Modeling Guessing Components in the Measurement of Political Knowledge

Tsung-han Tsai[1] and Chang-chih Lin[2]

[1] *Assistant Professor, Department of Political Science, National Chengchi University, No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC. Email: thtsai@nccu.edu.tw*
[2] *Adjunct Assistant Professor, Department of Political Science, National Chengchi University, No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC. Email: lincc@nccu.edu.tw*

Abstract

Due to the crucial role of political knowledge in democratic participation, the measurement of political knowledge has been a major concern in the discipline of political science. Common formats used for political knowledge questions include multiple-choice items and open-ended identification questions. The conventional wisdom holds that multiple-choice items induce guessing behavior, which leads to underestimated item-difficulty parameters and biased estimates of political knowledge. This article examines guessing behavior in multiple-choice items and argues that a successful guess requires certain levels of knowledge conditional on the difficulties of items. To deal with this issue, we propose a Bayesian IRT guessing model that accommodates the guessing components of item responses. The proposed model is applied to analyzing survey data in Taiwan, and the results show that the proposed model appropriately describes the guessing components based on respondents' levels of political knowledge and item characteristics. That is, in general, partially informed respondents are more likely to have a successful guess because well-informed respondents do not need to guess and barely informed ones are highly seducible by the attractive distractors. We also examine the gender gap in political knowledge and find that, even when the guessing effect is accounted for, men are more knowledgeable than women about political affairs, which is consistent with the literature.

1 Introduction

Political knowledge is a central construct in theories of democracy, most of which suggest that knowledgeable citizens are one of the components of a well-functioning democracy. The logic behind these theories is as follows: the more knowledgeable, informed a citizen is, the more deliberate decisions s/he makes and the more s/he participates in democratic processes (see e.g., Campbell *et al.* 1960; Galston 2001; Lassen 2005). To test these theories empirically, many efforts have been made to measure political knowledge in the studies of public opinion and political behavior (e.g., Luskin 1987; Delli Carpini and Keeter 1996; Mondak 1999). The conventional approach to measuring levels of political knowledge is conducted by asking survey respondents a brief battery of factual questions about politics and counting the number of correct answers. Common formats used for these survey questions include closed-ended (multiple-choice and true–false) items and open-ended identification items.

The selection of item formats involves two related issues that may affect the estimates of political knowledge: items that are likely to be guessed correctly and the levels of knowledge behind "don't know" (DK) responses in the measurement procedure (Mondak 1999, 2001). On closed-ended items, one primary issue is that respondents who have no knowledge at all or

Political Analysis (2017)

DOI: 10.1017/pan.2017.21

Corresponding author
Tsung-han Tsai

Edited by
R. Michael Alvarez

Authors' note: The authors would like to thank Chia-yi Lee, Kevin A. Clarke, T. Y. Wang, Simon Jackman, Michelle Torres, Chi Huang, Dean Lacy, Brett Benson, Emerson Niou, Chung-li Wu, and Ching-ping Tang for helpful comments. Earlier versions of this manuscript were presented at the 2014 APSA Annual Meeting, 2015 Asian Political Methodology Conference, and 2015 APSA Annual Meeting. The research is financed by the Ministry of Science and Technology, R.O.C. under grant MOST 104-2410-H-004-091-MY2. We are also grateful to the editor, Michael Alvarez, and two anonymous reviewers for helpful suggestions. The replication materials are available at doi:10.7910/DVN/80WUQB (Tsai and Lin 2017).

those who have partial knowledge about a question are able to guess an answer from provided choice options, which leads to overestimated knowledge scores if they have a successful guess. On the other hand, although open-ended questions are more difficult than closed-format items for respondents to guess, barely and partially informed respondents who are uncertain of the answer to a question are likely to say "DK" due to individual-level characteristics, which suggests that there might be some knowledge hidden within DK responses (Mondak and Halperin 2008; Mondak 2010).

This article examines guessing behavior by focusing on multiple-choice items with the DK-neutral condition and argues that "informed guessing" (i.e., guessing the correct answer based on partial knowledge) is more common than "lucky guessing" or "blind guessing" (i.e., guessing the correct answer randomly). Building on the methods of item response theory (IRT), we propose an IRT model with a guessing component that accommodates the chance of guessing the correct answer for a multiple-choice item based on a respondent's level of political knowledge and the item's difficulty. We term this model "two-parameter logistic guessing" (2PL-G) model and show that the proposed 2PL-G model successfully identifies items likely to be correctly guessed. The 2PL-G model is constructed and estimated by a Bayesian approach, which offers flexibility for complex model specifications. In the estimation process, we treat DK responses as "missing values" rather than collapsing DK responses and incorrect responses into the same category, which reflects our ignorance of these respondents' knowledge levels.

The 2PL-G model is applied to analyzing survey data conducted in 2012 from the Taiwan's Election and Democratization Study (TEDS) project. The results show that the proposed model accurately describes the characteristics of the political knowledge items employed in the TEDS2012 project. In particular, the guessing property of these items corresponds to what we expect for the public's political knowledge in Taiwan. This article has two primary contributions. First, substantively, this article contributes to our understanding of guessing behavior in multiple-choice items. The evidence shows that informed (and/or misinformed) guessing is actually more common than blind guessing. One implication of this fact is that the responses in multiple-choice items can appropriately reflect actual knowledge when informed guessing is recognized and modeled accordingly. Second, methodologically, the proposed 2PL-G model appropriately describes the guessing components based on respondents' levels of political knowledge and item difficulties. While the focus of this article is on political knowledge, the model proposed here can be applied to the multiple-choice format for measuring other types of knowledge in surveys.

The remainder of this article proceeds as follows. Section 2 reviews the extant literature on the measurement of political knowledge and discusses guessing behavior in multiple-choice items, followed by an illustration of the proposed 2PL-G model in Section 3. Section 4 introduces the analyzed data and presents the results of analysis. Section 5 concludes the paper.

2 Measurement of Political Knowledge and Guessing Behavior

Due to the crucial role of political knowledge in democratic participation, the measurement of political knowledge has been a major concern in the discipline of political science (Luskin 1987, 1990; Delli Carpini and Keeter 1996; Mondak 1999). In most of the survey research in political science, the common procedure to measure political knowledge is that respondents are asked a series of questions about politics, such as their awareness and cognitions of "textbook facts," "current events," and "historical facts" (Delli Carpini and Keeter 1991, 1993). The responses to political knowledge questions basically fall into three categories: correct responses, incorrect responses, and DKs. Conventionally, DKs are treated as incorrect responses and then the levels of political knowledge are constructed by counting the number of correct responses. This conventional procedure for the measurement of political knowledge involves at least two related issues. One issue is that the selection of item formats is closely associated with guessing behavior; the other is about whether there is knowledge hidden within DK responses.

To diminish the contamination effect of guessing propensity on the measurement of knowledge, Delli Carpini and Keeter (1996) advocate that the DK option should be encouraged to increase the validity of measurement in the survey because the DK-encouraged strategy will somewhat reduce the tendency for respondents to guess, especially for the completely uninformed or partially informed respondents.[1] However, Mondak (1999, 2001) argues that respondents' propensity to guess is not completely eliminated by encouraging the DK option which still threatens the validity of knowledge measures. Rather, it is recommended that one should discourage DK responses in multiple-choice items because propensity to guess is eliminated if all respondents are forced to choose an answer despite their tendency to guess and, thus, knowledge scores reflect only one systematic factor, that is, actual knowledge levels (Mondak 2001; Mondak and Davis 2001).[2]

In addition to the propensity to guess, Mondak's forced-choice procedure implicitly assumes that there is variation among the levels of political knowledge for respondents who initially say they do not know. When these respondents are forced to guess, the partially informed and uninformed respondents will guess with successful guessing determined by partial information and by chance (Mondak and Davis 2001, p. 207).[3] Following this logic, Miller and Orr (2008) claim that the DK-omitted strategy is better than both the DK-encouraged and DK-discouraged strategies to eliminate guessing propensity and to reveal partial knowledge despite slightly inflated knowledge estimates due to blind guessing.[4] Employing DK-neutrality as a reference strategy, Luskin and Bullock (2011) show that there is concealment of partial knowledge within DK responses under the DK-encouraged condition rather than under the DK-discouraged condition.[5] One implication is that the multiple-choice format with the DK-neutral strategy not only eliminates the propensity of guessing but also reveals partial knowledge, with blind guessing reduced.[6]

Several factors are associated with guessing behavior in the measurement of political knowledge, including the levels of respondents' certainty about politics, the format of items, the presence/absence of the DK option, and respondents' personality (e.g., the propensity to guess). Most of the existing literature on the measurement of political knowledge, however, emphasizes the importance of the latter three factors but ignores the first one.[7] To fill this gap, we focus on respondents' certainty about politics, which involves respondents' levels of knowledge and the difficulty of political knowledge questions. In the following discussion, we assume that the levels of knowledge construct a spectrum, with complete ignorance and perfect informativeness as two extremes. Therefore, the majority of respondents are partially informed and a successful guess depends on the level of partial knowledge. In this sense, guessing behavior is actually referred

1 In that way, survey interviewers will give respondents some prompts like that "many people don't know the answers to these questions, so if there are some you don't know just tell me and we'll go on" (Delli Carpini and Keeter 1996, p. 305).

2 In this version of knowledge questions, interviewers will give a preamble such as "many people don't know the answers to these questions, but even if you're not sure, I'd like you to tell me your best guess" (Mondak 2001, p. 226).

3 When DK responses are discouraged, the remaining DKs account for only a few proportions of responses, which suggests a firm resistance to guessing (Mondak 1999, 2001). To further eliminate DK responses, Mondak (1999, 2001) suggests that one can do a simple post hoc correction by randomly assigning DKs to the substantive choice categories.

