval	n	Wide Interval	n	p**
<b>Midterm</b>					
<i>Unanchored</i>	6.7	19	14.6	8	.006
<i>Narrow/wide</i>	9.0	18	<b>16.7</b>	3	.069
<i>Wide/wide</i>	8.5	19	<b>25.0</b>	3	.026
<b>RegFee</b>					
<i>Unanchored</i>	165.00	5	1280.40	19	.004
<i>Narrow/wide</i>	0.00	1	763.90	13	*
<i>Wide/wide</i>	20.00	1	608.33	18	*
<b>Homework</b>					
<i>Unanchored</i>	2.9	12	6.6	10	.033
<i>Narrow/wide</i>	2.8	10	<b>6.4</b>	10	.030
<i>Wide/wide</i>	2.7	9	<b>7.6</b>	8	.020

\*Narrow interval group n=1; no test of significance possible.

\*\*p-values are included for those readers for whom p-values are sensible. We recognize this will not be true for all readers.

(2) For both the homework question and the midterm question, the mean of the wide intervals for respondents given the wide-wide anchor was longer than the mean of the intervals for respondents given the narrow-wide anchor. This relationship did not hold for the question about registration fee, for which most respondents seem to have been very much lacking in knowledge about how much the actual fee was (which resulted in low confidence).

(3) It is interesting to note that there seems to be a relationship between respondents’ confidence and the salience of the question. A large majority of respondents were quite confident that they remembered their midterm grade correctly, a large majority lacked such confidence for the registration fee, and for the homework grade the respondents split about half and half.

(4) Table 2 further checks the manipulation, asking whether there was indeed a correlation between respondents’ confidence in the accuracy of their recall and their actual accuracy in reporting their usage quantities. The actual accuracy is measured as the absolute value of the differences between the reported usage quantity and recorded “truth.” Large values of these differences represent *inaccuracy*, so if there is a relationship between accuracy and confidence we would expect negative correlations, as indeed we see in Table 2. (We might have labeled the absolute value of the difference between truth and the usage quantity as “inaccuracy,” but calling it accuracy simplifies our discussion as long as the reader keeps in mind how the variable is measured and that we hope for negative correlations between it and interval length.)

**Table 2: Correlations between Confidence and “Accuracy” (usage-truth)\* -- All Respondents**

	r	n	p (1-tail)**
<b>Midterm</b>			
<i>Unanchored</i>	-.110	27	.305
<i>Narrow/wide</i>	-.518	21	<b>.008</b>
<i>Wide/wide</i>	-.443	23	<b>.017</b>
<b>RegFee</b>			
<i>Unanchored</i>	-.110	27	.291
<i>Narrow/wide</i>	-.049	14	.434
<i>Wide/wide</i>	-.375	21	<b>.047</b>
<b>Homework</b>			
<i>Unanchored</i>	-.385	20	<b>.047</b>
<i>Narrow/wide</i>	-.355	20	.062
<i>Wide/wide</i>	-.184	17	.289

\*We expect high levels of confidence to go with greater accuracy (small error, the absolute difference between the usage quantity offered by the respondent and truth); thus increasing confidence should go with decreasing error. Hence, if our hypothesis is correct, the correlations should be negative. They are.

\*\*p-values are included for those readers for whom p-values are sensible. We recognize this will not be true for all readers.

While these correlations are hardly enormous, we see that there is a relationship between accuracy and confidence in all cases. Each group of respondents contributed at least one low correlation – the unanchored group showing a low correlation for both the midterm question and the registration fee question, the narrow-wide anchor group showing a low correlation for the registration fee question, and the wide-wide anchor group showing a low correlation for the homework question. Hence we cannot

attribute the low correlations either to a particular group of respondents or to the difficulty of a particular question. We do suspect, however, that the correlations coming from the registration fee question are influenced by the fact that very few respondents were confident about their answers to this question – see the *n*'s in Table 1. We speculate that respondents knew more about the total fees they paid than about the specific registration fee, about which they knew almost nothing, so they guessed wildly. There is also some evidence from student comments that if their parents paid their fees or if they received financial aid, they have little knowledge about the amount of any fees.

