Political Analysis (2016) 24:226–242
doi:10.1093/pan/mpw010

Guessing and Forgetting: A Latent Class Model for Measuring Learning

M. Ken Cor
Faculty of Pharmacy and Pharmaceutical Sciences, 3-209 Edmonton Clinic Health Academy, University of Alberta, Edmonton, CA T5M1B3
e-mail: mcor@ualberta.ca (corresponding author)

Gaurav Sood
Independent Researcher
e-mail: gsood07@gmail.com

Edited by: R. Michael Alvarez

Guessing on closed-ended knowledge items is common. Under likely-to-hold assumptions, in the presence of guessing, the most common estimator of learning, difference between pre- and postprocess scores, is negatively biased. To account for guessing-related error, we develop a latent class model of how people respond to knowledge questions and identify the model with the mild assumption that people do not lose knowledge over short periods of time. A Monte Carlo simulation over a broad range of informative processes and knowledge items shows that the simple difference score is negatively biased and the method we develop here is unbiased. To demonstrate its use, we apply our model to data from Deliberative Polls. We find that estimates of learning, once adjusted for guessing, are about 13% higher. Adjusting for guessing also eliminates the gender gap in learning, and halves the pre-deliberation gender gap on political knowledge.

1 Introduction

A nontrivial number of participants appear to *lose* knowledge over informative processes.[1] For instance, nearly 9% of the participants of Deliberative Polls,[2] which provide vetted briefing materials and include moderated discussion, appear to go from knowing a piece of information to not knowing it over the course of the poll. Could such informative processes somehow be backfiring, perhaps by replacing information with misinformation? A far likelier explanation, we suspect, lies in the existence of guessing—lucky guessing on the pretreatment wave, and unlucky guessing on the posttreatment wave.

More generally, guessing inflates the scores—when measured as proportion correct—of the uninformed. The less informed a respondent, the greater the positive bias in their score, on

Author's note: We thank Ed Haertel for his advice and encouragement, Robert Luskin for improving some of the language, and Pablo Barberá, Ying Cui, Kabir Khanna, Marc Meredith, and Pete Mohanty for useful comments. Replication files for this study are available on the Harvard Dataverse at http://dx.doi.org/10.7910/DVN/HZHVCU. See Cor and Sood (2016). Supplementary materials for this article are available on the *Political Analysis* Web site.

[1]Our evidence for the claim comes from a broad set of cases: (1) Over an experiment that gave students opportunity to learn basic algebra, roughly 10% of the students appear to go from knowing the answer to not knowing it (Cor 2012); (2) Between 6% to 10% of respondents appear to "forget" what they know, sometimes across closely spaced waves overlapping periods when correct pertinent information is being widely disseminated (Lenz 2009); and (3) Nearly 9% of Deliberative Poll participants suffer the same fate over the course of a poll.

[2]Deliberative Polls bring together a random sample to deliberate about an issue or set of issues. Participants are provided with balanced briefing materials in advance of deliberations and an opportunity to quiz experts (for greater detail, see Fishkin 2009; Luskin et al. 2002).

average.[3,4,5] Over an informative process, where people typically know more at the end of the process than they do before it, guessing induced bias in the preprocess scores is typically greater than the bias in the postprocess scores. Ergo, raw estimates of knowledge gain (post minus pre) are downwardly biased. For a few respondents guessing induced inflation in the preprocess scores (average over multiple items) can even exceed the joint total of learning, and lucky guessing on the postprocess instrument.

So, what can we do to correct for guessing? One reasonable place to start is "the standard correction," which assumes that the probability of getting an item correct when guessing is squarely a matter of chance, estimates the true number of guesses based on the number of incorrect responses, and docks appropriately from the observed score. The technique, however, fails to account for differential opportunities to narrow alternatives across items ($\sim$ eliminate nondiagnostic red herrings) before guessing (Nunnally 1967; Lord 1975). Many variations of the standard correction have been proposed to account for nonuniform chance of lucky guessing across items (see, for instance, Hansen et al. 1975). But all of the methods rely on strong assumptions about the data-generating process.

A more recent strategy has been to model responses using a three-parameter Item Response Theory (IRT) model (see Lord et al. 1968). The model, rather than assume that the ignorant guess at random, estimates an item-level conditional probability of getting an item correct given that the respondent lacks "ability." In doing so, the model accounts for item idiosyncrasies such as the presence of red herrings in response options that can lead the ignorant to guess correctly at rates very different from chance. IRT models, however, assume that "ability" is unidimensional, contrary to what some evidence in political science suggests (see, for instance, Iyengar 1990; Stolle and Gidengil 2010). And while multidimensional IRT models can be estimated, such models generally need lots of items and respondents for stable estimates; most surveys on politics carry but a few political knowledge questions.

More importantly, perhaps, all these measurement models typically estimate (and correct for) guessing using data from a single wave (which covers a vast majority of the cases). However, in places where we have data from two or more waves, better methods for accounting for guessing may exist. In this article, we develop a new method for accounting for guessing for such data; we estimate item-level guessing parameters by exploiting informative within-item transitions across waves. We conduct a Monte Carlo simulation over a broad range of informative processes to assess how our method compares to the conventional estimator. We find that the conventional estimator is negatively biased and estimates have a large mean squared error (MSE), and that our method yields unbiased estimates with a low MSE. We next apply our model to data from the Deliberative Polls. We find that estimates of the proportion of people who learn various specific pieces of information, once adjusted for guessing, are considerably higher. Relatedly, we find that accounting for guessing removes the gender gap in learning and halves the pretreatment gender gap on political knowledge.

2 Traditional Pre- and Post-Estimates of Learning

For reasons to do with their popularity, we focus on closed-ended questions. To better reflect the nature of the data we analyze later, and given that items with a "don't know" option can easily be translated to "forced choice" data—one need only treat "don't know" as "don't know" (see Luskin and Bullock 2011; Luskin and Sood 2012; Sturgis et al. 2008)—we focus on closed-ended items that carry a "don't know" option (Table 1). (For completeness, we also describe a model of responses on items that do not offer a "don't know" option in Online Appendix A.) We further assume that there

[3]Guessing by the ignorant contributes solely to error, while guessing by those with related knowledge, under nominal assumptions, increases proportion correct proportional to the amount of related knowledge. In our discussion here, when we refer to guessing, we mean guessing by those who do not know.
[4]Note that when there is an opportunity to narrow options without the use of substantive knowledge, guessing by the ignorant can net correct answers at better than chance levels. For instance, asked to identify the person in a photo, respondents can take the gender of the person in the photo into account when selecting between options.
[5]The same point is stated as an assumption in Sniderman et al. (1993, 105).

is no missing data (though no changes to the logic need to be made if the data are missing completely at random).

Let "0" indicate an incorrect answer, "1" a correct answer, and "c" a "don't know" response. Additionally, let i denote a response to a knowledge item (0, 1, or c), and let x_{ii} denote the proportion of people who display the subscripted response pattern on a particular item across the two administrations. Responses to an item over two waves can be described as follows:

Table 1 Relationship between manifest responses across two waves

		Post		
		0	*1*	*c*
Pre	0	x_{00}	x_{01}	x_{0c}
	1	x_{10}	x_{11}	x_{1c}
	c	x_{c0}	x_{c1}	x_{cc}

The conventional estimator of knowledge gain on any single item is the difference between the proportion of people who got the item right on the posttest $(x_{01} + x_{11} + x_{c1})$ and the proportion of people who got the item right on the pretest $(x_{10} + x_{11} + x_{1c})$. Simplifying, the conventional estimate of learning for a single item is the difference between the proportion of people moving from an incorrect or "don't know" response to a correct response, and the proportion of people moving from a correct response to either an incorrect or a "don't know" response $(x_{01} + x_{c1} - x_{10} - x_{1c})$.