4 The DK-omitted strategy advocated by Miller and Orr (2008) was conducted via web surveys, in which the DK option was not provided in the political knowledge batteries.

5 In the DK-neutral version of knowledge items, respondents are asked the knowledge questions without any preface encouraging or discouraging DKs or any DK options (Luskin and Bullock 2011, p. 549). But DK responses will be recorded if respondents say so.

6 In a recent work, Jessee (2017) shows that DK responses do not conceal much knowledge and that the tendency to give DKs does not differ by the types of personality, which implies that respondents' personality is not associated with guessing behavior. However, since the issues about the concealment of partial knowledge within DKs and that about the association between personality and guessing behavior are unsettled, we take a conservative stand for these issues. In other words, we recognize the possibility of hidden knowledge within DKs and, following Luskin and Bullock (2011), we employ the multiple-choice format with the DK-neutral strategy to deal with these issues.

7 For an exception, see Prior and Lupia (2008).

to as *informed guessing* and blind guessing could be considered as a special case of informed guessing, which relies upon no information.

To investigate the guessing components in the measurement of political knowledge, we consider multiple-choice items with the DK-neutral condition suggested by the literature. We do so for the following reasons. First of all, the choice options in a multiple-choice item can provide the relevant information which motivates respondents to think about the survey questions when they are asked to choose the correct one.[8] Rather than being completely ignorant, many respondents may be momentarily unable to quickly recall the right answer or may have vague information about the correct answer. These choice options can serve for respondents to either recognize (Tedin and Murray 1979; Gibson and Caldeira 2009) or recall (Prior and Lupia 2008) the correct answer.[9] We do not consider open-ended items because, according to the conventional wisdom, respondents are much less likely to guess when they are asked to answer open-ended items. Even if they do, retrieving partial knowledge from open-ended items depends on the coding procedure rather than knowledge questions since there may be some knowledge hidden in the incorrect responses under the conventional coding process (Gibson and Caldeira 2009; Luskin and Bullock 2011).

Second, we assume that, when the DK-neutral strategy is employed along with multiple-choice questions, informed guessing occurs more frequently than blind guessing. Neither encouraged to admit ignorance nor forced to select one among choice options, respondents will choose a substantive answer based on their partial knowledge, which determines the probability of a successful guess/selection. In contrast, respondents who are completely uninformed are very likely to say they "DK" rather than randomly choose an answer. For example, Luskin and Bullock (2011) show that the DK-neutral strategy is not very different from the DK-discouraged strategy in terms of the proportion of DK responses and the amount of information concealed within DK responses, which implies that the DK-neutral strategy successfully makes partially informed respondents guess and leaves uninformed ones on DKs. Furthermore, Luskin and Bullock (2011, p. 552) show that the DK-neutral strategy dramatically reduces blind guessing, compared to the DK-discouraged strategy.

Finally, following a substantial body of literature indicating that the multiple-choice format with the absence of the DK options (either DK-discouragement or DK-omission) effectively excludes respondents' propensity to guess from the measurement of political knowledge (e.g., Mondak 1999, 2001; Miller and Orr 2008), we assume that the multiple-choice format with the DK-neutral condition achieves this goal as well based on the work of Luskin and Bullock (2011). Consider, for example, that when asked to answer a standard political knowledge question without the DK option, respondents who tend to guess will behave just like what they do in answering questions with the presence of the DK option, that is, guessing an answer. However, respondents with a low tendency of guessing will be implicitly rather than explicitly forced to guess. As a result, all but completely uninformed respondents will guess an answer despite their guessing propensity.

The above discussion suggests that the use of multiple-choice questions with the DK-neutral strategy bolsters our emphasis on guessing behavior associated with knowledge levels and item difficulties, without the need to worry much about confounders from item formats, the

8 For example, in exploring whether or not partial knowledge lies hidden in DK responses, Sturgis, Allum, and Smith (2008) show that there is an trivial amount of knowledge concealed in DK responses based on an experiment of true/false knowledge items. By conducting experiments on multiple-choice items, however, Miller and Orr (2008) find partial knowledge hidden within DK responses. The contrast between these findings, as Miller and Orr (2008, p. 779) argue, results from the availability of choice options in multiple-choice items, which may motivate respondents to draw on their partial knowledge.

9 Prior and Lupia (2008) discuss a monetary incentive which motivates respondents' searches of "declarative memory." Here we argue that the choice options can also serve as an incentive for respondents to draw declarative memory.

Table 1. Possible responses to multiple-choice items without DK options.

Knowledge level	Item difficulty		
	Difficult items	Moderate items	Easy items
Fully informed	Correct response almost due to understanding	Correct response almost due to understanding	Correct response almost due to understanding
Well informed	Correct response due mostly to understanding/ partly to informed guessing	Correct response almost due to understanding	Correct response almost due to understanding
Moderately informed	Correct response due mostly to informed guessing; incorrect response due to misinformed guessing	Correct response due mostly to understanding/ Partly to informed guessing	Correct response almost due to understanding
Barely informed	Misinformed guessing or blind guessing	Misinformed guessing or blind guessing	Correct response due partly to informed guessing
Uninformed	Misinformed guessing or blind guessing	Misinformed guessing or blind guessing	Misinformed guessing or blind guessing

presence/absence of the DK option, or the personality of respondents. Mondak and Davis (2001) illustrate possible relationships between knowledge levels and responses to standard multiple-choice questions, but they are more concerned with DK responses than guessing behavior. Modifying Table 1 in Mondak and Davis (2001), we investigate guessing behavior by delineating possible responses corresponding to the level of knowledge and the difficulty of items, which is summarized in Table 1.

As can be seen in Table 1, unlike the four levels of knowledge categorized in Mondak and Davis (2001), we consider five groups of respondents based on the level of (partial) knowledge, given our assumption of an underlying knowledge spectrum. The first two groups consist of fully informed and well-informed respondents who are knowledgeable about politics and, thus, know the answers to easy, moderate, and most of the difficult political knowledge questions. Even though well-informed respondents are not fully certain of the answers to some difficult questions, they are able to rule out some options that are certainly incorrect based on their knowledge. When respondents can rule out implausible options, the probability of a correct response is higher than that of a random guess (Embretson and Reise 2000, p. 71). The exclusion of implausible options indicates that, when they "guess" one among the remaining options, they actually do this through partially informed guessing rather than blind guessing.

Consider, for example, that a well-informed respondent is asked to answer a four-category multiple-choice question. Theoretically, each of the choice options has a probability of 0.25 to be correct by chance. Suppose that, since this respondent is relatively informative, s/he knows that two of the options are incorrect but is uncertain which of the remaining two is correct. When s/he chooses one of the probable options, s/he has a probability of 0.5 to correctly answer the question, which is higher than the probability of a random guess.

The third group is moderately informed respondents who have partial information on politics and are uncertain of the answers to most of the questions except for easy ones. Since moderately informed respondents are less knowledgeable than well-informed ones, the former are more likely to answer questions correctly by guessing than the latter. For items with the same level of difficulty, however, moderately informed respondents have a lower probability of a correct response. Consider the same political knowledge question mentioned above, but now suppose that a moderately informed respondent is able to rule out only one choice option instead. Since this respondent is uncertain of the correct answer, s/he guesses one among the remaining three options. Therefore, the probability of a correct response is 0.33, which is higher than the probability of random guessing but is lower than 0.5.

One implication of the above two examples is that, given the item difficulty, a relatively informed respondent has a higher probability of a correct response. Another implication is that guessing contributes less to the probability of a correct response for a relatively informed respondent (e.g., $0.25/0.5 < 0.25/0.33$). In other words, a respondent with a relatively low level of knowledge is more likely to have correct responses due to guessing. These two implications make intuitive sense but have not been discussed much in this discipline.

The last two groups include barely informed and uninformed respondents who hardly have knowledge about politics. These respondents are almost unable to correctly answer difficult questions even by guessing. For moderate and easy questions, they are not knowledgeable enough to choose the correct answers or even to rule out some choice options. As a result, the probability of a correct response is mostly due to guessing. Moreover, unlike Mondak and Davis (2001), we do not consider misinformed respondents as a distinct group. Instead, we assume that well-informed, moderately informed, barely informed, and uninformed respondents might be misinformed, with the extent determined by the level of knowledge. That is to say, barely informed and uninformed respondents are more likely to be mistaken about political events than partially informed and well-informed respondents. In this regard, when they are uncertain of correct answers, they will guess, probably with the exclusion of the correct answer, based on misleading information rather than by chance.

3 Item Response Theory for Guessing Behavior

In this article, we propose an IRT model with a guessing component that accommodates the chance of guessing the correct answer for a multiple-choice item based on a respondent's level of political knowledge and the item's difficulty. The methods of IRT have been developed in education testing and applied by political scientist to a variety of studies such as the measurement of political knowledge (Delli Carpini and Keeter 1996; Jackman 2000), congressional roll-call data analysis (Jackman 2001; Clinton, Jackman, and Rivers 2004), decisions of the Supreme Court (Martin and Quinn 2002), and levels of democracy (Treier and Jackman 2008). Political knowledge is typically assumed to be an unobserved, latent variable that can be measured by a number of manifest variables asking a respondent's awareness of officeholders and political systems (Pietryka and MacIntosh 2013). However, the estimates of latent traits and item parameters from conventional IRT models would be biased due to guessing behavior. In the following, we introduce the proposed model and its properties.

3.1 The IRT guessing model

The Rasch model and the two-parameter logistic (2PL) model are two of the most widely used IRT models in various applications (Rasch 1960; Lord and Novick 1968; Embretson and Reise 2000; Baker and Kim 2004). The IRT models can be used to assess the item properties through item characteristics curves (ICCs), which describe how changes in the trait level relate to changes in the probability of successful responses. However, without considering the effect of guessing, items that have been correctly guessed appear relatively easier than they would be. Since the difficult items are easier than they would be, the knowledge levels of informed respondents who are able to answer difficult items correctly will be underestimated while those of uninformed and/or partially informed ones will be overestimated.