(5) Table 3 examines the relationship between interval length and accuracy (measured as explained above, that is, as “inaccuracy”). If, as we hope, respondents who are less

**Table 3: Correlations between Interval Length and “Accuracy” (usage-truth)**

All Respondents with Useable Data			
	r	n	p (1-tail)*
<b>Midterm</b>			
<i>Unanchored</i>	.311	25	.065
<i>Narrow/wide</i>	.149	19	.272
<i>Wide/wide</i>	.069	22	.381
<b>RegFee</b>			
<i>Unanchored</i>	.395	24	.028
<i>Narrow/wide</i>	.671	12	.008
<i>Wide/wide</i>	.286	19	.128
<b>Homework</b>			
<i>Unanchored</i>	.011	20	.482
<i>Narrow/wide</i>	.273	18	.138
<i>Wide/wide</i>	.320	17	.105
Obedient Respondents Only			
	r	n	p (1-tail)*
<b>Midterm</b>			
<i>Unanchored</i>	.311	25	.065
<i>Narrow/wide</i>	.624	11	.020
<i>Wide/wide</i>	.349	14	.110
<b>RegFee</b>			
<i>Unanchored</i>	.395	24	.028
<i>Narrow/wide</i>	.638	9	.032
<i>Wide/wide</i>	.247	16	.178
<b>Homework</b>			
<i>Unanchored</i>	.011	20	.482
<i>Narrow/wide</i>	.274	12	.194
<i>Wide/wide</i>	.305	11	.181

\*p-values are included for those readers for whom p-values are sensible. We recognize this will not be true for all readers.

accurate give longer intervals, we would expect positive correlations. The correlations in Table 3 are all positive. There are two panels for Table 3 – the top panel includes all respondents who gave the 4 pieces of data requested – confidence rating, usage quantity, lower bound, and upper

bound – and whose usage quantity properly fell within the bounds. The bottom panel includes only what we call “obedient” respondents – those who followed the directions given in the anchoring instructions and gave a wide interval at least as wide as prescribed, or a narrow interval at least as narrow as prescribed. Two comments are in order for this table. First, we seem to have been successful in increasing the correlation between interval length and accuracy from the level obtained from respondents without any special instructions regarding interval length. Most of these correlations are larger than those reported in Press and Tanur (2002), where the median of 18 correlation coefficients (for 18 items) was 0.13; 6 of the 18 were negative; and the only correlations exceeding 0.40 were those relating to the frequencies of behaviors (a case where those who really had no occurrences of the requested behavior could easily remember that they had none, and could be quite confident about their recall). Second, limiting ourselves to obedient respondents seems to be useful. (Note that, because the unanchored group was not given a suggested length of interval, the “obedient” vs. “disobedient” distinction does not pertain to this group and the data for this group in the lower panel of Table 3 simply repeat the data in the upper panel.) When we omit those respondents who were “disobedient” we find that the correlations never decrease substantially and two correlations that were originally small increase considerably.

(6) Table 4 shows the results of the estimation process

**Table 4: Point Estimate Results Using Vague Prior and Extended Average and Extended Range Procedures Obedient Respondents Only**

	n	Average Truth	x-bar	Extended Average	Extended Range
<b>Midterm</b>					
<i>Unanchored</i>	25	83.88	83.04	<b>83.79</b>	83.17
<i>Narrow/wide</i>	11	81.36	79.64	<b>79.94</b>	79.70
<i>Wide/wide</i>	14	86.57	86.71	<b>86.68</b>	86.70
<b>RegFee</b>					
<i>Unanchored</i>	24	239.00	1366.46	<b>1202.25</b>	1328.28
<i>Narrow/wide</i>	9	239.00	1090.78	<b>974.77</b>	1047.50
<i>Wide/wide</i>	16	239.00	1190.88	<b>1122.65</b>	1176.97
<b>Homework</b>					
<i>Unanchored</i>	20	16.91	18.00	18.06	<b>17.99</b>
<i>Narrow/wide</i>	12	16.72	<b>17.92</b>	18.04	18.01
<i>Wide/wide</i>	11	16.69	<b>18.63</b>	18.81	18.65

using the Bayesian estimators for the obedient respondents only<sup>4</sup>. In Table 4 the estimator that is closest to the truth is

<sup>4</sup> Minitab macros that simplify computing the numerical results required for RGI Bayesian estimates were written by Dr. Diane Miller. For information about these macros contact her at [diane.m.miller@ngc.com](mailto:diane.m.miller@ngc.com).

presented in boldface. We see that although all the estimates were very close to one another, the extended average estimator is closest to truth for the midterm grades and for the registration fee. The sample mean seems to work best for the homework question, except for the unanchored condition where the extended range estimate is a tiny bit closer to truth. Note that the median correlation between accuracy and interval length for the midterm question is 0.349; for the registration fee question its is 0.395; but for the homework question it is only 0.274. Hence we should not be surprised that the Bayesian estimator works better for the midterm and registration fee questions than it does for the homework question. The findings for the extended average estimator are also shown graphically in Figure 6. Here what is graphed is the absolute value of the difference between the estimated value and average truth. G1 refers to the unanchored condition, G2 to the narrow-wide anchor condition, and G3 refers to the wide-wide anchor condition.