The "raw" (conventional) estimator, however, may be a biased estimator of the quantity of interest—proportion of people who actually learned the item between the two waves. Inference about the quantity of interest can be compromised in various ways. For instance, if the observed 0–1 transition reflects a transition from an unlucky guess to a lucky one, counting such transitions will lead us to overestimate the amount of learning. More generally, if a greater share of correct answers comes from lucky guessing by the ignorant on the pretest than on the posttest, the raw estimate will be negatively biased. In processes designed to impart information, where respondents know as much or less before the intervention as after, the proportion of correct answers attributable to lucky guessing in the pretest is liable to be either equal to or larger than in the posttest. Thus, the conventional estimator is a biased estimator of learning in informative processes where people actually learn.

Our goal is to develop an unbiased estimator of learning, the number of true transitions from ignorance to knowledge. To do that, we begin by describing how the latent states of knowledge and ignorance are connected to manifest responses.

3 Manifest Responses and Latent States

A person either knows a piece of information or does not. This is not to say that someone who does not know a specific piece of information does not have any related cognitions about it. For instance, an individual may not know that Barack Obama is the president of the United States but may know that he is not the president of France. But knowing that Mr. Obama is not the president of France is not the same as knowing who the president of the United States is. We take the knowledge of a specific piece of information as the quantity of interest. (We come back to measuring related cognitions later in this section.) Assume further that people do not know the wrong thing. This leaves knowledge and ignorance as the only possible latent states of cognition about a particular piece of information.

Let u indicate the manifest response. We expect someone who knows the specific piece of information to be able to recognize it among the few response options and mark the right answer. Hence, we assume that the conditional probability of a respondent marking the right answer given that the respondent knows the answer is unity, or $p(u = 1|K) = 1$, where K denotes that the respondent actually knows the item (the latent state is knowledge). This also means that the

Table 2 Relationship between manifest responses and latent classes

		Latent classes		
		K	G	C
Observed	$u=0$	0	$(1-\gamma)$	0
	$u=1$	1	γ	0
	$u=c$	0	0	1

respondent's probability of getting the question wrong, or marking "don't know" given that she knows the answer, is zero: $p(u=0|K)=p(u=c|K)=0$.[6] Given that we have only two latent states, previous results also mean that only the uninformed get an item incorrect or offer a "don't know" response. The uninformed, however, can also guess luckily. So, we have a world where some of the uninformed confess to their ignorance, marking "don't know," and some guess. It is easier to think of the uninformed who guess and who confess as two kinds of people, say "Guessers," denoted by G, $p(u=c|G)=0$, and "Confessors," denoted by C, $p(u=c|C)=1$. When the Guessers do guess, γ of their guesses are lucky, $p(u=1|G)=\gamma$. We assume γ to be a fixed item-level parameter, constant across time.[7] Table 2 summarizes the information more compactly.

The relationships between manifest responses and the set of latent classes described in Table 2 describe one question at a single point in time. The probability of different response patterns across two waves of measurement, given a set of underlying latent class transitions, can be found by taking the Kronecker Product of the pretest and posttest classification matrices:

$$\begin{pmatrix} & G & K & C \\ u=0 & 1-\gamma & 0 & 0 \\ u=1 & \gamma & 1 & 0 \\ u=c & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} & G & K & C \\ u=0 & 1-\gamma & 0 & 0 \\ u=1 & \gamma & 1 & 0 \\ u=c & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} & GG & GK & GC & KG & KK & KC & CG & CK & CC \\ u=00 & (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=01 & (1-\gamma)\gamma & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=0c & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\ u=10 & (1-\gamma)\gamma & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ u=11 & \gamma^2 & \gamma & 0 & \gamma & 1 & 0 & 0 & 0 & 0 \\ u=1c & 0 & 0 & \gamma & 0 & 0 & 1 & 0 & 0 & 0 \\ u=c0 & 0 & 0 & 0 & 0 & 0 & 0 & (1-\gamma) & 0 & 0 \\ u=c1 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma & 1 & 0 \\ u=cc & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

[6]It is reasonable to worry that when a "don't know" option is offered, some of the respondents who know the answer will still mark "don't know." However, research suggests that there is very little "hidden knowledge" (about 2–3%) behind "don't know" responses on closed-ended items (Sanchez and Morchio 1992; Sturgis et al. 2008; Luskin and Bullock 2011; Luskin and Sood 2012). Claims to the contrary (Prior and Lupia 2008) have been shown to be founded on a misreading of the data—interpreting results in terms of percentage increases rather than increases in percentages (Luskin and Bullock 2011). In Online Appendix B, we describe a latent class model that assumes the probability of a person marking "don't know" when holding the relevant piece of information as .03.

[7]In order for γ to vary over an informative process, the process must affect participants' ability to use nondiagnostic cues to answer closed-ended questions correctly. Leaving aside 'informative' processes focused on teaching test-wiseness, we do not expect the ability to answer correctly without the aid of substantive knowledge to change over the course of an informative process.

The resulting 9×9 latent class transition matrix describes the conditional probabilities of specific response patterns given different latent class transitions. For example, given a latent transition from guess to guess (GG), the response patterns $0c$, $1c$, $c0$, $c1$, and cc are impossible. The probabilities of 00, 01, 10, and 11 response patterns given an underlying GG transition are represented by the remaining entries in the first column of the 9×9 matrix.

Next, the latent class transition matrix is multiplied by a vector of unknown latent class transition parameters to define the system of equations that describe the multinomial distribution of observed response patterns. There are nine latent class transitions that could hypothetically occur: GG, GK, GC, KG, KK, KC, CG, CK, and CC.

To identify the item-level conditional probability of getting an item correct if the respondent chooses to guess, we make the same assumptions as above—that the respondent who knows the right answer does not forget it over the course of the informative process, and that they do not become misinformed. Under these assumptions, the proportion of people moving from knowing to guessing (KG) and from confessing to ignorance (KC) is zero.

Multiplying the 9×9 matrix from equation (1) by the latent class transition vector, λ, and setting λ_{KG} and λ_{KC} to zero, we get a system of equations, relating latent parameters to expected response pattern proportions, π:

$$\begin{array}{rc} & \begin{array}{ccccccccc} GG & GK & GC & KG & KK & KC & CG & CK & CC \end{array} \\ \begin{array}{r} u=00 \\ u=01 \\ u=0c \\ u=10 \\ u=11 \\ u=1c \\ u=c0 \\ u=c1 \\ u=cc \end{array} & \begin{pmatrix} (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1-\gamma)\gamma & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\ (1-\gamma)\gamma & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ \gamma^2 & \gamma & 0 & \gamma & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (1-\gamma) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{pmatrix} \lambda_{GG} \\ \lambda_{GK} \\ \lambda_{GC} \\ \lambda_{KG}=0 \\ \lambda_{KK} \\ \lambda_{KC}=0 \\ \lambda_{CG} \\ \lambda_{CK} \\ \lambda_{CC} \end{pmatrix} \tag{2}$$

$$= \begin{pmatrix} (1-\gamma)^2\lambda_{GG} \\ (1-\gamma)\gamma\lambda_{GG} + (1-\gamma)\lambda_{GK} \\ (1-\gamma)\lambda_{GC} \\ \gamma(1-\gamma)\lambda_{GG} \\ \gamma^2\lambda_{GG} + \gamma\lambda_{GK} + \lambda_{KK} \\ \gamma\lambda_{GC} \\ (1-\gamma)\lambda_{CG} \\ \gamma\lambda_{CG} + \lambda_{CK} \\ \lambda_{CC} \end{pmatrix} = \begin{pmatrix} \pi_{00} \\ \pi_{01} \\ \pi_{0c} \\ \pi_{10} \\ \pi_{11} \\ \pi_{1c} \\ \pi_{c0} \\ \pi_{c1} \\ \pi_{cc} \end{pmatrix}$$