For the guessing property of an item, it has been observed that the lower tail of the empirical ICC sometimes is asymptotic to a value greater than 0. To show the asymptoticity, suppose that, for each item $k = 1, \dots, K$, respondent $i = 1, \dots, N$ provides a response $(y_{i,k})$, which is either correct $(y_{i,k} = 1)$ or incorrect $(y_{i,k} = 0)$. We assume that these items measure an unidimensional latent variable θ_i, i.e., the level of political knowledge here. The three-parameter logistic (3PL) model which describes this asymptotic behavior is presented as follows:

$$\begin{aligned} \Pr(y_{i,k} = 1|\theta_i, \alpha_k, \beta_k, c_k) &= c_k + (1 - c_k)\Lambda[\beta_k(\theta_i - \alpha_k)] \\ &= c_k + (1 - c_k)\left(\frac{\exp[\beta_k(\theta_i - \alpha_k)]}{1 + \exp[\beta_k(\theta_i - \alpha_k)]}\right), \end{aligned} \tag{1}$$

where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function (cdf), α_k the item-difficulty parameter, β_k the item-discrimination parameter, c_k the asymptotic probability of correct response for $\theta \to -\infty$ (Birnbaum 1968). In the literature, c_k is commonly referred to as the "guessing" parameter, which indicates the probability of an item success for respondents with the lowest trait level (Hambleton and Cook 1977).

Empirically, however, the estimates of c_k from the 3PL model are not equivalent to, and usually smaller than, the randomly guessing probability (Embretson and Reise 2000, p. 73). It lacks an explanation of why it is higher or lower than the randomly guessing probability for respondents with the lowest trait level. Therefore, some research argues that c_k should not be interpreted as a guessing parameter (Lord 1968, 1970). Instead, c_k should be considered as the lower bound for the ICC. This becomes more clear when we rearrange Equation (1) as the following:

$$\Pr(y_{i,k} = 1|\theta_i, \alpha_k, \beta_k, c_k) = \Lambda[\beta_k(\theta_i - \alpha_k)] + (1 - \Lambda[\beta_k(\theta_i - \alpha_k)])c_k. \tag{2}$$

Equation (2) shows that the probability of a correct response to item k for respondent i results from two components. The first component represented by $\Lambda[\beta_k(\theta_i - \alpha_k)]$ is the probability that a respondent works on the item to find the correct answer based on his/her level of latent trait, which is the functional form of a standard 2PL model. The second component indicated by $(1 - \Lambda[\beta_k(\theta_i - \alpha_k)])\, c_k$ is that a respondent answers the item correctly due to how likely the item is to be guessed, which is the value of c_k moderated by the probability of an incorrect response $1 - \Lambda[\beta_k(\theta_i - \alpha_k)]$. Furthermore, Equation (2) also shows that the greater the value of the first component, the smaller the value of the second component and, thus, the smaller the impact of guessing on the item (Andrich, Marais, and Humphry 2012, p. 427).

Considering the second component on the right-hand side of Equation (2) as the probability of guessing the correct answer, some research attempts to represent the guessing property of items by allowing individual differences and/or item characteristics in the functional form. For example, building on the Rasch model, San Martín, Del Pino, and De Boeck (2006) propose a one-parameter logistic model with ability-based guessing (1PL-AG), which is formulated as follows:

$$\Pr(y_{i,k} = 1|\theta_i, \alpha_k, \gamma_k, b) = \Lambda(\theta_i - \alpha_k) + [1 - \Lambda(\theta_i - \alpha_k)]\frac{\exp(b\theta_i + \gamma_k)}{1 + \exp(b\theta_i + \gamma_k)}, \tag{3}$$

where b is the general weight of the trait level in the guessing component and γ_k is termed "guessing" parameter corresponding to a respondent with an average trait level on the logistic scale.[10]

10 By the same token, Cao and Stokes (2008) work on the extension to the 2PL model to deal with the guessing components. From a different perspective, some research works on procedures for removing guessing in the estimation of item parameters and latent trait levels (Waller 1976, 1989; Andrich, Marais, and Humphry 2012; Andrich and Marais 2014).

According to San Martín, Del Pino, and De Boeck (2006), the 1PL-AG model differs from the 3PL model in several ways. First, a successful guess is related to latent traits and, thus, the guessing parameter c in Equation (2) is formulated as a function of latent traits. Second, a general parameter b is used to represent the weight of the ability in the guessing component and, thus, the discrimination parameter β_k is excluded. Finally, the parameter γ_k is used as the guessing parameter, which determines the probability of a successful guess when $b = 0$, that is, guessing is not ability related.

Considering a successful guess as a function of trait levels and item properties, we extend the 1PL-AG model to include the item-discrimination parameter β_k and make a modification with a replacement of γ_k by $-\alpha_k$. Moreover, we allow the weighting parameter to vary across items. As a result, the proposed model, termed the *two-parameter logistic guessing (2PL-G) model*, is specified as

$$\begin{aligned}\Pr(y_{i,k} = 1|\theta_i, \alpha_k, \beta_k, b_k) = {} & \Lambda[\beta_k(\theta_i - \alpha_k)] \\ & + (1 - \Lambda[\beta_k(\theta_i - \alpha_k)]) \left(\frac{\exp[b_k(\theta_i - \alpha_k)]}{1 + (M - 1)\exp[b_k(\theta_i - \alpha_k)]} \right), \end{aligned} \tag{4}$$

where M is the number of choice options for a multiple-choice item. Like Equation (3), the model presented by Equation (4) has two terms contributing to the probability of item success. The first term contributes to the success probability due to ability and the second term contributes to the success probability due to guessing.

The 2PL-G model differs from the 1PL-AG model in the following ways. First, we use the item-difficulty parameter α_k rather than γ_k in the guessing component because a successful guess is not only determined by ability but also by how hard an item is. Second, the weighting parameter varies across items to show differences between items, which is the main reason why IRT models are better than the classical test theory. Third, we include a constant M in the guessing component to reflect the fact that guessing is related to the number of choice options. Finally, the inclusion of M limits the highest probability of guessing to $1/(M - 1)$ because the probability of a successful guess higher than $1/(M - 1)$ reflects partial knowledge, which should be accounted for in the first component. Moreover, this setting also solves the interpretation problem suffered by the 1PL-AG model because both components depend on ability (San Martín, Del Pino, and De Boeck 2006, p. 187).[11]

3.2 Properties of the proposed 2PL-G model

Some properties of the proposed 2PL-G model, compared to the conventional 2PL and 3PL models, are discussed in this subsection. Following the literature on IRT, ICCs are displayed to describe how changes in the trait level relate to changes in the probability of a specified response. First, the inclusion of $M - 1$ in the guessing component of Equation (4) indicates that the asymptotic probability of a successfully random guess for $\theta \to -\infty$ is equal to $1/M$, given that $\alpha = 0$ and $b = 0$. For example, suppose that $M = 4$, that is, there are four options for a multiple-choice item. A respondent with the lowest trait level has a probability of 1/4 of randomly choosing the correct response ($b = 0$) to an item with average difficulty ($\alpha = 0$), which is represented by the black solid curve in the left panel of Figure 1.

Second, the weighting parameter b indicates how important the level of ability is for the probability of a successful guess for an item, given that item parameters α and β are fixed. The left panel of Figure 1 shows that, as the value of b increases, the probability of a successful guess for respondents with the lowest trait level decreases and the density becomes more concentrated

11 For the purpose of comparison, we also apply the 1PL-AG model to analyzing the ANES-EGSS4 data and the results are presented in Appendix G (see the Online Appendix in Supplementary Materials for details).

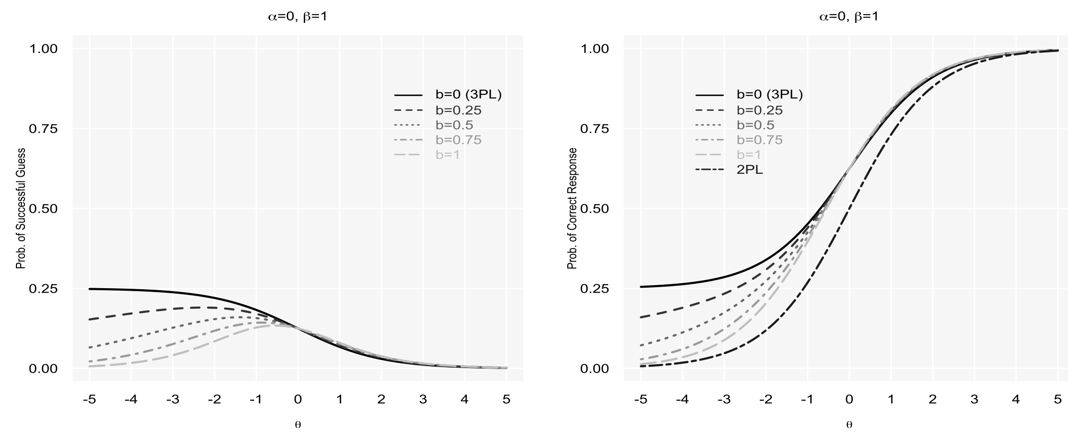

Figure 1. The probability of a successful guess and that of correct response for a four-category multiple-choice item. Item parameters α and β are fixed while b is varied.

around the average trait level. In the context of political knowledge, this property basically reflects the fact that both partially informed and barely informed respondents are more likely to guess than well-informed ones, but partially informed respondents are more likely to have a successful guess than barely informed ones. Barely informed respondents would make guesses with a low probability of being correct and the chance of guessing the correct answer is lower than randomly guessing because they are highly seducible by the attractive distractors. Furthermore, the reason why well-informed respondents have a relatively low probability of a successful guess is that their high trait levels play the major role in contributing to the probability of item success. Therefore, partially informed respondents have a higher probability of a successful guess than both barely informed and well-informed ones.