(7) Table 5 relates respondents' ratings of their memory to

**Table 5: Average Confidence Scores by Respondents' Memory Rating**

	All Respondents				
	Good Memory	n	Poor Memory	n	p-value (1-tail)*
<b>Midterm</b>					
<i>Unanchored</i>	<b>8.75</b>	6	<b>7.05</b>	22	<b>.028</b>
<i>Narrow/wide</i>	<b>9.58</b>	6	<b>7.83</b>	15	<b>.017</b>
<i>Wide/wide</i>	8.21	7	9.22	16	
<b>RegFee</b>					
<i>Unanchored</i>	<b>4.29</b>	7	<b>3.50</b>	20	.257
<i>Narrow/wide</i>	<b>4.00</b>	5	<b>2.78</b>	9	.266
<i>Wide/wide</i>	<b>3.33</b>	6	<b>1.17</b>	15	.074
<b>Homework</b>					
<i>Unanchored</i>	<b>7.50</b>	6	<b>6.84</b>	19	.276
<i>Narrow/wide</i>	4.58	6	6.25	14	
<i>Wide/wide</i>	4.50	5	6.35	13	

\*p-values are included for those readers for whom p-values are sensible. We recognize this will not be true for all readers.

their confidence as rated on the confidence scales for the several questions. We see those respondents who rated their memory good (that is, those who claimed never or seldom to be mistaken when they were sure they knew something) in many cases give higher average confidence ratings than those respondents who say their memory is less good (those who claimed sometimes or frequently to be mistaken when they were sure they knew something). This finding (shown in boldface) holds true for all questions for the unanchored condition, for the midterm and the registration fee questions for the narrow-wide anchored condition, and only for the registration fee for

the wide-wide anchored condition. Confidence ratings were higher on average for the "good memory" group than in the "poor memory" group in 6 of the 9 possible comparisons. Two of these 6 "wins" reached statistical significance at conventional levels and another one was marginally significant.

(8) Table 6 shows the accuracy achieved by respondents at different levels of self-rated memory. Note that "accuracy" is again measured by the absolute value of the difference between a respondent's reported usage quantity and truth. Thus large values represent inaccuracy, and smaller values are more accurate. We see that on the average respondents who rated themselves to have good memories were closer to the truth in 5 of the 9 possible comparisons (shown in boldface). None of these "wins" reached statistical significance at conventional levels.

**Table 6: Accuracy (|usage-truth|) by Respondents' Memory Rating**

	All Respondents				
	Good Memory	n	Poor Memory	n	p-value (1-tail)*
<b>Midterm</b>					
<i>Unanchored</i>	<b>2.33</b>	6	<b>5.10</b>	21	.134
<i>Narrow/wide</i>	7.33	6	6.00	16	
<i>Wide/wide</i>	<b>0.57</b>	7	<b>0.69</b>	16	.424
<b>RegFee</b>					
<i>Unanchored</i>	<b>961</b>	7	<b>1245</b>	19	.253
<i>Narrow/wide</i>	<b>685</b>	5	<b>1519</b>	10	.112
<i>Wide/wide</i>	1405	6	729	15	
<b>Homework</b>					
<i>Unanchored</i>	<b>1.50</b>	5	<b>1.60</b>	15	.453
<i>Narrow/wide</i>	3.00	6	2.13	15	
<i>Wide/wide</i>	2.90	5	2.26	13	

\*p-values are included for those readers for whom p-values are sensible. We recognize this will not be true for all readers.

#### 4. Discussion

We can, we believe, claim some success with these new directions. We seem to have affirmed the need to ask confidence questions separately for each usage quantity, for while respondents' estimates of their own memory seem to be fairly good predictors of that confidence, those memory estimates do not relate nearly as well to actual accuracy as do the confidence ratings themselves. We had hoped to minimize respondent burden by asking a single memory question, but it seems the burden of asking separate confidence questions is a necessary one.

We have established that respondents directed to give wide intervals give wider ones on the average than do respondents directed to give narrower ones. There does not seem to be much effect of the length of the anchoring interval, but we await the results of a considerably larger sample size experiment to see if that lack of effect is real. We were able to improve the correlation between accuracy

and interval length through the use of the confidence scale. We would like to increase that correlation even more, as it is the sine qua non for the successful application of the RGI protocol. We shall be trying other ways of asking for respondents' confidence, but of course we will be limited by any imperfections in respondents' understanding of their own accuracy. Respondents who are honestly confident but nevertheless inaccurate, and respondents who honestly lack confidence but are nevertheless accurate, will continue to haunt us. Even in this small test, however, we do see that our manipulation of respondents' interval length, based on their confidence, results in the RGI Bayesian estimator showing less bias than the sample mean in a majority of the cases we examined.

**References**

Chu, Liping, S. James Press, and Judith M. Tanur (2004) "Confidence Elicitation and Anchoring in the Respondent-Generated Intervals (RGI) Protocol" *Journal of Modern Applied Statistical Methods* Vol 3 No 2: (November) In press.