Based on the model, the proportion of respondents who learned an item is simply the sum of the estimates of the two transition parameters that track movement from ignorance to knowledge, λ_{GK} and λ_{CK}.[8]

[8]Rather than estimating γ using the above model, one can estimate γ under other assumptions, or using alternate methods. The value of γ can be plugged into the equation and then used to estimate the other unknown parameters. For instance, one can use estimate of γ from the standard guessing correction, inverse of the number of options, or the 3-PL IRT model. Note, however, that data with "don't know" do not lend themselves to correct estimation of γ in the

3.1 *Related Knowledge*

As we note above, a person either knows a particular piece of information, or he does not. A person, however, may have any number of relevant cognitions related to the piece of information. For instance, a person may not know that Pakistan has nuclear weapons, but may know that Poland does not. And over the course of an informative process, the repository of related cognitions may grow even while the respondent never learns the particular piece of information. For instance, a respondent may learn that Japan does not have nuclear weapons, but still not learn that Pakistan does. Growth in such related cognitions may be important to the researchers. And given how we motivate the model, it is not immediately clear whether the model we propose captures such learning.

Closed-ended items are, naturally, limited instruments for capturing the learning of related cognitions. The only kind of related knowledge that they can capture is growth in cognitions in one of the options in the response set. Staying with the example we use above, if Japan is within the response set of a closed-ended item (with a single correct answer) asking the respondent which of the countries has nuclear weapons, the respondent who has learned that Japan does not have nuclear weapons (but is still ignorant of the fact that Pakistan has nuclear weapons) is more likely to pick Pakistan than another similar respondent without that piece of related knowledge. Such learning is liable to lead to an increase in proportion correct.

Within the multinomial distribution, this means a greater number of people going from 0 (or c) to 1, and lower probabilities of going from 1 to 0 (or c). Our model does not disambiguate between a correct response as a result of related knowledge, and as a result of knowledge of the relevant piece of information. And as proportions of transitions from c to 1 and 0 to 1 increase, the probability of CK and GK also increase. Thus, while we motivate our model without acknowledging related knowledge, the latent class model that we have developed implicitly accounts for gains in related knowledge.

4 Estimation

We estimate the parameters of the system of equations that define the multinomial distribution for a single item (equation (2)) via maximum-likelihood. Letting n denote the sample size, m the number of multinomial cells, x_i the observed counts in each cell i, and $p_i(\theta)$ the individual cell probabilities, the log-likelihood of a multinomial distribution is

$$l(\theta) = log(n!) - \sum_{i=1}^{m} log x_i! + \sum_{i=1}^{m} x_i log(p_i(\theta)) \tag{3}$$

To apply this general form to our model, we need to swap x_i and $p_i(\theta)$ with u and π, respectively. Given that n and $\sum_{i=1}^{m} log u_i!$ for any one item are constant, parameters are estimated by maximizing the last term in the likelihood function, with the constraint that the transition parameters (λ) (see equation (2)) sum to 1:

$$\begin{gathered}
\text{max.} \sum_{i=1}^{m} u_i log(\pi_i) = u_{00}((1-\gamma)^2 \lambda_{GG}) + u_{01}(1-\gamma)\gamma\lambda_{GG} + (1-\gamma)\lambda_{GK}) + \\
u_{00}((1-\gamma)\lambda_{GC}) + u_{10}(\gamma(1-\gamma)\lambda_{GG}) + u_{11}(\gamma^2\lambda_{GG} + \gamma\lambda_{GK}, + \\
\lambda_{KK}) + u_{1c}(\gamma\lambda_{GC}) + u_{c0}((1-\gamma)\lambda_{CG}) + u_{c1}(\gamma\lambda_{CG} + \lambda_{CK}) + u_{cc}(\lambda_{CC}) \\
\text{s.t.} \sum_{i=1}^{m} \lambda_i = 1
\end{gathered} \tag{4}$$

conventional 3-PL model. Without reconstituting the conventional 3-PL model, there are two ways to handle "don't know" responses within it—treat them as missing, or convert them to zeros. Doing either requires us to make untenable assumptions. In the first case, we must assume that "don't know" responses are missing at random when, in fact, we know that they are near-perfect indicators of ignorance (Luskin and Bullock 2011). Doing the latter means the 3-PL IRT model sees them as unlucky guesses, which they are not.

We perform the optimization using a general nonlinear optimizer, Rsolnp (Ghalanos and Theussl 2010), which uses the Augmented Lagrange Multiplier Method. (An additional nonlinear optimizer, Alabama [Varadhan and Grothendieck 2011], using Adaptive Barrier Minimization, produced virtually identical results.) The sum of λ_{GK} and λ_{CK} gives us the guessing-adjusted estimate of the proportion of people who learned a piece of information. We calculated the errors of the parameters (and composites of parameters) using nonparametric bootstrap. In particular, we took 100 random samples with replacement of observed response patterns and re-estimated the results. The standard deviation of these estimates was used as the estimate of the standard error.

5 Monte Carlo Study

To assess how well our model fares in recovering actual knowledge gains over an informative process, we simulated a broad range of informative processes and knowledge items used to measure learning. We simulated informative processes that varied in their efficacy (average learning) and in how the information gains were distributed. And, to measure learning over these processes, we simulated knowledge items that varied widely in their difficulty and in the ease with which people who did not know the right answer could correctly guess it.

We started with a model of the underlying data-generating process. In particular, we chose a 1-PL IRT model—as our quantity of interest in learning over an item, discrimination parameters are unnecessary—which assumes whether or not a person knows a particular piece of information, ($p(x = 1)$), is a stochastic function of their ability (θ), and the difficulty of the piece of information (or, more precisely, the difficulty of the item used to measure it) (δ):

$$p(x = 1) = \frac{1}{1 + e^{-(\theta - \delta)}} \tag{5}$$

The stochastic component of the model is best interpreted as follows: a person of a given ability will know some pieces of information of a given difficulty and not know others. We further assume that the items used to test knowledge of the pieces of information are closed-ended, and that those who know the answer always mark the right answer.

We further assumed that the abilities had a standard normal distribution, $N(0, 1)$. From the standard normal, we randomly drew n respondents. Given that what is easy for the more able is difficult for the less able, ability and difficulty are on the same scale. And, since we had no prior knowledge about the typical difficulty of the items used to measure learning, we took item difficulty to be uniformly distributed. From the uniform distribution across the entire observed range of θ, we randomly drew k item difficulty parameters. Following this, we simulated responses on the preprocess wave using the 1-PL IRT model we specified above.

Next, we simulated a broad range of learning processes that varied in their average efficacy, which we take to be the proportion of respondents learning the item, and in how the gains in learning were distributed. Given that we were modeling an informative process, we took average gains (α) to be non-zero, ranging from 5% to a hefty 50%. And since we did not know whether less efficacious informative processes are more or less probable than more informative processes, we took α to be uniformly distributed between 0.05 and 0.50. From this uniform distribution, we randomly drew k values of α.