Third, related to the second property, respondents with relatively low trait levels are considerably affected by the weighting parameter b, in terms of the probability of item success. The right panel of Figure 1 shows the ICCs for different values of b given that $\alpha = 0$ and $\beta = 1$. As can be seen, the probability of item success for respondents with lower trait levels decreases when correct guessing requires high ability levels, which is reflected by the increase in the value of b. Moreover, no matter how large b is, the ICCs suggest that the success probability due to guessing always contributes to the probability of item success for partially informed respondents, compared to the ICC of a 2PL model displayed by the black two-dashed curve, which has no guessing component. In contrast, the probability of guessing contributes less to the probability of item success for well-informed respondents since they are knowledgeable enough to correctly answer these items.

Fourth, the black solid curve in Figure 1 is equivalent to a 3PL model with $c = 0.25$. In a traditional 3PL model, i.e., Equation (1), the lower asymptote theoretically lies between the interval 0 and 1 regardless of the number of options in a multiple-choice item and is an unknown parameter to be estimated. More importantly, it implies that the guessing component is an item property that applies to all respondents regardless of differences in item difficulty and levels of latent trait. In contrast, in the proposed 2PL-G model, the asymptotic probability is determined by both item characteristics (M and α) and levels of latent trait (depending on the magnitude of b), which makes the 3PL model a special case of the 2PL-G model.

Finally, the 2PL-G model allows not only differences in trait levels but also those in item difficulty in the guessing component. The left panel of Figure 2 shows how changes in item difficulty relate to changes in the probability of a successful guess, which suggests that, as items become more difficult, higher ability is required for respondents to have a successful guess. As can be seen, the effects of b and α on the probability of a successful guess are quite different but jointly

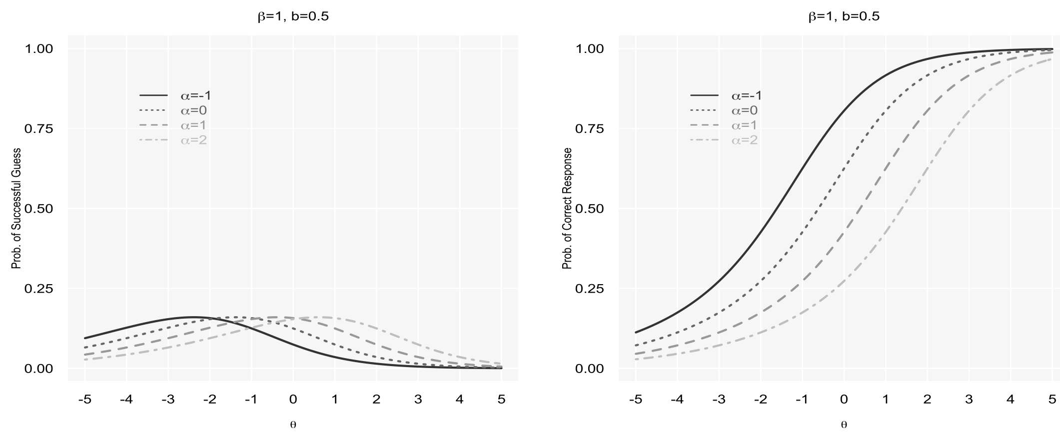

Figure 2. The probability of a successful guess and that of correct response for a four-category multiple-choice item. Item parameters b and β are fixed while α is varied.

determine the probability of a successful guess for an item. Moreover, the corresponding ICCs are shown in the right panel of Figure 2. We can see that, when $b \neq 0$, the probability of a correct response is slightly higher than 0.5 for a respondent with an average knowledge level ($\theta = 0$), given that $\alpha = 0$ and $\beta = 1$. The difference indicates the probability of a successful informed guess.

3.3 Prior distributions and identification

It is well known that the IRT models suffer from two identification problems: *scale invariance* and *rotational invariance* (Albert 1992; Johnson and Albert 1999). The problem of scale invariance occurs because the metric (location and scale) of the latent trait is only known up to a linear transformation. Therefore, one must anchor the metric of the latent traits. The problem of rotational invariance refers to the fact that, for the unidimensional case, multiplying all of the model parameters by -1 would not change the value of the likelihood function.

We estimate the model presented in Equation (4) by a Bayesian approach, so we complete the model specification by defining the prior distribution. In the Bayesian context, the use of informative prior distributions can resolve these two identification problems (Johnson and Albert 1999; Martin and Quinn 2002). First, in typical IRT models, latent variables are assumed to be sampled from a normal distribution with mean 0 and finite variance σ_θ^2, that is,

$$\theta_i \sim \mathrm{N}(0, \sigma_\theta^2), \quad \text{for } i = 1, \ldots, N. \tag{5}$$

To solve the problem of scale invariance, we can simply assume that $\sigma_\theta^2 = 1$ (Jackman 2009, p. 460).

Second, for item parameters β_k, α_k, and b_k from Equation (4), we assume that

$$\beta_k \sim \mathrm{N}(1, 2)\boldsymbol{I}(\beta_k > 0), \quad \text{for } k = 1, \ldots, K, \tag{6}$$
$$\alpha_k \sim \mathrm{N}(0, \sigma_\alpha^2), \quad \text{for } k = 1, \ldots, K, \tag{7}$$
$$b_k \sim \mathrm{Uniform}(0, 1), \quad \text{for } k = 1, \ldots, K, \tag{8}$$

where $\boldsymbol{I}(\cdot)$ denotes an indicator function and σ_α^2 follows the conjugate inverse gamma prior:

$$\sigma_\alpha^2 \sim \text{Inverse-Gamma}(0.01, 0.01). \tag{9}$$

The rotation invariance problem is solved by restricting item-discrimination parameters to be positive. This is because ICCs with positive item-discrimination parameters indicate that respondents answer test items correctly if they have higher ability. This form of constraint is sometimes known as "anchoring" (Skrondal and Rabe-Hesketh 2004, p. 66). The truncated normal prior with mean 1 and variance 2 for β_k reflects the fact that item-discrimination parameters usually take values between the interval 0.5 and 3 (Fox 2010, p. 21). The hyper prior Inverse-Gamma(0.01, 0.01) for σ^2_α is set to present noninformativeness. We assume a uniform prior between 0 and 1 for parameter b_k. If the estimates of b_k are different from 0, it shows evidence supporting the proposed model against the 3PL model, which implies informed guess rather than random guess.[12]

Increasingly, scholars agree that it is inappropriate to simply pool DKs and incorrect responses together as a single absence-of-knowledge category (Mondak and Davis 2001; Mondak and Anderson 2004; Miller and Orr 2008; Gibson and Caldeira 2009). Therefore, in the estimation process, DKs are treated as missing values rather than incorrect responses.

4 Data and Analysis

The analyzed dataset is survey data conducted by the TEDS project and included in Module 4 of the Comparative Studies of Electoral Systems (CSES): the presidential and legislative elections of 2012 (TEDS2012).[13] This dataset has 1,826 samples, collected by face-to-face interview after the 2012 election (from January to March), and includes three open-ended items and four multiple-choice items about political affairs in order to investigate the Taiwanese public's political knowledge. These knowledge questions are DK-neutral items, which makes the TEDS data opportune to investigate guessing components in the measurement of political knowledge. In the survey process, interviewers were instructed to accept but not to offer the DK option in advance. In other words, interviewers neither encouraged nor discouraged respondents to answer DK and just recorded respondents' answers including DKs. As discussed before, this format of political knowledge items, to some extent, excludes the effects of respondents' personality and leaves little knowledge within the DK responses. To show the generalizability of the proposed 2PL-G model, we also analyze the dataset from the American National Election Studies: Evaluations of Government and Society Study, Survey 4 (ANES-EGSS4) and present the results in Appendix A.

4.1 Descriptive statistics

Table 2 shows the seven political knowledge items and corresponding distributions of responses. Based on the distributions of responses, these items can be roughly classified into three groups: (1) *US President, Premier*, and *Second Legislative Party*; (2) *Constitutional Interpretation, Finance Minister*, and *Unemployment Rate*; and (3) *UN Secretary-General*. We can observe guessing behavior in the multiple-choice items. For example, the three items of the first group are relatively easy for survey respondents to answer. The percentages of correct responses for these three items are 75.85%, 63.14%, and 87.02%, respectively. Furthermore, comparing Second Legislative Party to US President and Premier, we observe a higher percentage of correct responses and a lower percentage of DKs in the multiple-choice item (Second Legislative Party) than in the open-ended ones (US President and Premier), which implies guessing behavior in the multiple-choice item.

12 In the data analysis presented in the next section, the priors of item parameters α and β and the latent variable θ for the 2PL and 3PL models have the same specification. Regarding the c parameter for the 3PL model, it is assumed to follow a uniform distribution between 0 and 1.

13 Data analyzed in this article are from "Taiwan's Election and Democratization Studies, 2009–2012 (III): The Survey of the Presidential and Legislative Elections, 2012 (TEDS2012) (NSC 100-2420-H002-030). The coordinator of the multi-year project is Professor Chi Huang (National Chengchi University). TEDS2012 is a yearly project on the election of presidency and legislators in 2012. The principal investigator of TEDS 2012 is Professor Yun-han Chu. More information can be found on the TEDS website (http://www.tedsnet.org). The authors appreciate the assistance in providing data by the institute and individuals aforementioned. The authors are alone responsible for views expressed herein. The replication materials can be found in Tsai and Lin (2017).

Table 2. The distributions of responses to knowledge items in TEDS2012.