Press, S. James (2004). "Respondent-Generated Intervals (RGI) for Recall in Sample Surveys", *Journal of Modern Applied Statistical Methods*, Vol.3 No.1:104-116.

Press, S., James and Judith M. Tanur (2002) "Cognitive and Econometric Aspects of Responses to Surveys as Decision-Making," Technical Report #271, Department of Statistics, University of California at Riverside, Riverside, CA 92521-0138.

----- and -----, (2003) "The Relationship between Accuracy and Interval Length in the Respondent-Generated Interval Protocol", *Proceedings of the Survey Research Methods Section of the American Statistical Association*. Washington, D.C.: American Statistical Association.

----- and -----(2004). "Relating Respondent-Generated Intervals Questionnaire Design to Survey Accuracy and Response Rate", *Journal of Official Statistics*, Vol.20 No. 2, pp. 265-288 (June), Special Issue on Questionnaire Development, Evaluation, and Testing Methods.

**Figure 1—New Form for RGI Protocol**

- 1) What is your best guess as to what your score was on your first exam in this class? (Please don't answer if you've missed the first exam).\_\_\_\_\_.
- 2) How confident are you about your answer to Question 1? Please answer on the following confidence scale. (Place a check in the first column next to the answer you prefer.)

Confidence Scale

Place a check somewhere this column	Numerical Score	Interpretation of confidence rating	Which question should I answer next?
<input type="checkbox"/>	0	I have absolutely no idea what my exam score was	Go to Question 3b
<input type="checkbox"/>	2.5	I am uncertain what my exam score was	Go to Question 3b
<input type="checkbox"/>	5.0	I might be right and I might be wrong about what my exam score was	Go to Question 3b
<input type="checkbox"/>	7.5	I think that I know what my exam score was	Go to Question 3a
<input type="checkbox"/>	10.0	I am absolutely certain what my exam score was	Go to Question 3a

**Figure 2 -- Unanchored Bounds Condition**

3a) If your answer to Question 2 is 7.5 or more, please give the **smallest possible interval** in which you believe that the exam score is included. Please fill in:

The smallest my exam score could have been is % \_\_\_\_\_,

The largest my exam score could have been is % \_\_\_\_\_.

**NOW GO TO QUESTION 4.**

3b) If your answer to Question 2 is 5 or less, give a **sufficiently wide interval** so that the interval will most likely include the actual exam score

Please fill in:

The smallest my exam score could have been is % \_\_\_\_\_,

The largest my exam score could have been is % \_\_\_\_\_.

**Figure 3 --Narrow Wide Anchor Bounds Question**

3a) If your answer to Question 2 is 7.5 or more, please give the **smallest possible interval** in which you believe that the exam score is included. For example, if your best guess about your exam score is 75%, give a narrow interval of no more than 4 points in length, such as (73%,76%).

Please fill in:

The smallest my exam score could have been is % \_\_\_\_\_,

The largest my exam score could have been is % \_\_\_\_\_.

**NOW GO TO QUESTION 4.**

3b) If your answer to Question 2 is 5 or less, give a **sufficiently wide interval** so that the interval will most likely include the actual exam score. For example, if your best guess about your exam score is 75%, give a wide interval of at least 20 points in length, such as (65%, 85%).

Please fill in:

The smallest my exam score could have been is % \_\_\_\_\_,

The largest my exam score could have been is % \_\_\_\_\_.

**Figure 4 – Wide Wide Anchor Bounds Question**

3a) If your answer to Question 2 is 7.5 or more, please give the **smallest possible interval** in which you believe that the exam score is included. For example, if your best guess about your exam score is 75%, give a narrow interval of no more than 4 points in length, such as (73%,76%).

Please fill in:

The smallest my exam score could have been is \_\_\_\_\_,

The largest my exam score could have been is \_\_\_\_\_.

**NOW GO TO QUESTION 4.**

3b) If your answer to Question 2 is 5 or less, give a **sufficiently wide interval** so that the interval will most likely include the actual exam score. For example, if your best guess about your exam score is 75%, give a wide interval of at least 30 points in length, such as (60%, 90%).

Please fill in:

The smallest my exam score could have been is \_\_\_\_\_,

The largest my exam score could have been is \_\_\_\_\_.

**Figure 5 – Memory Evaluation Scale**

**Does it ever happen that when you are sure you know something, it turns out that you are mistaken? Please check one:**

Never \_\_\_\_\_

} Good Memory

Seldom \_\_\_\_\_

Sometimes \_\_\_\_\_

} Poor Memory

Frequently \_\_\_\_\_

Figure 6: Bias (estimate - truth) for Extended Average Bayesian Estimate Compared with Bias of Sample Mean