At one end are egalitarian informative processes in which each respondent who does not know the item has an equal chance of learning it. Of course, some respondents will learn the item and some will not. Most naturalistic informative processes, however, produce learning that is correlated with preprocess ability (Dooling and Lachman 1971; Bransford and Johnson 1972; Eckhardt et al. 1991; Cooke et al. 1993; Hambrick 2003; Recht and Leslie 1988). We model this kind of process by systematically giving those with greater preprocess ability a greater chance of learning an item. Thus, we specified another parameter (β) that was multiplied to preprocess ability, going from 0 (no correlation with preprocess ability) to 0.6. Among people who did not know the item before the

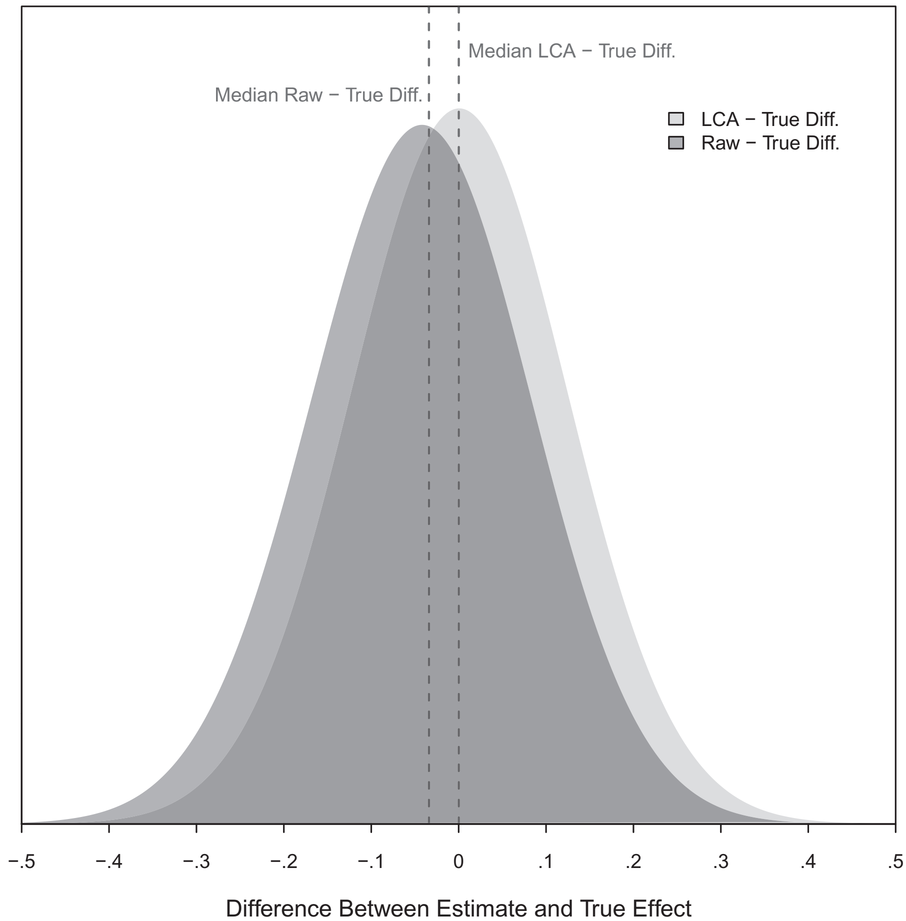

Fig. 1 Distribution of difference between LCA and raw estimates and actual learning.

process ($x = 0$ on the preprocess instrument), we drew a random binomial with probability equal to $\alpha + \beta * \theta$. Thus, learning added to preprocess responses gave us the (true) postprocess indicators of whether the respondents knew the item or not.

Next, we simulated confessing and lucky guessing on top of true indicators of knowledge. On items that the respondent did not know, we first simulated the choice between guessing and confessing. We assumed that the decision to guess or to confess was random, with probability of confession varying between 0.05 and 0.25 (in line with observed rates of confession to ignorance in data we use later). We assumed the chance of guessing an item correctly ranged from .1 to .6. For coverage, we again drew from a random uniform. Simulating these two processes on both waves gave us the observed responses.

As expected, the raw estimator is negatively biased (Bias = –0.044) (Fig. 1). To put the number in perspective, bias was, on average, 31.7% of the true estimate. However, using the latent class analysis (LCA) model that we develop here yields nearly unbiased estimates (Bias = 0.001), or an average of less than 2% of the true estimate. Not only was our model vindicated on bias, it was also vindicated on mean squared error (MSE). The MSE of the raw estimate was four times the LCAs ($MSE_{raw} = 0.0035$, $MSE_{LCA} = 0.0008$). Last, as expected, differences between the raw estimates and actual learning were negatively correlated with item difficulty: the harder the item, the greater the underestimate ($r = -0.67$). However, estimates from our model were uncorrelated with item difficulty ($r = -0.004$).

6 Application

Next, we apply the model to data from the Deliberative Polls. We begin by describing the data.

6.1 *Data and Measures*

Data are from nineteen face-to-face and four online Deliberative Polls (see Fishkin 2016). (See Online Appendix C for details about each of the polls included in the study.) In a face-to-face Deliberative Poll, a random sample is interviewed. Sometime later the entire interviewed sample, or a random subset, is invited to deliberate. Invitees are typically offered a small honorarium. However, not everyone who is invited comes. People who take part in deliberations tend to be more knowledgeable and better educated than those interviewed initially (see O'Flynn and Sood 2014). Apart from that, participants are similar to nonparticipants on a variety of demographic characteristics.

If the site of the deliberation is nonlocal, travel to the site is reimbursed, and a free hotel stay is arranged. Invitees are mailed balanced briefing materials prior to deliberation. Of those who accept the invitation, typically a reasonably large proportion turns up to deliberate. At the site, respondents are randomly assigned to small groups. Deliberation in small groups is facilitated by trained moderators. Participants also get a chance to quiz experts in plenary sessions. At the end of deliberation, participants fill out a survey featuring many of the same questions they were posed initially, along with some new ones.

Online Deliberative Polls work much the same way—like face-to-face Deliberative Polls, a random sample is interviewed, and a random subset of those interviewed is invited to deliberate; briefing materials are provided in advance of deliberations. But a few things are different in online polls. Participants cannot see one another, and rely on voice alone to communicate. And small groups are not allocated randomly but decided upon opportunistically, based on times that groups of participants find convenient.

In each of these polls, respondents' knowledge of policy-relevant facts is measured using, mostly, closed-ended items (we found only two open-ended items in the twenty-three polls we analyze here). Almost always, all the knowledge items that are fielded in the pre-deliberation survey are repeated in the post-deliberation survey. In analyses below, we subset on closed-ended questions that are asked both pre- and post-deliberation. In all, we have data on 177 items. All the closed-ended items offered a "don't know" option. For placement items, correct absolute placement is scored as correct (see Luskin and Bullock 2004). (See Online Appendix D for question wording and response options.)

6.2 Results

6.2.1 Learning specific pieces of information

Based on differences in proportion correct, a significant proportion of the respondents learned the pieces of information that they were surveyed on.[9] Across all 177 items in twenty-three polls, on average, approximately 16% of the respondents learned the piece of information they were surveyed about based on the conventional estimator of knowledge gain (Mean $= 0.158$, SE $= 0.01$). The estimated proportion of respondents who learned the piece of information varied widely across items ($s = 0.158$). Regardless, even based on the "raw" estimator, which we expect to be negatively biased, participants learned a fair bit on items they were surveyed on.

As we reason above, accounting for guessing is liable to lead us to (correctly) infer that a larger proportion of respondents learned items they were surveyed on. But before we present guessing-adjusted estimates of learning, we present some evidence consistent with our heuristic account of why adjusted estimates are liable to be higher than raw estimates. Our claim rests upon the insight that when people know less, they have a greater opportunity to guess, and that people know as much, or less, before an informative process, than after. Hence, there is likely more guessing (and very likely more successful guessing) in the survey conducted before the process than after it.

[9]Note that we lack data on how much people did not deliberately learn over the same time. While we think it is unlikely that people would have learned the pieces of information they were surveyed on without the Deliberative Polling treatment, it is possible that the causal effects of Deliberative Poll on knowledge are lower.

