PICK YOUR POISON: ON SELECTING ROLL CALLS TO SAY WHATEVER YOU WANT

by

ALICE E. KISAALITA

(Under the Direction of Michael S. Lynch)

ABSTRACT

This dissertation takes up the question of how interest group ratings contribute to our understanding of legislator ideology. By judiciously picking which roll calls are included in our estimation techniques, we can gain a more accurate perspective of how legislators line up on the liberal to conservative or other issue-specific scale. This behavior becomes problematic when the standard for distinguishing the "important" roll calls from the multitude of available votes rests on biased, partisan, or otherwise unsound grounds. After fleshing out the consequences of the biased and/or non-standardized subsetting behaviors of interest groups (both general and special-issue) during the creation of interest group ratings, I consider a mathematical standard for creating subsets of ideologically relevant roll calls by which we can efficiently order legislators along a liberal-conservative line. Through an examination of the ADA, ACU, and NEA over the course of more than thirty years, I find that the selection habits of interest groups for which roll calls to include in their legislator ratings actually serve as a mirror into their own political motivations. Additionally, I analyze the consequences of subsetting the roll call record into smaller sizes for the purposes of ideal point estimation in general. I find that a small subset does not necessarily produce inconsistent estimations, but certain kinds of votes can substantially alter the resulting ideal points relative to NOMINATE scores.

INDEX WORDS: Interest Group Ratings, Legislator Ideology, Roll Call Votes, Ideal Point Estimation, Item Response Theory, US Congress

PICK YOUR POISON: ON SELECTING ROLL CALLS TO SAY WHATEVER YOU WANT

by

ALICE E. KISAALITA

B.A., The University of West Georgia, 2011

B.S., The University of West Georgia, 2014

M.A., The University of Georgia, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

PICK YOUR POISON: ON SELECTING ROLL CALLS TO SAY WHATEVER YOU WANT

by

ALICE E. KISAALITA

Major Professor: Michael S. Lynch

Committee: Scott H. Ainsworth
Anthony J. Madonna

Electronic Version Approved:
Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

ACKNOWLEDGEMENTS

No one earns a doctorate degree on their own, even if only one or two others show them support along the way. I am so lucky to have so many people in my life who have encouraged, supported, advised, and challenged me throughout my pursuit of this degree. I would like to sincerely thank each and every person who has helped me reach this goal. Several people deserve special acknowledgement and thanks in this regard.

I can honestly say that I would not be where I am today without the guidance, encouragement, and friendship of my major advisor, Dr. Michael S. Lynch. He has been a pillar of support for me since the very beginning of my graduate studies. He taught me to enjoy describing the world with statistics, probability, and uncertainty when at first I was sure I only liked "real" math like algebra and calculus. He challenged me to expand and develop my computer programming skills even though there were plenty of easier roads I could have taken to get the job done without improving myself. He encouraged me to go to EITM, he has included me on several working projects that have kept me interested in political science and enhanced my marketable skills immeasurably, and he showed me how to teach and how to conduct compelling research. Without Michael, this dissertation would not exist. As my mentor and guide in academia he has nurtured my innate desire for knowledge, and I have flourished under his care.

But Michael has been more than an academic mentor and guide to me. I speak with Michael about politics, economics, mathematics, education, and philosophy, but I also rely on him for advice on life in general. He took me seriously when I questioned whether to continue with the Ph.D. or stop at the Master's degree. He provided me with exceptional resources to help me decide between being a professor or taking a private sector job. He was there for me when I got married, when my mother fell ill, when I moved to Texas in the middle of my degree, and for everything else that happened over the last five years. He has always shown respect for my thoughts and concerns

on every matter I brought to him, and I count him among my greatest friends. Michael, it is hard for me to put in to words how much your guidance, care, and support has meant to me over these five years. I am truly, sincerely, and forever grateful for the time and attention you have given to me, for the unending support you have shown me, for the excellent discussions we carry on, for the knowledge you have imparted to me, for the confidence you place in my abilities, and for our friendship. Thank you.

Dr. Scott H. Ainsworth epitomizes what it means to be a professor. He exudes wisdom, thoughtfulness, consideration, and curiosity in all he does, and he serves as an exemplar for me when I think of the kind of person I want to be as I continue to grow and learn what it means to be human. Scott taught me that being a political scientist involves describing people's collective and individual behaviors philosophically, and I very much enjoyed our discussions about which philosophical frameworks apply in which contexts. He is willing to engage with me in conversations that get to the heart of these quandaries and provide a perspective filled with experience and expertise. Scott, I sincerely thank you for always making yourself available to me for advice, guidance, and interesting discussions. You always have a set of authors ready-to-hand for me to read, a follow-up thought from our previous conversations, or a new perspective from which I can learn. You have my utmost respect, and I am truly grateful for your part in my intellectual development.

Dr. Anthony J. Madonna showed me what it means to truly love your work and to be so passionate about a subject that every detail inspires and excites as you learn more about it. His attention to the minutia within the history and proceedings of Congress continually inspires me, and I plan to emulate his example and work ethic wherever I go. Thank you, Tony, for always being there to listen and give advice to me as I continually questioned my place in the world and where I am going. You helped me realize that there is not only one way to define success, and you always encouraged me to find my way towards happiness as I define it for myself.

Two other faculty deserve a special mention here. Dr. Janet Donohoe taught me philosophy at the University of West Georgia, and she was my primary advisor throughout my first and still favorite degree. She is also one of the reasons I aspired to go this far in academia. Dr. Donohoe

showed me just what it means to have a love of wisdom, and I am so truly grateful to her for her part in my growth as a thinker. Secondly, I would like to thank Dr. Keith Poole, with whom I had the honor of working at the University of Georgia. His work laid the groundwork for my dissertation, and I am grateful for the time and advice he gave to me.

I think it might be impossible to fully quantify the ways in which a husband supports his wife in the pursuit of her Ph.D. Ssempa Kisaalita has been there for me every day along the path to this degree, showing me unending love and support as I worked through classes, homework, studying for comprehensive exams, writing and defending the prospectus and the chapters that follow. He keeps me motivated when I feel like quitting, he walks through complex ideas with me, he listens to my problems and offers knowing advice, he believes in me even when I lose faith in my own abilities, and he makes amazing dinners every night while I work. He is a quiet, loving, intelligent, hard-working, dedicated, and diligent partner whom I can always look to for love, support, and encouragement. This dissertation would not exist without him. Thank you so much, Ssempa, for all this and more. I hope that I give you the same support for your dreams and goals as you have shown for mine.

Of the many people in my life whom I look up to, my mother is the first and most important inspiration to me. Dr. Lesley Barker has shown me what it truly means to be a strong, intelligent, admirable, independent, and thoughtful woman, and I am proud to be her daughter. Not only did she instill in me an insatiable desire to learn, she taught me how to do so humbly, so that I can gain insights without presuming to know everything. She showed me the value of cultivating an eclectic set of perspectives by which to evaluate the arguments presented by the world. She helped me think through the work I did for this dissertation and gave me invaluable feedback on the ideas herein. All my life she has encouraged me in my pursuit of knowledge, spurring me on and on to ever greater academic accomplishments as she herself followed a similar path simultaneously. My mother is an amazing woman, and I will be forever grateful to her for raising me to be the person I am today. Thank you, Mom. You are my role model and my inspiration.