Political knowledge items	Correct	Incorrect	"Don't Know"
1. Who is the current president of the United States?	75.85	2.14	22.01
2. Who is the current premier of our country?	63.14	8.87	27.99
3. What institution has the power to interpret the constitution?	28.81	27.77	43.42
4. Which of these persons was the finance minister before the recent election?	34.56	31.98	33.46
5. What was the current unemployment rate in Taiwan as of the end of last year (2011)?	33.68	33.30	33.02
6. Which party came in second in seat in the Legislative Yuan?	87.02	3.07	9.91
7. Who is the current Secretary-General of the United Nations?	18.67	27.44	53.89

Note: The first three knowledge items are open-ended while the other four are multiple-choice items. Row percentages are displayed.

Second, the percentages of correct responses for the three items of the second group are 28.81%, 34.56%, and 33.68%, respectively. These items are slightly hard for the public in Taiwan to answer. In other words, respondents require some extent of political knowledge to get these items correct. Table 2 shows that one third of the respondents choose the DK option in the two multiple-choice items (Finance Minister and Unemployment Rate), which suggests that these respondents may not know the answers. Furthermore, comparing Constitutional Interpretation with Finance Minister and Unemployment Rate, the proportion of DKs is a little higher in the open-ended item (Constitutional Interpretation) than that in the multiple-choice items (Finance Minister and Unemployment Rate). This result implies that some respondents choose DK in Constitutional Interpretation, but may guess the answers in Finance Minister and Unemployment Rate.

Finally, as shown in Table 2, only 18.67% of the respondents correctly answer the question about the current Secretary-General of the UN, which implies that this is a much harder question among these seven items. This highly difficult item also results in a high percentage of DKs (53.89%). Furthermore, incorrect responses are more than correct ones, which is the evidence of guessing behavior. It indicates that, for respondents who are uncertain about the answer, some of them choose the DK option and others choose the wrong answers since it requires higher levels of political knowledge to guess the correct one.

To illustrate (mis)informed guessing, in Table 3 we further display the distributions of choice options for the four multiple-choice items against the number of correct responses excluding the given item. Presumably respondents with relatively high levels of knowledge would have more correct responses and are more likely to select the correct option for a multiple-choice item and less likely to choose the DK option. For the purpose of illustration, we roughly classify respondents into three groups based on the number of correct responses (0 to 1, 2 to 4, and 5 to 6 correct responses). Consider Finance Minister as an example, first of all, for the first group in which respondents at most correctly answer one item among the remaining six items, there are only 3.99% of correct responses while there are 79.4% of respondents who choose DK. The percentage of correct responses rises to 31.45% for the second group and to 72.22% for the third group. In contrast, the percentage of DKs is 79.4% for the first group and goes down to 30.68% and 5.85% for the second and third groups, respectively. For the purpose of illustration, we temporarily

Table 3. The distributions of choice options against the number of item success in TEDS2012.

Fin. Minister No. of success	Jiang Yi-huah	Chen Chun	Mao Chi-kuo	Lee Sush-der (Correct)	"Don't Know"
0–1	6.98 (21)	6.31 (19)	3.32 (10)	3.99 (12)	79.40 (239)
2–4	11.07 (131)	17.58 (208)	9.21 (109)	31.45 (372)	30.68 (363)
5–6	7.60 (26)	12.28 (42)	2.05 (7)	72.22 (247)	5.85 (20)
Unemployment **No. of success**	**2.3%**	**4.3%** **(Correct)**	**6.3%**	**8.3%**	**"Don't Know"**
0–1	1.88 (6)	10.31 (33)	7.50 (24)	10.31 (33)	70.00 (224)
2–4	1.78 (20)	31.91 (359)	28.00 (315)	7.11 (80)	31.20 (351)
5–6	1.31 (5)	58.53 (223)	24.93 (95)	6.30 (24)	8.92 (34)
Second party **No. of success**	**KMT**	**DPP** **(Correct)**	**PFP**	**Non-Partisan** **Solidarity Union**	**"Don't Know"**
0–1	4.41 (23)	64.75 (338)	0.77 (4)	0.38 (2)	29.69 (155)
2–4	1.63 (17)	95.10 (989)	0.48 (5)	0.10 (1)	2.69 (28)
5–6	0.76 (2)	99.24 (262)	0.00 (0)	0.00 (0)	0.00 (0)
UN Secretary **No. of success**	**Kofi Annan**	**Kurt Waldheim**	**Ban Ki-Moon** **(Correct)**	**Boutros** **Boutros-Ghali**	**"Don't Know"**
0–1	2.31 (7)	4.29 (13)	4.29 (13)	0.99 (3)	88.12 (267)
2–4	20.96 (227)	6.00 (65)	12.74 (138)	1.85 (20)	58.45 (633)
5–6	34.32 (151)	1.59 (7)	43.18 (190)	0.68 (3)	20.23 (89)

Note: Row percentages are presented and the corresponding counts are in parentheses.

assume that the three groups represent uninformed/barely informed, moderately informed, and well-informed/fully informed respondents, respectively.

Table 3 shows that barely informed respondents do not randomly select an answer among the options, but misinformedly select wrong answers, except for the easiest item. For example, for the question of Finance Minister, the proportions of respondents with relatively low-level knowledge choosing one of the first two options are higher than that choosing the correct one. For Unemployment Rate and UN Secretary-General, although the correct answer is one of the highest chosen options, the other highest chosen one is an absurd answer. This evidence suggests that these respondents do not randomly but mistakenly choose the answer when they lack knowledge.

Table 3 also displays informed guessing for moderately informed and well-informed respondents. For example, about one fourth of moderately informed or well-informed respondents think the unemployment rate was 6.3% in 2011 instead of 2.3% or 8.3%. The rate 6.3% is a reasonable guess because unemployment was a rising issue and the unemployment rate has been between 4% and 6% since 2000 in Taiwan. Furthermore, for UN Secretary-General, numerous respondents can easily rule out two of the options (Kurt Waldheim and Boutros Boutros-Ghali). Some of them selected Kofi Annan as the answer to the question because they did not know Ban Ki-Moon succeeded Kofi Annan in 2007. This is the obvious evidence of informed guessing.

4.2 Results of analysis

We apply the 2PL-G model to analyzing the four multiple-choice items in the TEDS2012 dataset.[14] The model is estimated via Markov chain Monte Carlo (MCMC) methods (Albert and Chib 1993; Fox 2010) implemented in JAGS 4.2.0 (Plummer 2003) called from `R` version 3.3.1 (`R2jags`, Su and

14 For the purpose of comparison between different models, we only analyze multiple-choice items. In practical applications, however, all items—open-ended and multiple-choice items—should be analyzed in order to obtain more accurate estimates of knowledge levels and item parameters.

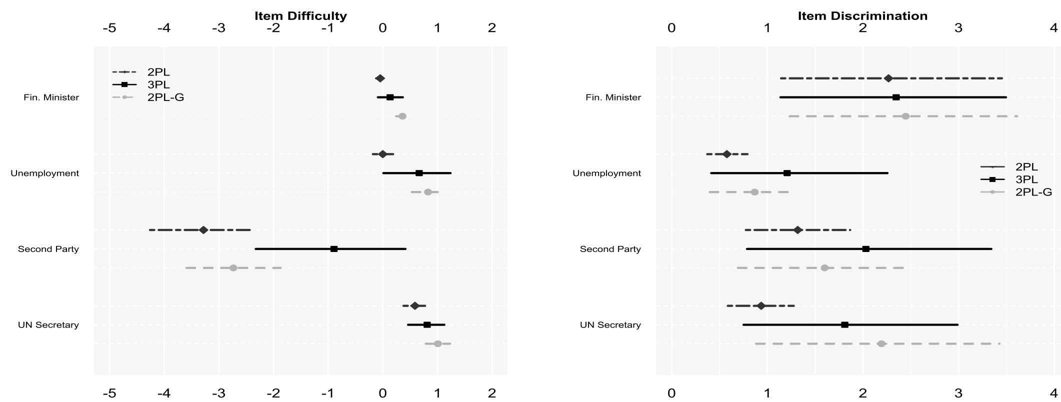

Figure 3. The 90% credible intervals of item parameters for the 2PL, 3PL, and 2PL-G models.

Yajima 2012).[15] Before we investigate the guessing components in the four multiple-choice items, we compare the estimates of item parameters from three models: the 2PL model, the 3PL model in Equation (1), and the proposed 2PL-G model in Equation (4).

Figure 3 shows the 90% credible intervals of item-difficulty parameters on the left and those of item-discrimination parameters on the right for the four multiple-choice, political knowledge questions. As can be seen in Figure 3, the left panel shows that Second Legislative Party has the least value of item difficulty among these items, which indicates that it is a relatively easy question. On the contrary, Finance Minister, Unemployment Rate, and UN Secretary-General are relatively hard questions. According to Figure 3, the 3PL model and the 2PL-G model have relatively larger estimates of item-difficulty parameters, particularly Finance Minister and Unemployment Rate, compared to the 2PL model. This result suggests that, without considering the guessing components, the 2PL model might underestimate difficulty parameters for multiple-choice items and reflects the fact that some of the multiple-choice items are actually harder if we take potential guessed responses into account.

Compared to the 2PL-G model, the 3PL model has larger estimates of difficulty parameters for easy items but has smaller estimates of difficulty parameters for difficult items due to parameter c. For easy items, the guessing components contribute too much to the probability of item success, which leads to a low probability of ability-based item success. Therefore, the difficulty parameters of easy items would be overestimated. By the same token, for hard questions, the guessing component does not contribute much to the probability of item success and, thus, a high probability of ability-based item success makes these items easier than they actually are. In contrast, the 2PL-G model does not suffer from this problem and provides more reasonable estimates of difficulty parameters when guessing behavior is taken into account.