Across all items, 29% of the respondents guessed incorrectly on the preprocess wave. Observable lucky guessing in the preprocess wave—percent "10" and "1c" response patterns—was about 8.7%. In all, at least 38% of the participants guessed in the pretreatment wave. As expected, somewhat fewer respondents guessed incorrectly on the posttreatment wave (25%).

Another indicator of greater guessing-related error on the preprocess wave, vis-à-vis the postprocess wave, would be its lower reliability. We find as much in the data. Across the twenty-three polls, reliability of preprocess instrument was significantly lower than the reliability of posttreatment instrument (Mean $\alpha_{T_1} = 0.495$, Mean $\alpha_{T_2} = 0.561$, Diff. $= 0.065$, $p < .05$). (The p-value is based on the sign test. In eighteen of twenty-three polls, the preprocess instrument had lower reliability than the postprocess instrument.) In all, the evidence is consistent with the idea that more people guessed (luckily) on the preprocess instrument than on the postprocess instrument.

Next, we present estimates of learning adjusted for guessing using two different methods: the standard guessing correction,[10] and the latent class model we have developed in the paper. We estimated learning using both the models for each of the 177 closed-ended knowledge items. As we had expected, once we account for guessing, the estimates of learning are larger ($\text{Mean}_{\text{LCA}} = 0.210$, $\text{SE}_{\text{LCA}} = 0.016$; $\text{Mean}_{\text{stnd}} = 0.182$, $\text{SE}_{\text{stnd}} = 0.015$). Altogether, estimates of learning based on the LCA were greater than the raw estimate for 78.5% of the items. And standard-guessing-correction based estimates of learning were larger than the raw estimates for 65% of items. The average difference between LCA and raw estimates was about 5% ($p < .01$) (see also Fig. 2). (We got virtually identical results when we fixed the probability of knowing given a don't know response as 3%, per the model in Online Appendix B.) To put this in perspective, accounting for guessing using the LCA model we develop here, we find that nearly 30% more respondents learned a piece of information than what the raw estimates suggest. The average difference between estimates based on standard guessing correction and raw estimates was smaller, about 2% ($p < .01$), but substantively meaningful—in terms of percentage change, the difference was approximately 16%.[11]

6.2.2 Heterogeneous effects

Next, we analyze how accounting for guessing changes inferences about how much different groups of people learn over the Deliberative Poll. In particular, given long-standing concerns about gender and guessing on knowledge questions, we analyze how inferences about how much men learn vis-à-vis women change when we adjust for guessing.

When a "don't know" option is offered, rates of guessing are perforce equal or lower than when such an option is not available. However, different problems may ensue—the opportunity to confess one's ignorance systematically privileges the "kinds of people" (men, according to Mondak 1999, 2001) who choose to guess when they do not know, or choose not to be reticent when in fact they do know. (Reticence by the knowledgeable on closed-ended items has been shown to be extremely infrequent (see Sturgis et al. 2008; Luskin and Bullock 2011). Over an informative process, differences in guessing rates can cloud estimates of subgroup differences in learning.

[10]The standard guessing correction assumes that lucky guessing is squarely a matter of chance. While expositions on standard guessing correction have not typically dealt with "don't know" responses, they are easily handled. Calculations of how many items an individual guessed on need omit "don't know" responses.

Where n denotes the number of respondents, i tracks respondents, k is the number of alternatives, u_1 and u_2 are responses on the pretest and the posttest, respectively, and 1 and 0 indicate a correct and incorrect response, respectively, the standard-guessing-correction based estimate of learning runs as follows:

$$\frac{1}{n}\left(\sum_{i=1}^{n}(u_{2_i}=1) - \frac{1}{k-1}\sum_{i=1}^{n}(u_{2_i}=0) - \sum_{i=1}^{n}(u_{1_i}=1) + \frac{1}{k-1}\sum_{i=1}^{n}(u_{1_i}=0)\right) \tag{6}$$

[11]As we note above, the latent class model implicitly accounts for gains in related knowledge. Under some implausible circumstances, however, the latent class model may miss some gains in related knowledge (these latter gains will also be missed by the conventional estimator). We discuss these implausible circumstances, and empirically assess whether such circumstances are cause for concern in Online Appendix E. We find little evidence of these implausible circumstances compromising inference.

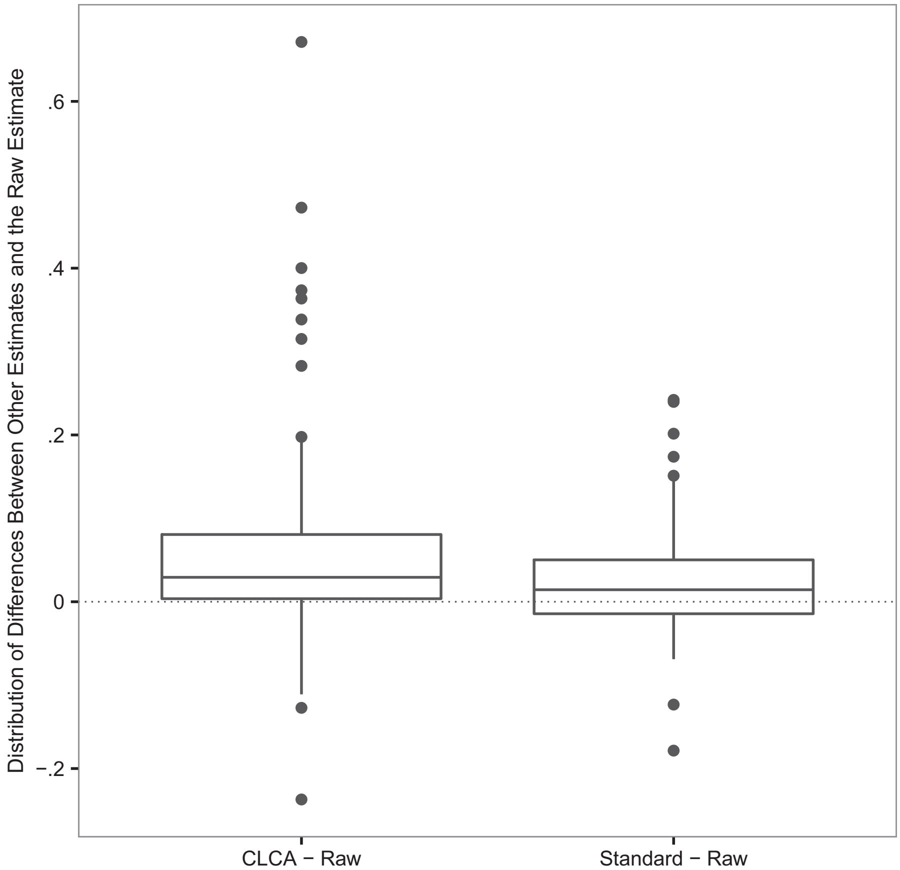

Fig. 2 Distribution of differences between guessing-adjusted and raw estimates.

To account for differences in guessing by gender, we calculated the probability that an observed correct answer reflects knowledge, $P(K|u=1)$, at T_1 and at T_2 for men and women, separately. To calculate this, we first calculate the item-level probability of a lucky guess, $P(u=1|G)$, which is constant, followed by the propensity to guess within each subgroup at each time. Next, we replaced all the 1s at $T1$ and $T2$ with the probability of knowing given an observed 1 and recalculated simple pre–post differences. This results in guessing adjusted estimates of learning that also account for subgroup differences in guessing rates.