I am lucky to have two sets of parents whom I can look to for love, support, and encouragement. My parents-in-law, Dr. William and Rose Kisaalita, have welcomed me into their family as if I were one of their own. I am so grateful to them for showing me such love and support in all aspects of life, including the pursuit of my Ph.D. They have read chapters, discussed ideas, given me advice, and helped me think through how I will use this degree after it is finished, and they have done so with genuine care for my well-being. Thank you both for being there for me along this journey. I cannot tell you how much your love and support means to me.

I want to thank my many siblings for the myriad ways in which you have all been there for me as I completed this degree. Esther, Nancy, Audrey, Lottie, Colin, Roger, Ntumwa, Nkaku, Nami, Aditya, Jeff, Dave, Joe, Whitney, Dani, and Erik, you are each so important to me and I love you all. You have done so much for me throughout my life, and I thank you one and all for loving me, supporting me, laughing with me, and being there for me as I accomplish this dream of mine.

Finally, I thank my friends, Simon McClelland-Williamson, Jordan McKissick, Sam Marcotte, Matthew Jordan Suwalski, and Haley Lattke, who have been there for me through thick and thin. Simon, you are the first friend I made at UGA, and I cannot imagine going through graduate school without you. From math camp to comprehensive exams, knitting to crocheting, and everything in between, it is my pleasure to know you. Jordan, you and I make a great team, whether it is studying together or co-authoring. I know you will be successful in whatever you choose to do. Sam, we study different parts of political science, but I always found your perspective to provide an interesting take on things that I would not have heard otherwise. Know that I appreciate you, and I am glad to have you as a friend. Matthew, you are a consistent delight to have in my life, and I am so glad you came to graduate school at UGA. Life there would have been much duller without you. Haley, you are the sweetest, gentlest person I know, and I wish you nothing but happiness for your life. You brought a softness to graduate school that made the roughest parts easier to handle. To each of you, thank you for spending your time with me, for growing and learning with me, and for being such good friends. Without your encouragement, I do not think I would be where I am

today. As I said, no one can really do this alone, and I am so glad that I gained your friendship as we went through graduate school together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

On April 6, 2017 three roll call votes took place in the Senate that paved the way for Justice Neil Gorsuch's confirmation to the Supreme Court the following day. The first was an unsuccessful attempt to invoke cloture[1]. The motion was defeated with 55 yeas and 45 nays. Notably, on this vote four Democratic senators (Heidi Heitkamp (D - ND), Joe Manchin (D - WV), Joe Donnelly (D - IN), and, surprisingly, Michael Bennet (D - CO)[2]) voted with the rest of the Republicans to invoke cloture, and one Republican (Senate Majority Leader Mitch McConnell) voted strategically against this motion so that he could bring it up again later[3].

The second roll call in this series of confirmation-related events changed the rules of cloture so that a simple majority, rather than the full 3/5ths majority of 60 votes, is now required to invoke cloture and stop debate for confirmation of presidential nominations. This rule change implemented what has become known as the "nuclear option" in the popular vernacular because it upended a centuries-long tradition in the Senate of respecting the right of the minority party to filibuster judicial nominations to the Supreme Court. It is important to note that this is not the first time that the nuclear option has been used. In November of 2013, Senate Democrats led by Majority Leader Harry Reid imposed this same rule change for most judicial nominations, with the notable exception of Supreme Court nominations. The nuclear option rule change that was implemented

[1]Invoking cloture means to stop debate, which, under the rules of the Senate, required a 3/5ths majority (60 votes) to proceed.

[2]This was the only vote on which Senator Bennet defected from the majority of his party, and it can be explained by noting that Justice Gorsuch is from Colorado. In a statement on FaceBook, Senator Bennet explained his position, saying, "I am proud Judge Gorsuch is from Colorado...He is a qualified judge who deserves an up-or-down vote. That is the tradition of the Senate, and it is why I opposed a filibuster before the rule change"(Meltzer (2017)).

[3]Under the rules of the Senate, a motion to reconsider a previously taken vote must be made by a member who voted with the prevailing side (in this case, the nays prevailed). So, Senator McConnell, alone among the Republicans, voted nay so that he could have the right to bring up the same question again later.

for Neil Gorsuch's nomination simply extended these changes so that all presidential nominations, including those for the Supreme Court, could now be confirmed by a simple majority rule vote. This second roll call changed the rules of the Senate by a straight party-line vote of 52 Republicans against 48 Democrats with no defections from either party.

Once the rules were changed, a third vote took place to invoke cloture under the new simple-majority rules. While this vote received the exact same vote total of 55 yeas and 45 nays, this time the result was that cloture was invoked, the minority party filibuster ended, and the up or down confirmation vote could take place on the following day. The second cloture vote under the new rules was approved with the support of every single Republican senator (this time Sen. McConnell voted with his party) as well as three Democratic senators (Heidi Heitkamp (D - ND), Joe Manchin (D - WV), and Joe Donnelly (D - IN)) each of whom represents a state that leans Republican. These three Democratic senators voted again with the Republicans on the actual confirmation vote on the following day, and Neil Gorsuch was confirmed as a Justice of the Supreme Court with a 54-45 roll call vote on April 7, 2017.

All four of the votes detailed above were necessary in order to confirm Justice Gorsuch, but only one of them, the straight party-line vote to impose the nuclear option rule change, was used in the interest group ratings created by the American Conservative Union (ACU). The ACU creates a yearly rating of US legislators in order to assess and display how liberal or conservative a particular legislator is based on a subset (usually 25) of their roll call votes. These ratings purportedly "reflect how elected officials view the role of government in an individual's life," and they are "designed to educate the public about how consistently their elected officials adhere to conservatism" (Letter from the Chairman, ACU 2017 Ratings of Congress). These ratings are often cited by prominent politicians in Congress, and they are frequently used to demonstrate to the voting public a commitment to the conservative agenda. A high rating offers conservative members of Congress bragging rights to the public, and a low rating can be used as electoral fodder for primary challengers. The ACU prides itself on selecting roll call votes that provide important distinctions among legislators and give a broad picture of the ideological leanings of Congress members from

a small set of votes. "The votes selected by the ACU are not always considered the 'most important' votes; instead, the selected votes we score help reflect the ideological distinctions among our elected officials" (Selecting the Votes, ACU 2017 Ratings of Congress).

With their stated goals in mind, however, it is useful to see exactly where the ACU places legislators like Senators Heitkamp, Manchin, and Donnelly, all three of whom voted with the Republicans on three out of the four votes related to the confirmation of Neil Gorsuch. Given that voting behavior, it might be reasonable to expect that the ACU conveys this information to the public and portrays these three as moderates who are sometimes willing to work across the aisle. Instead, in their ratings for 2017, the ACU placed all three senators in the "Coalition of the Radical Left," a category reserved for legislators who receive scores less than or equal to 10% agreement with conservative principles. Both Donnelly and Heitkamp received scores for the year of 0/21 on votes which are supposed to show the range of ideological distinctions among legislators, while Manchin voted only twice with the conservative positions chosen by the ACU and received an 8% rating for the year. In fact, with his 8% conservative rating, Senator Manchin achieved the highest (i.e., most conservative) rank among the Democrats and Independents in the Senate that year. By contrast, the lowest ranking Republican (Susan Collins (R-ME)) had a score of 48%.