The estimates of item-discrimination parameters are displayed in the right panel of Figure 3, which shows that the question Finance Minister performs well in discriminating different levels of political knowledge. Although the uncertainty of the item-discrimination parameter for Finance Minister is large, the three models show consistent results. In contrast, the estimates for the other three items are somewhat different. First, the 2PL model has less uncertainty, compared to the 3PL and 2PL-G models. Second, in terms of the means of the item-discrimination parameters, Second Legislative Party is the second best item in the 2PL and 3PL models while UN Secretary is the second best item in the 2PL-G model.

15 The estimation was performed with three parallel chains of 100,000 iterations each to be conservative. The first half of the iterations were discarded as a burn-in period and 10 as thinning and thus 15,000 samples were generated. The common diagnostic tests for the convergence of MCMC chains were conducted (Tsai and Gill 2012) and there is no evidence of nonconvergence in these chains. The results of convergence tests are presented in Appendix B (see the Online Appendix for details).

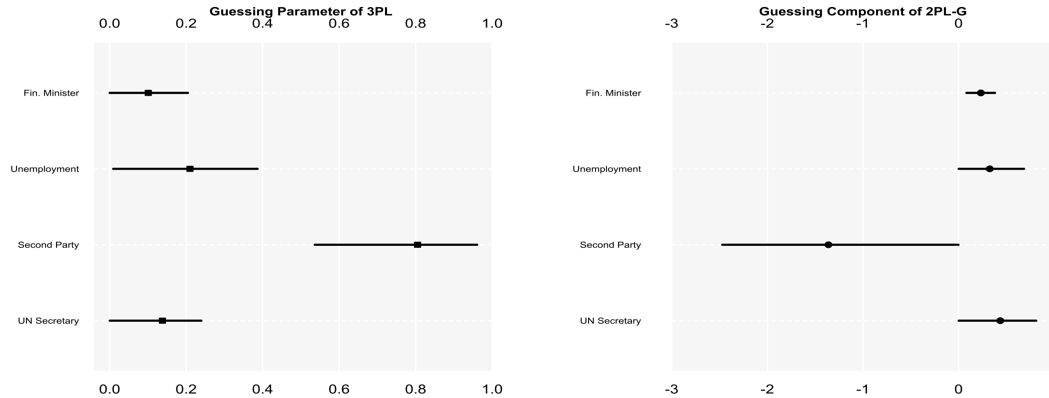

Figure 4. The 90% credible intervals of guessing effects for the 3PL and 2PL-G models.

The left panel of Figure 4 shows the estimates of parameter c_k for the 3PL model. The result suggests that Second Legislative Party is easy to guess for the uninformed and barely informed respondents. Regarding the 2PL-G model, because both b_k and α_k determine the guess components of multiple-choice items, we show the multiplication of the two parameters for each item.[16] Positive values of the multiplication mean that respondents need certain levels of knowledge to guess correctly while negative values indicate that respondents with low knowledge levels are able to have a successful guess. The result indicates that UN Secretary-General requires relatively high levels of political knowledge for respondents to correctly guess the answer.

Figure 5 displays the probability of a successful guess and the probability of a correct response across different levels of political knowledge from the 3PL model (on the top) and the 2PL-G model (on the bottom) for the four multiple-choice items. The two top panels suggest

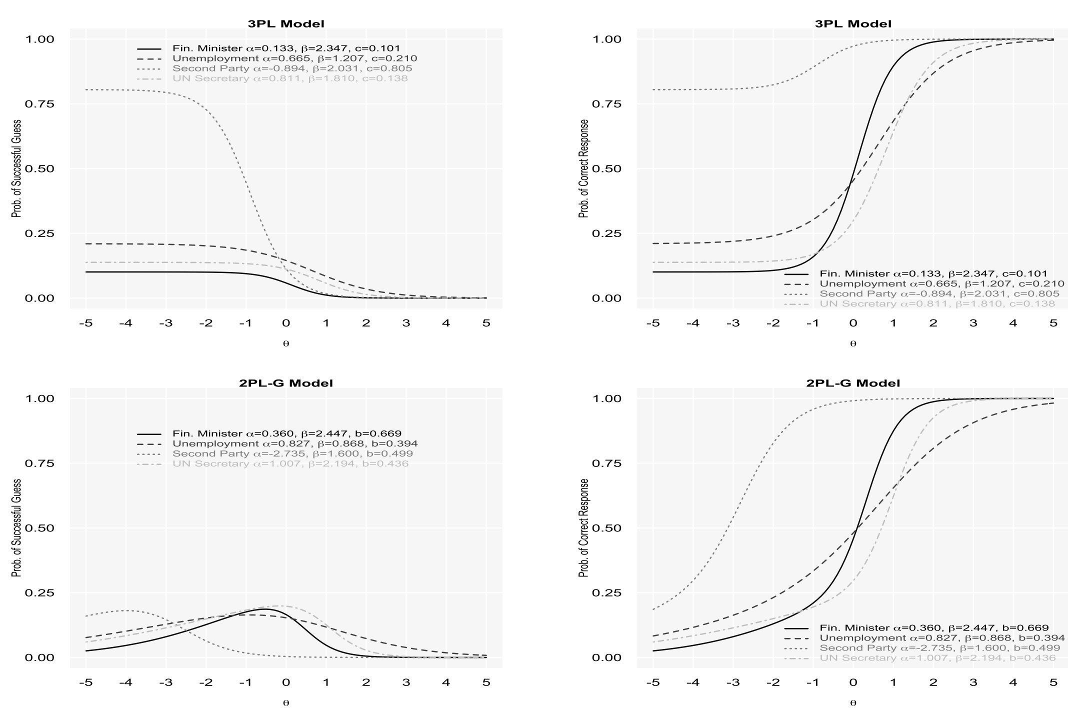

Figure 5. The probability of successful guesses and responses for the 3PL and 2PL-G models.

16 The estimates of b_k are presented in Appendix C (see the Online Appendix for details).

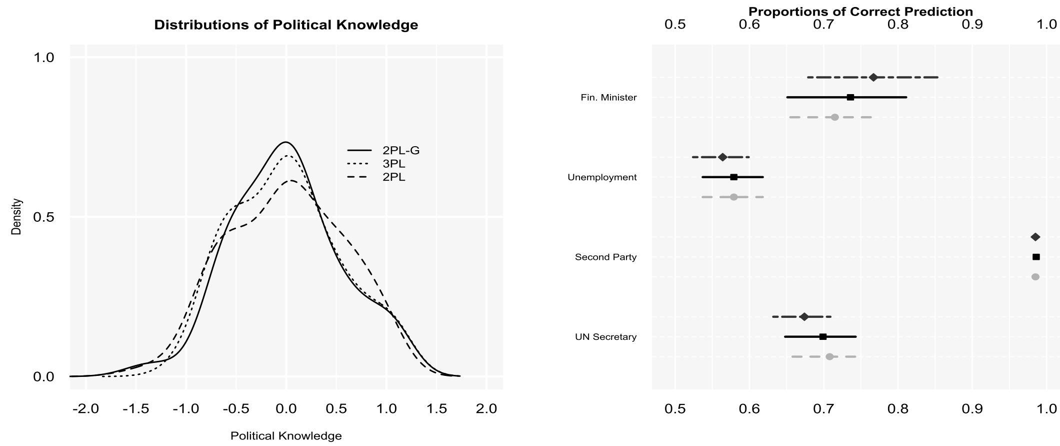

Figure 6. Comparisons between 2PL, 3PL, and 2PL-G. The left panel shows the distributions of political knowledge estimates. The right panel shows the 90% credible intervals of predictive accuracy.

that barely informed respondents have a higher probability to correctly guess the answer for multiple-choice items, compared to partially informed and well-informed respondents regardless of item difficulties. This setting is not supported by the real data shown in Table 3. Moreover, the two top panels demonstrate that the probability of a successful guess for uninformed respondents is higher than a random guess for Second Legislative Party and lower for the other three items. In contrast, the two bottom panels in Figure 5 illustrate an informed guess for items that require relatively high levels of political knowledge for respondents to guess the correct answers.

The results from the 2PL-G model are considerably consistent with the conventional wisdom about the status and amounts of Taiwanese information-obtaining channels which are associated with their levels of political knowledge. On the one hand, UN Secretary-General involves international affairs, which are relatively limited information in Taiwan's traditional and electronic media. Therefore, respondents have to be relatively knowledgeable to correctly guess or answer who the UN Secretary-General is.

On the other hand, Finance Minister, Unemployment Rate, and Second Legislative Party are questions about domestic politics. Second Legislative Party asks respondents which party came in the second in seats in the Legislature (Legislative Yuan). Information about party politics like this is more common in media reports and, therefore, citizens are more aware of party competition in both electoral and legislative arenas. So respondents with low-level knowledge are likely to guess/select the correct answer. Although Finance Minister and Unemployment Rate involve the fact of Taiwan's politics, the public in Taiwan is less aware of both. The finance minister is relatively unknown than the prime minister due to limited media exposure, which explains why large proportions of barely and partially informed respondents select the two prime ministers during the survey data collection, Jiang Yi-huah (the premier from 2/18/2013 to 12/8/2014) and Chen Chun (the premier from 2/6/2012 to 2/18/2013). With regard to Unemployment Rate, the public may have partial information from the media to understand domestic economic situation but not necessarily know the exact figure of the unemployment rate. Therefore, relatively high knowledge levels are required for respondents to correctly guess or answer Finance Minister and Unemployment Rate, but not Second Legislative Party.