The $P(K|u=1)$ for men and women separately at each time is calculated using a simple counting principles approach. Let n_l denote total number of lucky guesses, n_u total number of unlucky guesses, and n_g total number of items a person guessed on, the sum of n_l and n_u. The LCA model defines $P(u=1|G)$ as γ while the standard guessing correction model assumes $P(u=1|G)=1/k$, where k is the number of response options on an item. The conditional probabilities of someone picking the right answer when they do not know are then

$$P(u=1|G)=\gamma=\frac{P(G|u=1)P(u=1)}{P(G)} \tag{7}$$

and

$$P(u=1|G)=\frac{1}{k}=\frac{P(G|u=1)P(u=1)}{P(G)} \tag{8}$$

Using these equations, and that n_g is the sum of n_l and n_u, we can solve for the total number of lucky guesses in terms of γ and n_u and k and n_u. Doing so yields the following equations:

$$n_l = \frac{n_u \gamma}{1 - \gamma} \tag{9}$$

$$n_l = \frac{n_u}{k - 1} \tag{10}$$

Dividing the above equations by the total number of observed correct answers gives the probabilities that an observed correct answer is a lucky guess. Subtracting them from 1 produces the required probabilities $P(K|u = 1)$ for the two different types of corrections.

Across all twenty-three polls, based on raw estimates, women learned more than men (Mean Diff. $= -0.022$, SE $= 0.006$). As Figure 3 shows, accounting for guessing leads to larger positive corrections for men than for women. Hence, once we account for guessing, the "learning gap" nearly disappears (MeanDiff.$_{\text{LCA}} = -0.007$, SE $= 0.004$; MeanDiff.$_{\text{stnd}} = -0.003$, SE $= 0.005$).

One plausible explanation for why the learning gap between men and women disappears after we adjust for guessing is that men and women vary in their propensity to guess when they do not know. In a world where men are more likely to guess (as opposed to confess) when they do not know than women, guessing-related attenuation would be greater in men's scores than women's in the conventional estimates.

The differences that cloud subgroup differences in learning also cloud subgroup differences in knowledge at any particular wave. For instance, differences in guessing may partly explain the "knowledge gap" across genders (Delli Carpini and Keeter 1996; Verba et al. 1997; Frazer and Macdonald 2003; Mondak and Anderson 2004). To assess this possibility, we compared gender gap

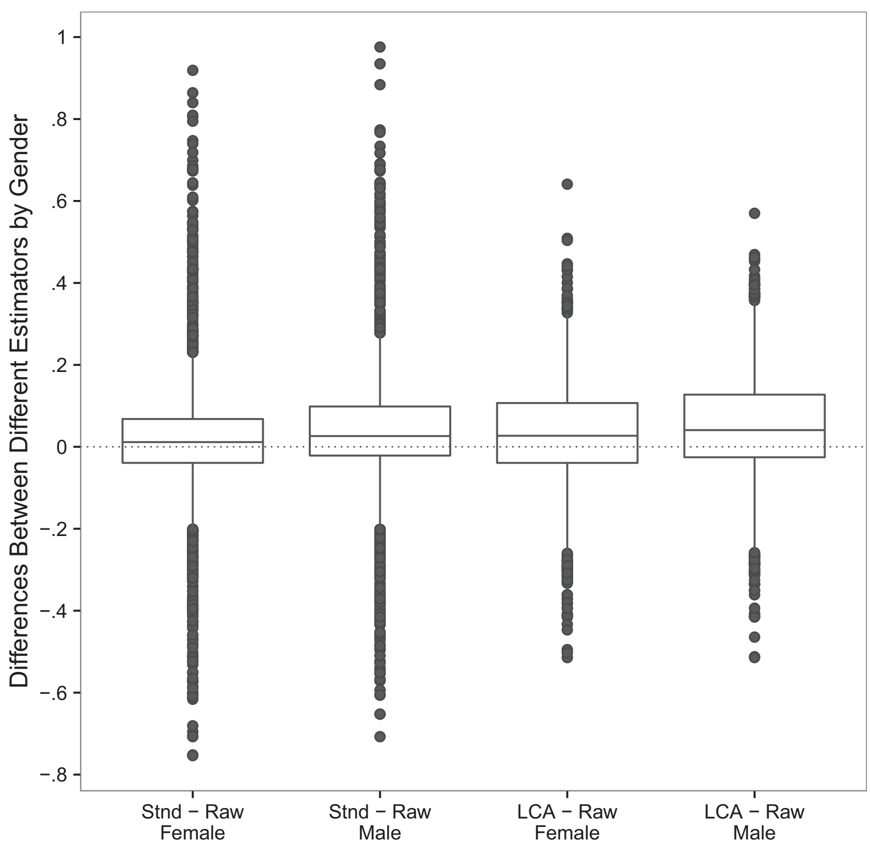

Fig. 3 Distribution of differences between guessing-adjusted and raw estimates by gender.

https://doi.org/10.1093/pan/mpw010 Published online by Cambridge University Press

in conventional estimates to gap in guessing adjusted estimates on the pre-deliberation wave. On average, based on the raw scores, women knew far less than men (Mean Diff. = 0.067, SE = 0.005). Once you adjust for guessing, however, the knowledge gap nearly halves ($\text{MeanDiff.}_{\text{stnd}} = .037$, SE = 0.004; $\text{MeanDiff.}_{\text{LCA}} = .039$, SE = 0.003). (The decline in estimated gender gap when you account for guessing is itself highly statistically significant, $p < .001$.) Significantly, our results match those obtained by Mondak and Anderson (2004), who also find that “approximately 50% of the gender gap is illusory.”

6.2.3 Learning over polls

One can use estimates of observed learning over specific pieces of information to infer learning over other unsampled pieces of information. Such inference rests upon one assumption—dimensionality of learning—and its generalizability is limited by biases in sampling. For dimensionality, we simply assume that learning is unidimensional. As for biases in sampling, like many other surveys, the pieces of information people are surveyed on in Deliberative Polls are considerably “easier” (Luskin et al. 2014). This limits inference to learning on similarly easy bits of information, though see Luskin et al. (2014) for some other issues that may sabotage inference.

We use the simple average of learning across items as an estimate of average learning of similar such items in a poll. On average, Deliberative Poll participants got a significantly greater proportion of items correct in surveys conducted right after the polls than in surveys conducted weeks before. While the increase over the polls—mean of item-level increases—varied considerably (Range: 0.027, 0.397), the average increase (inverse variance weighted) was a hefty .14 (SE = 0.003) (see Table F1 in Online Appendix F; Fig. 4). Expectedly, adjusting for guessing