Clearly there is some disconnect between the stated goals of the ACU and the actual ratings that result from their selection of important or ideologically distinctive votes. There is a 40-point gap in the ideological spectrum in which no senators are classified (between 8% and 48%), which, unless the Democrats in the Senate really are completely devoid of conservative principles, puts the lie to the ACU's claim to provide a broad ideological picture across the board. By contrast, the DW-NOMINATE scores for the 115th Senate also have Joe Manchin as the most conservative Democrat and Susan Collins as the most liberal Republican, but Senators Heidi Heitkamp and Joe Donnelly are listed as the second and third most conservative members of the Democratic party, rather than some of the most liberal (recall that the ACU gave both of those senators a conservative score of 0%). Indeed, the difference between Senators Manchin and Collins is 0.053 on a scale from -1 to 1, while the difference from Senator Manchin to the most liberal member of the Democratic party

is 0.707, showing that Senator Manchin is located much closer on this measure of ideology to the conservatives than to the liberals in the Senate. But if the motivation for the ACU to create these legislator ratings is not to provide a broad ideological picture of the legislature, as they claim, then what is it?

My project is to answer this question, not just for the ACU but for other interest groups as well, who likewise produce legislator ratings based on a selected sample of roll call votes. I posit that other motivations, besides purely educational ones, underlie the selection of roll calls that are used in interest group ratings. Such motivations could include, for example, electoral motivations, maintaining partisan brand names, forcing reluctant legislators to vote with the interest group's preferred policy position or risk a lower score, and making specific legislators look more or less extreme than they really are.

In order to find out exactly what motivations are really driving the creation of these scores, I treat each interest group rating as a "test" of a legislator's ideology. Then, using a technique borrowed from the educational testing literature, I assess, not the resulting position of the legislator, but the revealed selection biases of the interest group who created the test. This technique uses item response theory to provide information about how well a particular test is written. It estimates how difficult each question is written to be, how well each question discriminates among test-takers at different levels, and finally how the test takers line up relative to one another on the ability scale being measured. By analyzing the roll call votes (e.g., the "ideological test" questions), I can use the interest group ratings as a mirror to see into their creators' motivations. Are they picking votes that target specific political opponents, for example? Or do they choose votes that make one party look better as a whole than another? These kinds of questions can be answered by turning the interest group ratings back on the interest groups themselves.

The three essays in this dissertation each expose new insights into the behaviors of interest groups who use roll call votes to rate members of Congress. I argue that the set of roll calls which are used in any ideology estimation technique has an effect on the order of legislators that is produced by that method. When the full roll call record is subsetted, as it is in interest group

ratings, then the choice of roll calls used to create the ratings often reveals more about the motives, priorities, and potential biases of those doing the estimation than about the legislators' ideologies. In the first essay, I trace the motivations of interest groups through their choice of roll calls and expose the groups' potential partisan biases using item response theory. In the second essay, this technique is applied to special interest groups like the NEA to show how the motivations of interest groups can be tracked and analyzed even if they order legislators based on single-issue preferences instead of general ideological leanings. Finally, the third essay examines the consequences of subsetting the roll call record at all, as interest groups do with every set of legislator ratings.

In general, this dissertation points to the need to better understand the uses, limitations, and possibilities of ideal point estimation techniques as they appear in political science. In addition, it exposes interest group ratings as an important tools to produce insights into the political world, and ones that should not be left dormant and forgotten as they have been since the development of NOMINATE scores. The rest of this introductory chapter is devoted to spelling out in more detail each of the three essays in this dissertation.

1.1. Interest Group Ratings in the House of Representatives

The first essay in this dissertation examines how interest group scores reveal the motives of the interest groups themselves. I focus on the ADA and the ACU because these two groups both claim to provide an overall picture of legislators' ideology with their ratings. Additionally, these two groups have also been used historically by political science research on legislator ideology (see, among others, Canes-Wrone et al (2002), Berry et al (1998), Groseclose et al (1999), Poole (1981), Poole and Rosenthal (1998), Reeher (2001)). Analyzing these two groups in particular makes the comparability of my results to other interesting studies much more feasible, and exposing the historical motivations of these two groups sheds new light on many interesting older problems. Finally, these two groups are the oldest partisan interest groups of their kind in the US, so their ratings, spanning over three decades and more, provide enough variation to make interesting statistical conclusions about the data.

I begin my analysis starting from 1980 to the present day. Although the ACU began publishing their legislator ratings in 1970, they did not make their choice of roll calls available to the public for the first decade in which they created their scores. The first year in which the actual votes that the ACU used are available to the public is 1980. The ADA makes all of their votes public from 1947 onward, but I analyze their records beginning with 1980 as well in order to ensure comparability with the ACU's scores.

After collecting the votes that the ADA and ACU used in their yearly interest group ratings over the designated time period, I run them through a Bayesian IRT model, which produces three parameters of interest. I use these parameters (alpha, the difficulty of the question; beta, the question's ability to discriminate among test takers; and theta, the estimate of where the test takers line up relative to one another) in my analysis of the ADA and ACU yearly legislator ratings of the House of Representatives with the following questions in mind. Are the groups picking questions that allow them to distinguish liberals from conservatives along the whole range of ideology, or are they picking questions that largely only distinguish one party from the other? Are there roll calls that these groups could have picked, but did not, that would have made for a better test of legislator ideology? How do the interest group ratings change the order and cardinal locations of legislators on the ideological scale relative to their positions in NOMINATE? What do the roll calls themselves reveal about the interest groups who chose them? Are interest groups choosing questions in response to changing institutional features, electoral challenges, or to pressure legislators into supporting specific policy goals? How do the interest group ratings line up with the stated goals of the interest groups themselves?

I find that the interest group ratings are crafted strategically, with an eye to distinguishing among members of the groups' preferred party, and making individual House members look more powerful than they really are in the face of party politics. Rather than presenting a complete picture of ideology using a few key votes, as they claim to do, the ADA and ACU choose votes that produce artificial divisions among their preferred party. These divisions can be used to pressure members to vote more or less conservatively or face potential electoral consequences.

This chapter offers several important contributions to the field of political science. First, the field of political science is broadened by applying the item response theory model from the educational testing literature to the analysis of roll call votes. This analysis allows new insights into the behavior, ideology, and relative preferences of legislators and interest groups to be exposed. Additionally, I build on existing literature about the nature of interest group ratings, including their tendency to lump enemies together and create artificial polarization, and I illuminate the need to update these findings to better reflect the institutional changes that have taken place in Congress over the last four decades. These new insights and analyses reveal interest group ratings as useful tools for political scientists to remember, even though their use has declined since the development of NOMINATE scores. These ratings are a valuable set of data that bring to light useful information about legislators, roll calls, and the groups who create them.

1.2. Issue-Specific Scores: The Case of the National Education Association

In my second essay, I shift gears to focus on a special application of the technique employed in my first article. So far, the IRT methodology has been used to examine measurements of general ideology from liberal to conservative. This chapter will demonstrate its usefulness in political science as a tool to analyze tests of any underlying ability given a set of topically related roll calls by looking at the interest group ratings created by the National Education Association (NEA).

While it is not the only single-issue interest group that creates legislator ratings from subsets of roll calls, the NEA is an ideal candidate for this study for many reasons. First, the NEA has a long history of advocating for its single-issue, support for public education. This means that there is ample data with which to study and track their motivations and priorities through time. Second, while their stated goal has remained consistent over the duration of the NEA's existence, their means of achieving that goal may have changed over time. Employing the IRT technique to their legislator ratings can shed some light on which of these goals received particular emphasis over the course of the NEA's history. Finally, the high-profile nature of the NEA along with its vast influence in the American political process make understanding its motivations a high priority

among political scientists and politicians alike. The IRT methodology can provide this kind of insight through its unique perspective on the choices the the NEA makes, and this will contribute to our overall understanding of the political interactions between legislators and one of the most influential lobbying groups in the country.