To see how the guessing components affect the estimates of political knowledge, we show the distributions of estimated knowledge levels in the left panel of Figure 6. As can be seen, the 2PL-G model provides more estimates of average knowledge level. This is because, when we take informed guessing into account, some respondents who are considered to have moderate levels of knowledge by the 2PL model would be estimated downwards. In other words, these respondents may be overestimated because they correctly answer questions based on informed guessing rather than certain information. Moreover, the respondents with relatively low levels of knowledge would be overestimated by the 3PL model because a high value of parameter c leads to the overestimation of the easy item. These respondents would be estimated downwards by the 2PL-G model.[17]

Finally, we asses the fit of the three models to data by the deviance information criterion (DIC) and predictive accuracy. The DIC of the 2PL-G model is 8505.4, which is much smaller than that of the 2PL (15059.7) and 3PL (12066.0) models. The result indicates that the 2PL-G fits the data better than the other two conventional IRT models. Moreover, to assess the performance of the three models in predictive accuracy, we select respondents who have at least one correct response and at most three correct responses to the four multiple-choice items because, as shown in Table 1 and the left panel of Figure 6, the differences between the three models mainly lie in the respondents who are neither uninformed nor well-informed. The right panel of Figure 6 displays the 90% credible intervals of the proportions of correct prediction based on the 15,000 simulated samples from 1,493 respondents. As can be seen, the 2PL-G model is slightly better than the 2PL and 3PL models in Unemployment and UN Secretary, but not in Finance Minister, although there is no difference in terms of the overlapping of the 90% credible intervals.[18]

4.3 Differential item functioning by gender

Individual differences in guessing may be influenced by some aggregate-level variables such as gender, education, and partisanship. For instance, some research argues that, when answering multiple-choice, political knowledge questions, men are more likely to guess while women are more likely to say “DK” due to the differences in personality (Mondak and Anderson 2004; Lizotte and Sidman 2009). The differences in the propensity to guess might lead to the overestimation of men’s knowledge levels and underestimation of women’s knowledge levels, which induce the gender gap in political knowledge, because guessing increases the possibility of correct responses and DKs are conventionally treated as incorrect responses. The findings of this literature show that, when the differences in personality is taken into account, the gender gap in political knowledge decreases but does not disappear. In other words, even though the gender gap in political knowledge can be partially explained by the guessing effect, the finding that men are more knowledgeable than women is robust.

We find the results in the literature consistent with our argument, which is that a successful guess requires certain levels of knowledge. That is to say, if men could have a successful guess to multiple-choice items, that means they are partially knowledgeable to select the correct answers. To examine the gender gap in political knowledge in the TEDS2012 data, we first demonstrate the proportions of DK responses across gender in Table 4. As can be seen, the percentages of DK responses are higher for female respondents, which implies that women are less likely to guess. However, the differences in guessing propensity do not necessarily induce biased estimates if there is a trivial amount of knowledge concealed in DK responses (Sturgis, Allum, and Smith 2008) or if a successful guess relies on partial knowledge as we argue in this article.

The differences between men and women in the propensity to guess, if not accounted for, would lead to different estimates of item parameters for different groups, which is referred to as differential item functioning (DIF) (Lizotte and Sidman 2009). In other words, if the propensity to guess is accounted for, there should be little difference between groups in the parameter

17 These results can also be observed in the two scatterplots presented in Appendix D (see the Online Appendix).

18 The reason why the three models do not differ in the predictive accuracy in terms of the overlapping of the 90% credible intervals is that the range of estimated latent traits mostly lies between -1.5 and 1.5. In this range, the guessing components are not significantly different between the four items displayed in the bottom-left panel of Figure 5. We also do the same comparisons with the restrictions to item parameters and the results presented in Appendix E show that the 3PL and 2PL-G models are better than the 2PL model in terms of predictive accuracy.

Table 4. Percentages of DKs across gender in TEDS2012.

Item	Don't know	
	Men (921)	Women (905)
Fin. Minister	28.01 (258)	40.22 (364)
Unemployment	30.29 (279)	36.46 (330)
Second party	6.73 (62)	13.37 (121)
UN Secretary	47.67 (439)	60.77 (550)

Note: Percentages are presented and counts are in the parentheses.

estimates. Following this literature, we examine DIF by employing a multilevel framework to allow the item parameters to vary across gender.[19] A multilevel 2PL-G model is specified as follows:

$$\begin{aligned}\Pr(y_{i,k} = 1|\theta_i, \alpha_{j,k}, \beta_{j,k}, b_{j,k}) &= \Lambda[\beta_{j,k}(\theta_i - \alpha_{j,k})] \\ &+ (1 - \Lambda[\beta_{j,k}(\theta_i - \alpha_{j,k})]) \left(\frac{\exp[b_{j,k}(\theta_i - \alpha_{j,k})]}{1 + (M-1)\exp[b_{j,k}(\theta_i - \alpha_{j,k})]} \right). \qquad (10)\end{aligned}$$

And we assume that

$$\beta_{j,k} \sim \mathrm{N}(1,2)\boldsymbol{I}(\beta_{j,k} > 0), \quad \text{for } j = 1,2; k = 1,\ldots,K, \qquad (11)$$

$$\alpha_{j,k} \sim \mathrm{N}(0,\sigma_\alpha^2), \quad \text{for } j = 1,2; k = 1,\ldots,K, \qquad (12)$$

$$b_{j,k} \sim \text{Uniform}(0,1), \quad \text{for } j = 1,2; k = 1,\ldots,K, \qquad (13)$$

where $j = 1$ indicates male respondents and $j = 2$ female respondents. For the identification of the multilevel 2PL-G model, the item parameters of Finance Minister are held constant, that is, $\alpha_{1,1} = \alpha_{2,1} = -3$ and $\beta_{1,1} = \beta_{2,1} = 1.5$. For the purpose of comparison, we also estimate a multilevel 3PL model.

First, we calculate the differences between women and men in the mean knowledge scores. The 90% credible intervals of the mean differences for the 3PL and 2PL-G models are [0.016, 0.163] and [0.008, 0.156], respectively. The results are consistent with the findings in the literature, which indicates that the gender gap in political knowledge still exists even after the guessing effect is accounted for.

Next, we compare the 3PL and 2PL-G models by showing the estimates of item parameters in Figure 7. Two differences between these two models can be observed in the left panel of Figure 7. First, the 3PL model provides similar estimates of Unemployment for men and women, but the 2PL-G model provides different item-difficulty estimates between men and women. Second, in terms of the means, the 3PL model shows that Second Legislative Party is easier for men than for women while the 2PL-G model shows the opposite although they are not different with the credible intervals overlapped.

To see which model fits the data better, we present in Table 5 the distributions of correct and incorrect responses to the three unconstrained multiple-choice items against the number of item

19 Unlike Lizotte and Sidman (2009), we consider DK responses as missing values rather than incorrect responses to avoid the underestimation of women's political knowledge levels.

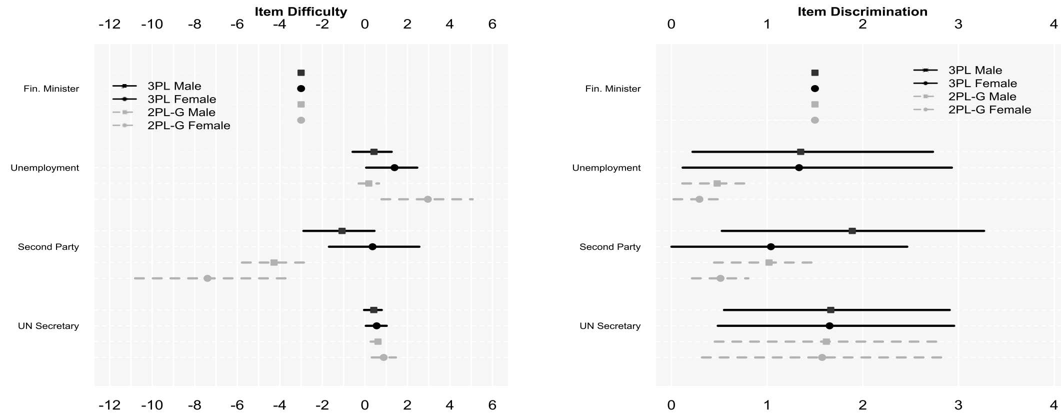

Figure 7. The 90% credible intervals of item parameters for the multilevel 3PL and 2PL-G models.

success for both men and women.[20] It is obvious that, for Unemployment, men are more likely to answer correctly than women at any level of political knowledge. Moreover, for Second Legislative Party, women have a higher rate of correct responses than men given that the number of success is 0 (96.02% > 93.81%). We can see that the 2PL-G model captures these facts, which is the evidence that the 2PL-G model performs better than the 3PL model.

Table 5. Distributions of responses across gender in TEDS2012.

Unemployment	**Men**		**Women**	
No. of success	**Correct**	**Incorrect**	**Correct**	**Incorrect**
0	60.00 (12)	40.00 (8)	26.92 (7)	73.08 (19)
1	45.61 (109)	54.39 (130)	39.62 (124)	60.38 (189)
2	59.27 (147)	40.73 (101)	45.66 (79)	54.34 (94)
3	74.07 (100)	25.93 (35)	58.73 (37)	41.27 (26)
Second party	**Men**		**Women**	
No. of success	**Correct**	**Incorrect**	**Correct**	**Incorrect**
0	93.81 (273)	6.19 (18)	96.01 (361)	3.99 (15)
1	97.38 (260)	2.62 (7)	96.90 (250)	3.10 (8)
2	99.00 (198)	1.00 (2)	98.21 (110)	1.79 (2)
3	99.01 (100)	0.09 (1)	97.37 (37)	2.63 (1)
UN Secretary	**Men**		**Women**	
No. of success	**Correct**	**Incorrect**	**Correct**	**Incorrect**
0	16.67 (1)	83.33 (5)	58.33 (7)	41.67 (5)
1	24.53 (26)	75.47 (80)	24.81 (33)	75.19 (100)
2	42.78 (83)	57.22 (111)	38.03 (54)	61.97 (88)
3	56.82 (100)	43.18 (76)	54.41 (37)	45.59 (31)

Note: DKs are excluded.