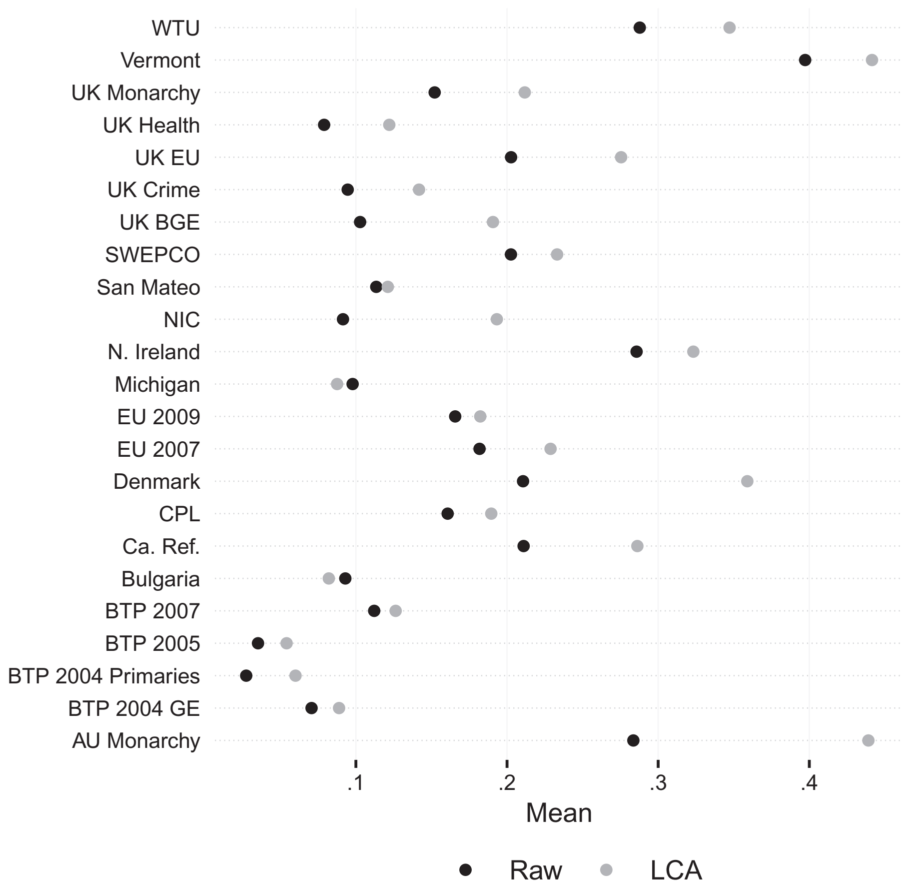

Fig. 4 LCA and raw estimates of learning by poll.

Table 3 Proportion of items in each poll with predicted patterns different from observed data

Poll	*LCA*
AU monarchy	0.3
BTP 2004 primaries	0
BTP 2004 GE	0.333
BTP 2005	0
BTP 2007	0.125
Bulgaria	0.143
Ca. Ref.	0
CPL	0.286
Denmark	0.222
EU 2007	0
EU 2009	0.167
N. Ireland	0.143
Michigan	0.222
NIC	0.125
San Mateo	0.25
SWEPCO	0.2
UK BGE	0.067
UK crime	0
UK EU	0
UK health	0.167
UK monarchy	0.125
Vermont	0.667
WTU	0.2
Weighted average	0.169

(using either standard guessing correction or LCA) boosts the estimates of learning—the average learning based on the LCA estimator is 0.171 (SE = 0.004; $p_{\mathrm{Diff.}} < .01$), and the estimate based on the standard guessing correction is 0.168 (SE = 0.003; $p_{\mathrm{Diff.}} < .01$).

6.2.4 Goodness of fit

Last, we check how good our model is at reproducing the observed distribution of response patterns. We find that the latent class model captures the data reasonably well. Based on the χ^2-test, the model reproduces response patterns that are consistent with the data on 83.1% of the items (see Table 3).

In all, the analyses suggest three things. First, conventional estimates of learning are negatively biased. Second, using any of the major models that adjust for guessing reduces bias. And third, adjusting for guessing using latent class models yields particularly good estimates of the true effect.

7 Discussion

When analyzing the impact of informative processes, one basic question that we often want an answer to is "What proportion of the respondents learned a particular item over the course of an informative process?" To answer this, one needs to know how many knew the information before the process, and how many knew it right after the process. Ask once, ask twice. Rinse and repeat. However, as we show above, this may not be enough. Our ability to detect whether a respondent knows something at each time point is hindered by lucky guessing by the uninformed.

Naive estimates of learning over an intervention are biased due to guessing. And while guessing by the uninformed may be random, its effects are distinctly nonrandom. Guessing-related error is negatively correlated with true knowledge. And not taking guessing into account leads us to underestimate learning over an informative process. Guessing may also bias inferences about differences in how much various subgroups learn over an informative process. This may happen simply because some groups of people start out knowing less than other groups of people or because a group has a greater propensity to guess than another. Adjusting for guessing using any of the popular methods, such as the standard guessing correction, and IRT, reduces bias. However, the latent class model developed here is unbiased and has the lowest MSE.

Accounting for guessing post hoc is, however, necessarily imperfect—for any particular response, there is no way to know the "latent state" without error. Thus, rather than account for guessing post hoc, it may be better to reduce guessing in the measurement instrument. As is readily intuited, guessing is largely a concern on closed-ended items. Thus, one cure may be to replace closed-ended items with open-ended items. However, that may be no panacea. Concerns about motivation to search memory for the answer become important on open-ended items. In particular, estimates of learning may be biased by differential motivation to recall the answer, pre- and postprocess (see also Tallmadge 1982). For instance, in the specific informative process we discussed—Deliberative Polling—it may well be the case that respondents are more motivated to recall a piece of information after having intensely discussed an issue than before deliberation. These kinds of concerns are not material when the task is merely to recognize the correct answer from a small set of options. Separately, coding responses to open-ended questions presents its own challenges (Gibson and Caldeira 2009). But perhaps more to the point, open-ended questions do not preclude guessing. The respondent who does not know the correct answer may nonetheless guess from among a set of possibilities. For example, a respondent may not know that the president of the United States is Mr. Obama, but may be pretty sure that it is either Mr. Obama or Mr. Clinton or Mr. Bush, and guess randomly from among those possibilities.

Another, perhaps more promising, way to preempt the need for building post hoc models to account for guessing may be to include additional questions that assess respondents' certainty about their responses (see, for instance, Alvarez and Franklin 1994; Pasek et al. 2015). Measuring certainty has the virtue of expanding the scope of what can be measured in three important ways. Assume, for instance, an absolute scale on which a politician's positions could be placed. Assume also that, for strategic reasons, the politician is deliberately unclear about his or her position, indicating just a range. And that the respondent, a particularly avid consumer of politics, knows the range exactly. In conventional closed-ended questions, such a respondent may be scored as uninformed. But certainty and range assessments can be used to capture such knowledge (see, for instance, Franklin 1991; Alvarez and Franklin 1994; Alvarez and Glasgow 1996). Not only that, certainty measures allow the tracking of changes in certainty, which may capture changes in how easy it is for the respondent to recall the information, over an informative process (see again Franklin 1991; Alvarez and Franklin 1994; Alvarez and Glasgow 1996).

Certainty assessments can also be used to measure misinformation. On conventional instruments, there is no way to distinguish misinformation from ignorance without making further assumptions. And, in the presence of misinformation (and the possibility of knowledge being replaced by misinformation), our identifying assumption would no longer hold. Supplementary questions that assess respondents' certainty about the responses, however, may provide one way to resolve the issue. One may take positions on which a respondent is certain and incorrect as indications of misinformation—belief in incorrect information (see, for instance, Pasek et al. 2015).

Including certainty measures has yet another benefit. A great deal of research suggests that certainty assessments are substantively important, capturing substantively important things like candidate strategies (Franklin 1991) and affecting respondent decision-making (Alvarez and Franklin 1994; Alvarez and Glasgow 1996; Bartels 1986).

Hitherto, we have focused on concerns related to measurement of learning of specific pieces of information. However, we often want to infer how much people learned in general over an informative process. Estimates of gross learning over a rich naturalistic informative process like Deliberative Polling are subject to multiple other concerns. For one, as Luskin et al. (2014) note, there is a great deal of item sampling bias in typical knowledge batteries. Add to it the tendency of the rich to get richer over these naturalistic informative processes and the fact that ceiling effects cap observable growth in learning. A combination of these issues can contrive to create a scenario where observed gains in knowledge are *negatively* correlated with true knowledge gain (Luskin et al. 2014).

Besides affecting estimates of learning, guessing likely also affects cross-sectional measures of political knowledge. Guessing likely abrades correlations with criteria. Guessing, as we note above, causes only positive error when the items are conventionally scored as proportion correct. And this positive error is negatively correlated with knowledge—the more uninformed one is, the greater the positive error in their score on average. When error is negatively correlated with the true measure, it weakens correlation with variables the latent concept ought to be correlated with. Guessing-related error in cross-sectional measures may be addressed by another measurement wave, followed by a latent class model, much like ours, except identified by the assumption that respondents cannot go from marking an item correctly to marking it incorrectly, and vice versa.