The NEA's legislator ratings are designed to provide the group's membership with an ordering of legislators on an educational dimension based on their relative preferences on this subject. I examine their ratings for both the House and the Senate over four Congresses, the 110th - 113th. I argue that the same technique that exposes the motivations of the ADA and ACU using their choice of votes can be applied to the NEA to evaluate their purposes in creating these scores and assess their success in accomplishing their goals over time.

After examining the NEA's explicit statements about the purposes and structure of their scores, I run the scores through the Bayesian IRT model and perform the same kind data analysis as in the first essay. This allows me to gain leverage on many important questions such as the following. Has the NEA computed scores in a consistent manner over time, or have they shifted from a bi-partisan, issue-focused group to a partisan group that merely distinguishes Republicans from Democrats in modern congresses? Is there a better subset of the roll call record that could provide a broader look at education-related preferences than the subset chosen by the NEA? Does the NEA construct their ratings differently for House members than for Senators due to differences in the two chambers' institutional features?

This technique can also illuminate many interesting features of the legislators themselves. I examine how the ordering of the legislators produced by the NEA's interest group scores compares to their general ideological line up using alternative sets of roll calls, and track this measurement through time. How often does the education dimension line up with the liberal-conservative dimension? In addition, I analyze how the content of education-related legislative preferences has changed over time. Has the NEA focused its priorities on roll calls related strictly to education, for example? Or has the NEA allowed other issues such as labor or finances to influence their ratings? These kinds of questions highlight the necessity of examining special interest groups like

the NEA in terms of tracking their motivations and priorities, and the IRT methodology borrowed from the educational testing literature and applied to the NEA's legislator ratings provides a unique perspective into the answers to these questions.

I find that the NEA does not choose roll calls that expose an education dimension that is separate from the liberal to conservative one that most political scientists rely on. They tend to include votes that are not strictly related to education in their ratings, which muddies their claim to be the authority on the educational preferences of members of Congress. Their ratings tend to line up with the liberal-conservative dimension found when analyzing all of the available roll calls in a particular Congress. Interestingly, a look at strictly educational roll calls as identified by recognized political experts[4] also fails to produce a distinct education-related dimension that differs significantly from the liberal to conservative ideological one.

This chapter offers several important contributions for political science. First, it provides an important counterexample to a recent article by Sarah Anderson (2012), who claimed that interest groups that focus on special issues should be used to augment conclusions about Congress that rely solely on NOMINATE scores or the full roll call record to estimate legislator preferences. At least in the field of education, I find evidence that counters her conclusions of a distinct dimension. This chapter also echoes a well-known but rarely used lesson for any science: that repetition is necessary for validating scientific conclusions. I present an important robustness check on previously established conclusions using new data and new techniques, and the conflicting conclusions reveal the need for more research, data, and techniques to finally establish the correct answer.

1.3. Abbreviated Ideology

My third essay begins with the premise that interest groups might actually have the right idea, although they are not necessarily going about it the right way. Given the miscellanea often included in the roll call record, a smaller roll call set could potentially give a cleaner, more precise calculation of legislators' ideology than the full roll call record. But, if this subset is to be useful

[4]The Comparative Agendas Project

practically, it cannot be based on any one person or group's subjective evaluations of the merits of particular roll calls. Having outlined the consequences of using biased or subjective standards to subset the roll call record in my first two essays, I use the third to consider more fully the idea of a mathematically derived standard for choosing a smaller set of roll calls for use in ideological calculations.

The main focus of this chapter is to find out if different subsets of votes (i.e., an "ideal" test, only important vote, or final passage votes, for example) result in a new ordering of legislators than the one found in NOMINATE. If so, are these orderings better, worse, or functionally equivalent? Do they solve the problem of repetitive or extraneous roll calls, which can muddy the ideological waters of the full roll call record, so to speak? If the answer to any of these questions is "yes", then this chapter will have contributed to the project of measuring politician's ideologies by providing a mathematical standard for distinguishing important or useful votes from the plethora of available options in the roll call record each year.

Using hundreds of random samples of both 25 and 100 sample sizes, I establish that a random subset of the modern day roll call record is sufficient to recover ideal points consistent with those from the full record. After finding the smallest minimum vote parameter that suffices for consistent estimates, I turn to an investigation of the kinds involved in the modern day roll call record to find out how each category influences the overall ordering of legislators. Procedural votes and votes on amendments correlate with the full record best, while final passage votes and journal votes correlate worst. Finally, I demonstrate that the minimum number of legislators needed to produce consistent estimates can be lowered from the established default of 50 voters without losing consistency in ideal point estimations. Instead of the 50 voters previously presumed necessary, voting bodies with as few as 4 people can now be analyzed and used to uncover more insights into the legislative process.

This chapter contributes significant new insights to the field of political science. These findings can be used to produce a test of the face validity of groups such as interest groups, who create ratings of legislators with the promise to deliver a clear picture of ideology with a small set

of important votes. If an analysis of their votes delivers ratings that differ significantly from the full record, then their ratings can be shown to be misleading and disingenuous misrepresentations of the Representatives' legislative behavior. Additionally, since journal votes produce totally inconsistent ideal points relative to the full record, these votes in particular should not be used to calculate ideology in any abbreviated subset like those that interest groups produce. But, journal votes might be the best tool for someone who is interested in making a party seem less unified, such as a member of Congress who does not want to be targeted for being a rubber stamp for his or her party. Finally, this analysis opens the door for smaller voting bodies, such as committees, bureaucracies, historical legislatures, and many others who take fewer votes, to be fruitfully analyzed using techniques that were previously reserved for larger voting bodies, such as the US Congress.

CHAPTER 2: CONSTRUCTING A MIRROR — INTEREST GROUP RATINGS IN THE HOUSE OF REPRESENTATIVES, 1980 - 2017[5]

[5]Kisaalita, Alice E. To be submitted to *Journal of Politics*.

2.1. Abstract

This chapter examines how interest group scores reveal the motives of the interest groups themselves. I focus on the ADA and the ACU because these two groups both claim to provide an overall picture of legislators' ideology with their ratings, and I apply a Bayesian item response theory model to the analysis of roll call votes they choose to include. This analysis allows new insights into the behavior, ideology, and relative preferences of legislators and interest groups to be exposed. Additionally, I build on existing literature about the nature of interest group ratings, including their tendency to lump enemies together and create artificial polarization, and I illuminate the need to update these findings to better reflect the institutional changes that have taken place in Congress over the last four decades. These new insights and analyses reveal interest group ratings as useful tools for political scientists to remember, even though their use has declined since the development of NOMINATE scores. These ratings are a valuable set of data that bring to light useful information about legislators, roll calls, and the groups who create them.

2.2. Introduction

The American Conservative Union, an interest group whose mission is "to educate the public about how consistently their elected officials adhere to conservatism," should have had ample opportunity to showcase the ideological behavior of the members of Congress in 2017, a year with unified Republican control of both chambers of Congress and the Presidency (ACU 2017 Ratings of Congress). Even though only 98 bills were passed into law that year, the House of Representatives alone recorded over 700 roll call votes [6]. From these recorded votes, the ACU creates ratings of the Representatives every year in order to display the ideological leanings and differences to the American people by selecting only the few votes that best exhibit members' adherence to conservatism.

[6] A roll call vote is one on which every member's individual vote is recorded. Other kinds of votes include voice votes and division votes, which are only recorded by their yea and nay totals.

In this way, the interest group ostensibly presents a short-cut for evaluating Representatives' legislative behavior to their constituents.