5 Conclusion

In the discipline of political science, the levels of political knowledge are measured by different item formats, including multiple-choice questions and open-ended identification items. The former are very likely to induce guessing behavior, which is associated with personal tendency

20 The distributions of choice options across gender are presented in Appendix F (see the Online Appendix for details).

to guess and blind guessing, and the latter usually lead to high proportions of DK responses, both of which affect estimates of political knowledge. Some research claims that multiple-choice items with the absence of the DK options effectively exclude respondents' guessing propensity and motivate partially informed guessing (Mondak 1999, 2001; Miller and Orr 2008; Luskin and Bullock 2011). Because it is impossible to eliminate all guessing (Mondak and Davis 2001), what is important is to clarify the association between guessing behavior and knowledge levels. There have been, however, very few studies conducted to investigate informed guessing in the measurement of political knowledge.

In this article, we examine guessing behavior by focusing on multiple-choice items with the DK-neutral condition and argue that informed guessing is more common than blind guessing. To deal with the issue of informed guessing, we propose a two-parameter logistic guessing model to accommodate the guessing components of successful responses based on respondents' levels of political knowledge and item difficulties. In particular, partially informed respondents are more likely to have a successful guess than both barely informed and well-informed ones. This is because barely informed respondents are highly seducible by the attractive distractors while well-informed ones have enough knowledge to correctly answer questions.

The proposed 2PL-G model is applied to analyzing survey data conducted in 2012 in Taiwan. The results show that the proposed model appropriately describes the characteristics of the political knowledge items, that is, a successful guess requires certain levels of knowledge depending on the difficulties of items. This article contributes to our understanding of guessing behavior in multiple-choice items. The evidence shows that informed (and/or misinformed) guessing is actually more common than blind guessing. We also examine the gender gap in political knowledge and the findings imply that men are more likely to guess than women because men are in general more knowledgeable than women. While the focus of this article is on the measurement of political knowledge, the proposed model can be applied to multiple-choice items for measuring other types of knowledge in surveys.

The results of the empirical analysis show that some problems appearing in the 2PL and 3PL models are fixed by the 2PL-G model. First, the 2PL model underestimates item-difficulty parameters due to the ignorance of the guessing effect, which leads to the overestimation of knowledge levels. Second, although the 3PL model accounts for the guessing effect, it overestimates the item-difficulty parameters of easy items while underestimating the item-difficulty parameters of hard items due to the implausible assumption of the probability of a successful guess through parameter c. Because the overestimated item-difficulty parameters of easy items and underestimated item-difficulty parameters of hard items influence the estimates of knowledge levels in different directions, the overall effects are uncertain. Therefore, we observe that the 2PL-G model is better than the 2PL model but is indifferent from the 3PL model in the analysis without considering DIF. The better performance of the 2PL-G model than the 3PL model can be observed in the analysis which takes DIF into consideration.

Supplementary Materials

For supplementary materials accompanying this paper, please visit https://doi.org/10.1017/pan.2017.21.

References

Albert, James H. 1992. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics* 17(3):251–269.

Albert, James H., and Siddhartha Chib. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(442):669–679.

Andrich, David, and Ida Marais. 2014. Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple choice items. *Applied Psychological Measurement* 38(6):432–449.

Andrich, David, Ida Marais, and Stephen Humphry. 2012. Using a theorem by Andersen and the Dichotomous Rasch model to assess the presence of random guessing in multiple choice items. *Journal of Educational and Behavioral Statistics* 37(3):417–442.

Baker, Frank B., and Seock-Ho Kim. 2004. *Item response theory: Parameter estimation techniques*. 2 ed. Boca Raton, FL: Chapman & Hall/CRC.

Birnbaum, Allan. 1968. Some latent trait models and their use in inferring an examinee's ability. In *Statistical theories of mental test scores*, ed. M. Lord Frederic and Melvin R. Novick. Reading, MA: Addison-Wesley, pp. 397–479.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E Stokes. 1960. *The American voter*. New York: John Wiley.

Cao, Jing, and S Lynne Stokes. 2008. Bayesian IRT guessing models for partial guessing behaviors. *Psychometrika* 73(2):209–230.

Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review* 98(2):355–370.

Delli Carpini, Michael X., and Scott Keeter. 1991. Stability and change in the US public's knowledge of politics. *Public Opinion Quarterly* 55(4):583–612.

Delli Carpini, Michael X., and Scott Keeter. 1993. Measuring political knowledge: Putting first things first. *American Journal of Political Science* 37(4):1179–1206.

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

Embretson, Susan E., and Steven P. Reise. 2000. *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fox, Jean-Paul. 2010. *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.

Galston, William A. 2001. Political knowledge, political engagement, and civic education. *Annual Review of Political Science* 4(1):217–234.

Gibson, James L., and Gregory A. Caldeira. 2009. Knowing the Supreme Court? A reconsideration of public ignorance of the High Court. *The Journal of Politics* 71(2):429–441.

Hambleton, Ronald K., and Linda L. Cook. 1977. Latent trait models and their use in the analysis of education test data. *Journal of Educational Measurement* 14(2):75–96.

Jackman, Simon. 2000. Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation. *Political Analysis* 8(4):307–332.

Jackman, Simon. 2001. Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. *Political Analysis* 9(3):227.

Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. Chichester, UK: Wiley.

Jessee, Stephen A. 2015. "Don't know" responses, personality, and the measurement of political knowledge. *Political Science Research and Method*, 1–21. doi:10.1017/psrm.2015.23.

Johnson, Valen E., and James H. Albert. 1999. *Ordinal data modeling*. New York, NY: Springer.

Lassen, David Dreyer. 2005. The effect of information on voter turnout: Evidence from a natural experiment. *American Journal of Political Science* 49(1):103–118.

Lizotte, Mary-Kate, and Andrew H. Sidman. 2009. Explaining the gender gap in political knowledge. *Politics & Gender* 5(2):127–151.

Lord, Frederic M. 1968. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement* 28(4):989–1020.

Lord, Frederic M. 1970. Item characteristic curves estimated without knowledge of their mathematical form–a confrontation of Birnbaum's logistic model. *Psychometrika* 35(1):43–50.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Luskin, Robert C. 1987. Measuring political sophistication. *American Journal of Political Science* 31(4):856–899.

Luskin, Robert C. 1990. Explaining political sophistication. *Political Behavior* 12(4):331–361.

Luskin, Robert C., and John G. Bullock. 2011. "Don't know" means "don't know": DK responses and the public's level of political knowledge. *Journal of Politics* 73(2):547–557.

Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. *Political Analysis* 10(2):134–152.

Miller, Melissa K., and Shannon K. Orr. 2008. Experimenting with a "third way" in political knowledge estimation. *Public Opinion Quarterly* 72(4):768–780.

Mondak, Jeffery J. 1999. Reconsidering the measurement of political knowledge. *Political Analysis* 8(1):57–82.

Mondak, Jeffery J. 2001. Developing valid knowledge scales. *American Journal of Political Science* 45(1):224–238.

Mondak, Jeffery J. 2010. *Personality and the foundations of political behavior*. New York, NY: Cambridge University Press.

Mondak, Jeffery J., and Belinda Creel Davis. 2001. Asked and answered: Knowledge levels when we will not take "don't know" for an answer. *Political Behavior* 23(3):199–224.
Mondak, Jeffery J., and Karen D. Halperin. 2008. A framework for the study of personality and political behaviour. *British Journal of Political Science* 38(2):335–362.
Mondak, Jeffery J., and Mary R. Anderson. 2004. The knowledge gap: A reexamination of gender-based differences in political knowledge. *Journal of Politics* 66(2):492–512.
Pietryka, Matthew T., and Randall C. MacIntosh. 2013. An analysis of ANES items and their use in the construction of political knowledge scales. *Political Analysis* 21(4):407–429.
Plummer, Martyn. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, ed. Kurt Hornik, Friedrich Leisch, and Achim Zeileis. March 20, Vienna, Austria. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.
Prior, Markus, and Arthur Lupia. 2008. Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science* 52(1):169–183.
Rasch, Georg. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute for Educational Research.
San Martín, Ernesto, Del Pino Guido, and Paul De Boeck. 2006. IRT models for ability-based guessing. *Applied Psychological Measurement* 30(3):183–203.
Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
Sturgis, Patrick, Nick Allum, and Patten Smith. 2008. An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly* 72(1):90–102.
Su, Yu-Sung, and Masanao Yajima. 2012. R2jags: A package for running JAGS from R. R Package ver. 0.03-08. http://cran.r-project.org/web/packages/R2jags/.
Tedin, Kent L., and Richard W. Murray. 1979. Public awareness of congressional representatives: Recall versus recognition. *American Politics Research* 7(4):509–517.
Treier, Shawn, and Simon Jackman. 2008. Democracy as a latent variable. *American Journal of Political Science* 52(1):201–217.
Tsai, Tsung-han, and Chang-chih Lin. 2017. Replication data for: Modeling guessing components in the measurement of political knowledge. doi:10.7910/DVN/80WUQB, Harvard Database, V1, UNF:6:OLHx9ShFd1/kZz7kxPYXZ==.
Tsai, Tsung-han, and Jeff Gill. 2012. Superdiag: A comprehensive test suite for markov chain non-convergence. *The Political Methodologist* 19(2):12–18.
Waller, Michael I. 1976. *Estimating parameters in the Rasch model: Removing the effects of random guessing*. Princeton, NJ: Educational Testing Service.
Waller, Michael I. 1989. Modeling guessing behavior: A comparison of two IRT models. *Applied Psychological Measurement* 13(3):233–243.