In all, our aim in this article was to draw attention to an important problem that affects measurement of learning, and to propose a way to solve the problem. The problem we highlight affects areas other than politics—most prominently, education. In a nontrivial number of cases, educational researchers still use naive estimates of learning when measuring gains from implementing certain programs. And accounting for guessing using a latent class estimate may provide a much better account of the efficacy of the programs. However, many opportunities for improving the model still remain. A natural extension of the latent class model developed in the article would be a two-wave latent trait model that allows for constraints on cross-wave transitions. Such a model of learning, which accounts for guessing by exploiting cross-wave transitions, is liable to provide better estimates of learning than simple averages, and perhaps even IRT. Second, extensions to the model that exploit data from more than two waves are liable to allow us to better account for guessing-related error by leveraging more data to estimate γ. (We present a model that exploits data from three waves in Online Appendix G.) And third, ancillary data, such as certainty of responses (for instance, Pasek et al. 2015), and data from experiments in which correct answers are incentivized (see, for instance, Prior and Lupia 2008; Prior et al. 2015), may prove useful in building larger (and better) models for measuring learning.

Conflict of interest statement. None declared

References

Alvarez, R. M., and C. H. Franklin. 1994. Uncertainty and political perceptions. *Journal of Politics* 56(3): 671–88.

Alvarez, R. M., and G. Glasgow. 1996. Do voters learn from presidential election campaigns? Technical report.

Bartels, L. M. 1986. Issue voting under uncertainty: An empirical test. *American Journal of Political Science* 30(4): 709–28.

Bransford, J. D., and M. K. Johnson. 1972. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior* 11(6): 717–26.

Cooke, N. J., R. S. Atlas, D. M. Lane, and R. C. Berger. 1993. Role of high-level knowledge in memory for chess positions. *American Journal of Psychology* 106(3): 321–51.

Cor, K., and G. Sood. 2016. Replication data for: Guessing and forgetting: A latent class model for measuring learning. http://dx.doi.org/10.7910/DVN/HZHVCU, Harvard Dataverse, V1 (accessed February 3, 2016).

Cor, M. K. 2012. An experimental test of the effect of norm-referenced and criterion referenced feedback on just world beliefs, motivation, and performance: Does social disadvantage matter? PhD thesis, Stanford University, Stanford, CA.

Delli Carpini, M. X., and S. Keeter. 1996. *What Americans know about politics and why it matters*. New Haven: Yale University Press.

Dooling, D. J., and R. Lachman. 1971. Effects of comprehension on retention of prose. *Journal of Experimental Psychology* 88(2): 216.

Eckhardt, B. B., M. R. Wood, and R. S. Jacobvitz. 1991. Verbal ability and prior knowledge contributions to adults' comprehension of television. *Communication Research* 18(5): 636–49.

Fishkin, J. 2009. *When the people speak: Deliberative democracy and public consultation*. Oxford: Oxford University Press.

——— 2016. Center for deliberative polling, Stanford University. http://cdd.stanford.edu (accessed February 4, 2016).

Franklin, C. H. 1991. Eschewing obfuscation? Campaigns and the perception of US senate incumbents. *American Political Science Review* 85 (04): 1193–214.

Frazer, E., and K. Macdonald. 2003. Sex differences in political knowledge in Britain. *Political Studies* 51(1): 67–83.

Ghalanos, A., and S. Theussl. 2010. *Rsolnp: General non-linear optimization*. R package version 1–0. https://cran.r-project.org/web/packages/Rsolnp/Rsolnp.pdf.

Gibson, J. L., and G. A. Caldeira. 2009. Knowing the Supreme Court? A reconsideration of public ignorance of the high court. *Journal of Politics* 71(2): 429–41.
Hambrick, D. Z. 2003. Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition* 31(6): 902–17.
Hansen, E. A., F. L. Schmidt, and J. C. Hansen. 1975. A model for the correction for guessing on multiple-choice tests. *ACM SIGSOC Bulletin* 7(1–4): 24–8.
Iyengar, S. 1990. Shortcuts to political knowledge: The role of selective attention and accessibility. *Information and Democratic Processes,* 160–85.
Lenz, G. S. 2009. Learning and opinion change, not priming: Reconsidering the priming hypothesis. *American Journal of Political Science* 53(4): 821–37.
Lord, F. M. 1975. Formula scoring and number-right scoring. *Journal of Educational Measurement* 12(1): 7–11.
Lord, F. M., M. R. Novick, and A. Birnbaum. 1968. *Statistical theories of mental test scores*. Addison-Wesley.
Luskin, R. C., and J. G. Bullock. 2004. Measuring political knowledge. Paper presented at the Annual Meeting of the Midwest Political Science Association.
———. 2011. "Don't know" means "don't know": DK responses and the public's level of political knowledge. *Journal of Politics* 73(2): 547–57.
Luskin, R. C., J. S. Fishkin, and R. Jowell. 2002. Considered opinions: Deliberative polling in Britain. *British Journal of Political Science* 32(3): 455–87.
Luskin, R. C., and G. Sood. 2012. Hidden knowledge (and veiled ignorance): Is the public distinctly less (or even more) ignorant than we have thought? In Annual Meeting of the Midwest Political Science Association.
Luskin, R. C., G. Sood, and A. Helfer. 2014. Measuring learning in informative processes. Paper presented at the European Consortium of Political Science Research Workshop, Salamanca, Spain.
Mondak, J. J. 1999. Reconsidering the measurement of political knowledge. *Political Analysis* 8(1): 57–82.
——— 2001. Developing valid knowledge scales. *American Journal of Political Science,* 224–38.
Mondak, J. J., and M. R. Anderson. 2004. The knowledge gap: A reexamination of gender-based differences in political knowledge. *Journal of Politics* 66(2): 492–512.
Nunnally, J. C. 1967. *Psychometric theory*. Tata McGraw-Hill Education.
O'Flynn, I., and G. Sood. 2014. What would Dahl say? An appraisal of the democratic credentials of deliberative polls and other mini-publics. In: *Deliberative mini-publics*, eds. K. Gronlund, A. Backtiger, and M. Setala, 41–58. Colchester: ECPR Press.
Pasek, J., G. Sood, and J. Krosnick. 2015. Misinformed about the affordable care act? Leveraging certainty to assess the prevalence of misinformation. *Journal of Communication* 65(4): 660–73.
Prior, M., and A. Lupia. 2008. Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science* 52(1): 169–83.
Prior, M., G. Sood, and K. Khanna. 2015. You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science* 10(4): 489–518.
Recht, D. R., and L. Leslie. 1988. Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology* 80(1): 16.
Sanchez, M. E., and G. Morchio. 1992. Probing "dont know" answers: Effects on survey estimates and variable relationships. *Public Opinion Quarterly* 56(4): 454–74.
Sniderman, P. M., R. A. Brody, and P. E. Tetlock. 1993. *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.
Stolle, D., and E. Gidengil. 2010. What do women really know? A gendered analysis of varieties of political knowledge. *Perspectives on Politics* 8(1): 93–109.
Sturgis, P., N. Allum, and P. Smith. 2008. An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly* 72(1): 90–102.
Tallmadge, G. K. 1982. The correction for guessing a case in which its use made treatment effects appear larger than they really were. *Evaluation Review* 6(6), 837–41.
Varadhan, R., and G. Grothendieck. 2011. *Alabama: Constrained nonlinear optimization*. R package version 1.
Verba, S., N. Burns, and K. L. Schlozman. 1997. Knowing and caring about politics: Gender and political engagement. *Journal of Politics* 59(4): 1051–72.