However, of the 25 roll call votes included in the ACU's 2017 Ratings of the House of Representatives, only four were on roll calls that actually became part of a law, and almost half of the votes included were on amendments that did not pass even in the House. For ratings which are meant to "serve as a guide showing you who you can rely on to fight for conservative principles and restore the role of government to what our nation's founding fathers envisioned," they seem, instead, to include mostly ineffectual legislative behavior (ACU 2017 Ratings of Congress). Relics of legislation that do not make a single policy difference to the American people are used to create a picture of the legislative process that the public uses to judge our elected officials and hold them accountable.

The ACU's ratings are often cited by prominent politicians in Congress, and they are frequently used to demonstrate to the voting public a commitment to the conservative agenda. A high rating from the ACU, for example, offers conservative members of Congress bragging rights to the public, and a low rating can be used as electoral fodder for primary challengers. The ACU prides itself on selecting roll call votes that provide important distinctions among legislators and give a broad picture of the ideological leanings of Congress members from a small set of votes. "The votes selected by the ACU are not always considered the 'most important' votes; instead, the selected votes we score help reflect the ideological distinctions among our elected officials" (ACU 2017 Ratings of Congress). But if the true motivations of the ACU are to present the actual policy-making distinctions among members of Congress, then they should focus their ratings on the issues that became a part of policy, and not tout distinctions based on failed amendments and bills that did not change any policies as they did in the 2017 ratings.

It is not just the American Conservative Union who creates ratings for the public in this way. Other interest groups likewise produce legislator ratings based on a selected sample of roll call votes. I posit that other strategic political motivations underlie the selection of roll calls that are used in interest group ratings. Such motivations could include, for example, electoral motivations,

maintaining partisan brand names, forcing reluctant legislators to vote with the interest group's preferred policy position or risk a lower score, and making specific legislators look more or less extreme than they really are.

In order to find out exactly what motivations are really driving the creation of these scores, I treat each interest group rating as a "test" of a legislator's ideology. Then, using a technique borrowed from the educational testing literature, I assess the revealed selection biases of the interest group who created the test. This technique uses item response theory to provide information about how well a particular test is written. It estimates how difficult each question is written to be, how well each question discriminates among test-takers at different levels, and finally how the test takers line up relative to one another on the ability scale being measured.

By analyzing the roll call votes (i.e., the "ideological test" questions), I can treat the interest group ratings as tests of the legislators' ideology, and then use the techniques from the educational testing literature to analyze whether their tests are good ones or could be improved. If they are not good tests of ideology, then what are these groups actually testing for? I can use the interest group ratings as a mirror to see into their creators' motivations. Are they picking votes that target specific political opponents, for example? Or do they choose votes that make one party look better as a whole than another? These kinds of questions can be answered by turning the interest group ratings back on the interest groups themselves.

Kisaalita & Lynch (manuscript) undertake a similar investigation focusing on the ACU and ADA (Americans for Democratic Action) ratings of senators over this time period, asking many of the same questions. We find preliminary evidence that these interest groups are not selecting the votes that best distinguish the relative ideological positions of the senators over the years, but are in fact biasing their ratings to make particular senators look more extreme or partisan than they would appear if the whole roll call record were considered. We look first into the Senate because the roll call generating process is not limited by the germaneness requirements that are present under the rules of the House of Representatives. This should conceivably allow for more votes that divide all parts of the ideological spectrum to be taken and therefore to be chosen by the interest

groups. In other words, the lack of a germaneness restriction in the Senate should allow for a better ideological test than in the House.

This chapter will extend this theory into the House of Representatives to see if the underlying strategies of the ACU and ADA ratings are reproduced in the House. I argue that the set of roll calls which are used in any ideology estimation technique has an effect on the order of legislators that is produced by that method. When the full roll call record is subsetted, as it is in interest group ratings, then the choice of roll calls used to create the ratings often reveals more about the motives, priorities, and potential biases of those doing the estimation than about the legislators' ideologies. Using item response theory, I can trace the motivations of interest groups through their choice of roll calls, as in Kisaalita & Lynch (manuscript), and expose the groups' potential partisan biases.

This chapter proceeds with a review of the scholarly literature on interest group scores, roll call analysis, and legislator ideology in political science, as well as a discussion of the scores' continued relevance in public affairs. Next, I will present both the substantial and mathematical theory behind my use of interest group ratings to investigate the motivations of these groups. I will then present the results of my analysis of the ADA and ACU ratings of the House of Representatives from 1980-2017, comparing their choice of roll calls to the results produced by a test comprising the twenty most informative votes available per year. Together, these analyses expose the actual motivations of the two groups that their own selection habits reveal.

2.3. Literature Review

For most of the 1970s and 1980s political scientists relied on interest group ratings to study the link between congressional voting behavior and legislators' ideology. These scores are constructed by individual interest groups who hand select approximately twenty to thirty roll call votes from a specific calendar year based on the votes' relevance to the group's stated goal. The groups then organize legislators together based on how they voted on these particular roll calls and create a "rating" for each legislator, usually from 0 to 100, based on how often the legislator voted with the group's stated position.

Interest groups create these ratings for a variety of purposes. Special interest groups, like the National Rifle Association or the Sierra Club, choose the votes to include in their ratings based on the votes' relevance to specific causes, like gun rights or the environment. Other groups, like the Americans for Democratic Action (ADA) and the American Conservative Union (ACU) choose their votes in order to provide the public with a broad account of each legislators' conservative or liberal ideological preferences. The common goal for all these groups is to present an abbreviated picture of ideology to the public that is easily digested and understood because it is based on a subset of votes that are deemed important.

2.3.1. Interest Group Ratings in Political Science

For many years these scores were the only tools available for political scientists and other observers to analyze the ideological behavior of members of Congress, and they were put to a variety of uses. To name just a few, they were used to assess institutional features of Congress, like the distribution of preferences on committees relative to that of the full chamber (Ray 1980; Weingast and Marshall 1988; Krehbiel 1990). They served as proxies for legislators' ideology in the study of voting behavior (Kau and Rubin 1979, 1982; Chappell 1981, 1982; Kalt and Zupan 1984; Peltzman 1984; Coughlin 1985; MacArthur and Marks 1988; Dougan and Munger 1989; Davis and Porter 1989; Richardson and Munger 1990, among many others). They were used to study how well members represented their constituents (Markus 1974; Schwartz and Fenmore 1977; Erikson and Wright 1980; Johannes and McAdams 1981). And by pooling several groups' scores together, they were also used to identify the ideological positions of members of Congress themselves (Kritzer 1978; Poole 1981; Poole and Rosenthal 1984; Poole and Daniels 1985).

While the intentions of these interest groups may be admirable, numerous scholars have since shown conclusively that interest group ratings are inherently biased and unreliable as measures of ideology. Several of these critiques emphasize the role of issue selection and content bias in the ratings. For example, Fowler (1982) shows that emphasis on salient roll calls weights the ratings toward a small set of issues and results in a polarized view of congressional decision-making.

Additionally, she shows that the ratings may be inconsistently constructed through time, making long-term analyses of congressional behavior untenable when the ratings are the basis for the analysis.

Other scholars emphasize the statistical issues associated with the scores, such as inconsistency of methodology (Jackson and Kingdon 1992) and inconsistency and bias in estimates which use the ratings as independent variables in regression analyses (Brunell et al. 1999; Herron 1999). Snyder (1992) reveals that the scores tend to misrepresent the ideological positions of many moderate members, making them appear more extreme than they do when all the votes are considered. Finally, Poole and Rosenthal (2007) show that when interest groups are mapped onto the same space as the legislators, they usually exist at the extreme ends of the scales. Given these criticisms, it is not surprising that when the NOMINATE method for analyzing the full roll call record emerged, which produced consistent estimates of ideology over time, the popularity of interest group ratings as a tool for political science research declined.

For these reasons, interest group scores have lost their relevance among political scientists who want to work with legislator ideology. In recent years, interest group ratings have only occasionally appeared in research on congressional behavior. Anderson (2012) uses ratings from special interest groups to better understand legislators' ideological positions on specific issue areas, like the environment, foreign policy, or agricultural politics. Others have used the ratings as robustness checks in addition to NOMINATE scores when making conclusions about societal preferences (Berry et al. 2010), shirking behavior (Tien 2001), and historical changes in ideology (Anderson and Habel 2009). Chand and Schreckhise (2013) and Chand et al. (2014) take a similar approach as Poole and Rosenthal (2007) when they compile many groups' ratings together to place the groups on the same line as legislators. They are thus able to make conclusions about the interest groups themselves, to delineate a typology among interest groups, and to describe similarities in groups' rating behavior when they share the same type. So, while they do pop up here and there in the literature, the overwhelming success and applicability of NOMINATE scores to questions about congressional

ideology has largely drowned out interest group ratings as interesting tools for political scientists today.

2.3.2. Why Interest Group Ratings Still Matter

Despite their irrelevance to modern scholarly ideological research, interest group ratings continue to be produced, published, and touted by these groups as informative and useful guides for the public at large, and for voters especially, about legislator behavior. They offer an avenue for voters to acquire information and knowledge about their representatives without the need to cull meaning from the almost overwhelming amount of information that the full roll call record provides.

There is a well-documented connection between political knowledge and political participation and engagement in the scholarly literature on public behavior in the US (see among others Campbell et al. 1960; Converse 1964; Zaller 1992; and Delli Carpini and Keeter 1996). However, the acquisition of this political knowledge can be problematic, especially in regards to Congress wherein the process of lawmaking can be heavily obscured by its complexity. Many point to a lack of transparency in the legislative process as a potential source of political ignorance on the subject. For example, Lynch and Madonna (2008) document the prevalence of unrecorded voice votes on landmark legislation and find that members' willingness to record their positions on legislation varies with electoral and institutional pressures. Thus, although citizens nearly all have preferences over important bills in Congress, hold beliefs about how their members vote on recorded votes, and use these beliefs to hold their representatives accountable (Ansolabehere and Jones 2010), when members choose not to record their votes these considerations become irrelevant to holding members accountable for their actual behavior. This opacity places a barrier in front of the would-be informed citizen.

However, increasing transparency may itself also be an obstacle in acquiring political knowledge. In an effort to enhance accountability and transparency in governance, Congress has passed several laws, like the Freedom of Information Act, the Government in the Sunshine Act, and the Open Government Act of 2007, which required (with exceptions) that federal agencies be open to

public observation and make information available to the public upon request (Fenster 2006). But this wealth of available information is only transparent in name if it masks the process of political decision-making behind an over-abundance of unrelated details.

In a note on transparency's role in the administrative state, Adam Candeub argues that, "[First,] 'transparency' or 'access' does not really exist if obtaining and securing information is costly in either time or effort. Second, transparency has a 'computational' or 'complexity' dimension, which has an inevitable functional or normative dimension" (Cadeub 2013, p385). Applying this argument to Congress, deciphering the legislative process from an analysis of recorded votes has never been easy, even when Congress was only producing an average of 200 votes every two years. Current Congresses are now producing an average of almost 600 roll calls per year, and that number ranges from 353 to 1177 over the years in my study. If constituents wants to know how well their member of Congress is representing them, the full roll call record may actually be too informative to be helpful.

Fortunato and Stevenson (2016) describe how voters increasingly turn to heuristics in explaining citizens' political attitudes, beliefs, and voting behavior. Interest group ratings provide just such a heuristic about the legislative behavior of members of Congress with which the public engages and uses to inform their evaluations of their representatives. Although seeking political knowledge is made more difficult by the complicated structure of political institutions and behaviors[7], the public is still seeking out this information.

Take Project Vote Smart for example. Project Vote Smart is an organization that compiles evaluations of Congress members by over 400 national and 1300 state special interest groups. Their mission is to "ensure that tolerance will no longer be the only option to the millions of us who are tormented by the issueless rhetoric and often misleading attacks that define contemporary American politics...[and to replace] the most crucial component in our struggle to self-govern: access to abundant, accurate, and relevant information about those who govern us"(About Project Vote Smart). At the time of this writing, Project Vote Smart ranks in the first 50,000 top websites traf-

[7]E.g., contextualizing the massive number of roll calls taken each year.

ficked among almost 7.5 million sites tracked by website traffic monitor, SEMrush.com. Interest group ratings are used in this way to provide a shortcut to political information that would otherwise be intractable for the general public. By selecting only a few key votes taken by Congress each year, interest groups distill the vast amount of available information into digestible cues about how members behave. Thus, they are often used as a proxy for ideology by voters, competing candidates, donors, and other political actors when evaluating the ideological behavior of members of Congress each year.

Members likewise care about their scores. In an article on the power of interest group ratings by Eliza Carney in CQ Weekly, she reports that "[w]hether they win or lose key legislative battles, interest group activists have sent a clear message to lawmakers that they are being watched and graded. And, like nervous schoolchildren, lawmakers are eager to earn gold stars" (Carney 2013). Members know that they way they vote can affect their standing with an interest group that is popular among his or her constituency, and this knowledge thus has the potential to be an influential factor in their vote choice. Given that these scores continue to influence not only the public's perception of their representatives, but Representatives' lawmaking behavior as well, it is imperative that we understand how and why these ratings are being constructed.

It is this lingering connection and importance to the public and to members of Congress themselves that makes it imperative to continue the study of interest groups and their congressional ratings. Although political scientists have discounted their importance to understanding legislative behavior in favor of the more comprehensive look offered by NOMINATE scores, I argue that interest group ratings still have much to offer to our field insofar as they provide insight into the motivations and behaviors of the groups themselves, who have such influence over how the public interprets the actions of Congress.

2.4. Theory

2.4.1. Substantive Motivations

Interest groups scores can be an invaluable tool for assessing where an incumbent politician has stood on a series of votes and issues in a given year or two. These scores are easily digestible to a large audience because they distill the vast and complicated legislative process into a simple number by which we can compare and contrast legislators based on their stances on important issues. But when biases enter the picture, the usefulness of interest group scores is lessened.

If the motivation is truly to provide a picture of the relative liberal or conservative leanings of a group of legislators as they are and with no ulterior goal, then the project in which interest groups are engaged provides the public with an important tool for quickly and easily understanding the stances their representatives have taken on important issues. But when these same ratings are used, not simply to educate the public, but to influence how the public interprets the stances the legislator has taken, then the public's basis for evaluating their representatives becomes corrupted by the artificial stretching of the truth exhibited in the biased scores. It is therefore of utmost importance to discern the true motivations behind each set of interest group ratings. Once we understand whether a group assigns ratings simply in order to punish their opponents and help their friends, for example, then we can assign the appropriate discounts to the weight we give to their ratings.

Although the project of measuring ideology with a smaller set of roll call votes could prove extremely valuable to both scholars and the political public alike, Kisaalita & Lynch (manuscript) argue that interest group ratings, such as the ones created by ADA and ACU, do not accomplish this task in an unbiased, non-partisan way. We test our argument using interest group ratings for senators from 1980 to 2016, and we show that, on average, both groups choose suboptimal roll call votes relative to a mathematically constructed test of ideology that maximizes information across the entire range of the ideological scale.

In fact, a better use of the interest group ratings is to turn them around on the interest groups themselves. The votes they choose to use in their ratings are drawn from a larger set of available roll calls, and their decision to use one vote and not another can provide insight into the motivations of the groups themselves. We find that both groups respond in systematic, predictable ways to changes in institutional variables, like the majority/minority status of their preferred party and/or a change in the party of the President, by using roll call votes that prioritize information about one party over another as the principal components of their legislator scores. The underlying theory and methodology used in that manuscript will be extended and applied to the House of Representatives in this chapter in order to gain a better understanding of the motivations that interest groups employ in their choice of which roll call votes to include in their scores.

Thus, for each year an ideal test is constructed[8] and measured against the ADA and ACU yearly ratings in terms of information, intra-party variation, and inter-party distance in the House of Representatives. In order to assess the quality of this new measure of interest group motivations and effectiveness it is important to lay out some expectations for how the ideally constructed twenty-question test should behave relative to the interest group ratings.

If the interest groups are actually creating scores that accomplish their stated goals of reproducing the ideological spectrum of legislators using a few key votes, then they should produce similar results in terms of the level of intra-party variation and inter-party distance as the ideal test. If, instead, they have alternative motivations, as much of the literature on interest group ratings has already established, we should see significant differences in the levels of intra-party variation and inter-party distance from the levels we find using the most informative votes. Additionally, these differences should be in predictable directions.

For example, if the ADA is focused on informing the public *only* about the Democratic party, then the Republican intra-party variation might be less than that found in the ideal set, since the ADA would not be concerned with highlighting any variation among their opposition. At the same time the level of Democratic intra-party variation for their tests might exceed the variation

[8]Details about the construction of these tests are described further in the next section.

produced from the ideal test, because their main task would be to show or even exaggerate any distinctions, however subtle, among their preferred party. This expectation comports with the analysis of Brunell et al (1999) in which the authors found evidence that interest groups tend to produce scores that lump their enemies together while simultaneously distinguishing much more finely among their friends.

> **Lumping Hypothesis:** Among the interest group scores, intra-party variation will be smaller than the ideal test for the opposition party and greater than the ideal test for the interest groups' preferred party.

The mean distance between parties might also vary depending on the underlying motivations of the groups creating the scores. Snyder (1992) discusses how interest groups tend to include votes with close margins rather than lopsided ones in their ratings. This makes the resulting ratings look bimodal and forces artificial extremism into the ratings. This is slightly different from the Lumping Hypothesis in that Snyder's emphasis on the choice of party-line votes common in the interest group ratings he analyzes means that the ratings push *both* parties toward the extremes, rather than just the opposition party, thereby increasing measures of polarization. His analysis is based on ratings from the late 1970s and early 1980s, a period of time in which party power was weak and individual members of Congress were thriving in candidate-centered elections (Jacobson and Carson 2015). Thus, the motivations for the interest groups in this time period may have been centered on defining the differences between the parties.

Snyder concludes that artificial extremism is an inherent characteristic of interest group ratings, and his hypothesis has not been tested or updated to account for the changes in party power that we have witnessed in the decades following his analysis. I propose to do just this, and see if artificial extremism is in fact an intrinsic part of interest group ratings or just one possible response to institutional features of Congress like party power. If this strategy holds, then the mean inter-party distance produced from the interest group ratings should be greater than the distance produced from the ideal test.

Artificial Extremism Hypothesis: Interest group ratings should produce more polarized estimates than the ideal test on average.

By focusing on these two measures (intra-party variation and inter-party distance), we can begin to assess whether interest groups strategically rate legislators to make them appear different than they would if the best available information was used to rate them. The ideal test constructed from the Bayesian IRT framework will contribute to this literature by providing a way to systematically test and expose the inherent biases and motivations of interest group ratings.

2.4.2. Mathematical Motivations

In order to assess the motivations of the interest groups in my study, I will borrow from the educational testing literature's use of item response theory (IRT) to assess the quality of tests, test questions, and test takers. The IRT model in this setting returns three quantities of interest[9]. First, it returns a list of "ability" scores, called thetas, by which an ordering of legislators can be produced based on their responses, yea or nay, to each question included in the test. The ability that the interest groups are testing in this analogy is a legislator's ability to answer conservative or liberal votes correctly. It is important to note that the interest groups I examine, the ADA and ACU, count abstentions as "nay" votes, so the model will be between a dichotomous choice.

The other two quantities of interest provide information about the test questions, the votes, themselves. The difficulty parameter, called alpha, informs us about the vote's placement along the liberal-conservative scale. It is akin to a question on a math test that only A- and B-level students should be able to answer correctly with ease, while C-, D-, and F-level students would have trouble answering correctly. On a test of a legislator's ability to answer questions conservatively, a vote with a difficulty parameter on the right-hand-side of the liberal-conservative scale might be easy for extreme conservatives to answer "correctly", while it would be more difficult for moderates and liberals to answer the "correct" (conservative) way.

[9] A full description of the IRT model is provided in Appendix A.

The third parameter returned by the IRT model is the discrimination parameter, called beta. This parameter measures how well a particular question discriminates between those who answered the question correctly and those who answered incorrectly. A question designed for a specific difficulty level would have a high discrimination parameter if the people answering the question all answered it as expected. If all the people to the right of the question get the answer right and none of the people on the left do, then the question will have a high discrimination parameter. A question with a lower discrimination parameter would be a bad test question, because some of the people to the left of the question would get it right, while some of the people on the right of the question would get it wrong. The discrimination parameter can be thought of as the slope of the S-curve for each question in the test. The more vertical the slope is, the better the question is able to distinguish between liberals and conservatives.

The difficulty and discrimination parameters can be combined together into item information functions (IIFs), which provide important details about each question and allow for different questions to be compared on the same scale. In addition, the IIFs for each test can be added together into a test information function (TIF), which gives information about the test itself. These two functions place tests and test questions on the same scale, even if the same set of people did not take each test and even if the same questions were not asked on each test. They therefore provide a way to compare and contrast different groups' ratings over time even though they may comprise different roll call votes and different legislators voted on the questions.

The ideal tests used in this analysis are constructed by running the full roll call record through a Bayesian IRT model and recovering the alpha and beta parameters for each roll call in a given calendar year. A vote is considered "ideal" if it has the highest discrimination parameter (beta) at a given difficulty level (alpha). By dividing the difficulty levels into 20 evenly spaced segments and selecting the vote in each section with the highest discrimination parameter, I construct an ideal test for each set of roll calls in my analysis. The information from each roll call in the ideal test can be added together to provide a measure of how informative the test itself is overall (i.e., the TIF value).

Table 2.1: Average Information (TIF) Across Five Thetas (1980 - 2017)

Group	**Theta Level**				
	-2 (Extreme Liberal)	-1 (Liberal)	0 (Moderate)	1 (Conservative)	2 (Extreme Conservative)
Ideal	6.63	18.57	26.55	19.07	6.52
ACU	4.38	12.15	17.96	11.39	4.96
ADA	3.23	10.82	15.99	10.79	3.62

The ideal tests for each year are constructed from the twenty most informative votes that range across the ideological spectrum without relying on any subjective judgement as to their ideological importance. The ideal tests constructed in this way should be a better backdrop for comparing and assessing the motivations of the interest group ratings than the full record, not only because of their comparable size, but also because they are specifically designed to maximize the discriminating information between legislators each year. Thus, for a given year, I expect that the ideal test should be more informative than the interest group tests on average. A cursory glance at the average levels of information provided by each group shows that my ideal test does perform better at this task.

Table 2.1 shows the average level of information produced in the TIF for each set of votes from 1980 - 2017. Each of the tests are most informative about the middle of the scale, and are least informative about the extremes of the ideological scale. The ideal test contains significantly more information about the distribution of legislators on the liberal - conservative scale than either interest group rating at the conventional .05 level. This shows that interest groups do a suboptimal job choosing ideologically distinctive votes for their ratings, and reinforces the need for a more standard test, such as the one provided by the IRT model. If the goal is to choose the twenty or so most informative votes so as to reduce the clutter involved in distinguishing important differences in legislative ideology, then the IRT model can provide a more informative and less subjective test than the interest groups create. In the next section, I describe the specific ways in which I used the IRT model to analyze the ratings of the ACU and ADA and to construct the ideal tests to which the IG ratings can be compared over the 38 year time frame.

2.5. Data and Methods

This chapter extends Kisaalita & Lynch (manuscript) into the House to examine how interest group scores reveal the motives of the interest groups themselves. I focus on the ADA and the ACU because these two groups both claim to provide an overall picture of legislators' ideology with their ratings. Additionally, these two groups have also been used historically by political science research on legislator ideology (see, among others, Canes-Wrone et al. 2002; Berry et al. 1998; Groseclose et al. 1999; Poole 1981; Poole and Rosenthal 1998; Reeher 2001). Analyzing these two groups in particular makes the comparability of my results to other interesting studies much more feasible, and exposing the historical motivations of these two groups can also help to shed new light on many interesting older problems. Finally, these two groups are the oldest partisan interest groups of their kind in the US, so their ratings, spanning over three decades and more, provide enough variation to make interesting statistical conclusions about the data.

I begin my analysis starting from 1980 to the present day. Although the ACU began publishing their legislator ratings in 1970, they did not make their choice of roll calls available to the public for the first decade in which they created their scores. The first year in which the actual votes that the ACU used are available to the public is 1980. The ADA makes all of their votes public from 1947 onward, but I analyze their records beginning with 1980 as well in order to ensure comparability with the ACU's scores.

I initially ran all recorded votes through the IRT model to acquire alpha, beta, and theta parameters for every recorded vote available from which interest groups could have chosen. I then compiled the votes from the ADA and ACU from each year and ran them through their own IRT models using only those few votes to get the thetas, alphas, and betas produced from the interest group tests and to create individual IIFs and TIFs for each yearly set of ratings. All IRT models used the same individuals to set the polarity of the scaled thetas from -2 to 2. I constructed the "ideal" test by selecting twenty votes spaced evenly along the (-2,2) interval, choosing the vote in each .2-length space with the highest discrimination (beta) parameter. If a .2-length space did not

contain any votes, then my ideal test contains fewer than 20 votes for that year[10]. In this way, the three tests were compiled for each year from 1980 - 2017. I was then able to assess the levels of variation and inter-party distance produced among the thetas for each set of tests, which I compared using a series of t-tests between the groups, and the ideal tests. The results of my analysis are presented in the next section.

2.6. Results

Since the interest groups are not maximizing information across the board given their self-imposed 20-question limitation, I turn to an analysis of the variance and distance produced among the thetas estimated from each test to expose their true motivations. Perhaps the interest groups are not selecting the *most* informative votes, but are they still selecting votes that provide *enough* information to convey the same information as the ideal set? If they are still accomplishing their stated goal of reproducing the ideological spectrum, then the average intra-party variance and inter-party difference produced by their scores should not differ significantly from the ideal set. Table 2.2 presents the results of the comparisons between each test, broken down by group and party.

In terms of intra-party variance, both the ideal test and the ADA's interest group ratings expose more variance among the Democratic party than the Republican party on average during this time period. Even still, the ADA distinguishes among the Democratic party significantly more than the ideal test does (0.37 vs 0.24). The ACU, on the other hand, produces tests that are as informative of Republicans as they are of Democrats in general. This shows that the ACU is inflating the amount of intra-party variation among the Republicans relative to what the ideal test would normally produce.

Consistent with the Lumping Hypothesis prediction, both the ADA and the ACU show a tendency to distinguish among their friends more finely than is warranted by a comparison to the ideal test. However, the other side of the Lumping Hypothesis is that the groups will lump their enemies

[10]In other words, if no recorded vote among the entire set had a difficulty (alpha) parameter on that .2-length part of the interval, then no roll call was evaluated for that spot.

Table 2.2: Relative Intra-party Variance and Inter-party Distance, 1980 - 2017

Group 1 vs Group 2	Mean of Group 1	Mean of Group 2	t-Value
Intra-party Variance			
Ideal (R) vs Ideal (D)	0.19	**0.24**	-2.54
ACU (R) vs ACU (D)	0.30	0.34	-1.40
ADA (R) vs ADA (D)	0.24	**0.37**	-4.20
Ideal (R) vs ACU (R)	0.19	**0.30**	-6.30
Ideal (D) vs ACU (D)	0.24	**0.34**	-3.67
Ideal (R) vs ADA (R)	0.19	**0.24**	-2.80
Ideal (D) vs ADA (D)	0.24	**0.37**	-4.19
Inter-party Distance			
Ideal vs ACU	**2.07**	1.83	6.09
Ideal vs ADA	**2.07**	1.83	5.68

Notes. All differences are statistically significant at the p .05 level, except for the intra-party variance between the ACU (R) vs ACU (D). Entries in bold indicate which group has the higher mean value.

into a clump rather than distinguish finely among them. This part of the Lumping Hypothesis is not supported by my analysis. Instead, both interest groups also distinguish more finely among the opposition party than the most informative votes comprising the ideal test do at a statistically significant level.

The twenty most informative votes in each year produce a picture of legislators in which there is much less intra-party variation than the interest group ratings would have the public believe. For the most part, legislators in a party vote together. In addition, party unity and party power has been on the rise over the years in my analysis (see Smith 2007). So, while lumping may have been an important explanation for interest group behavior in the past, this analysis shows that tendency is no longer represented in today's interest group ratings and may be subject to changes in institutional variables like party power in Congress.

To understand this result better, recall that the IRT model is akin to a math test, where the questions reveal the underlying mathematical abilities of the test takers. If the students taking the test all have the same underlying ability, they should all receive the same score on the test. The clumping that is found in the ideal tests shows that, given the questions that are being asked in the

roll call record, the legislators taking this test all have basically the same underlying ability to be conservative or liberal. The most informative questions across the ideological spectrum test the legislators' conservative abilities, showing the most Democrats vote together non-conservatively while most Republicans vote together conservatively. The interest group ratings disguise this fact and produce ratings that show much greater variation among the parties than really exists.

This conclusion is further supported by the evidence for the average inter-party distance comparisons for this time period. Relative to the ideal test, both interest groups produce a lower rate of polarization on average[11]. By choosing votes that highlight the distinctions among the legislators in their preferred parties, they make that party look more ideologically diverse than it actually is. Thus, contrary to what the Artificial Extremism Hypothesis suggests, the ACU's ratings identify many more moderates among the Republican party than the ideal test. A similar tactic is used by the ADA to distinguish finely among the Democrats. This has the effect of decreasing the average distance between the parties in the interest group scores relative to the average distance found in the more informative test. This strategy allows the interest groups to put pressure on members of their own party. For example, the ADA can make a Democratic representative look more conservative than they might otherwise appear and thereby incite a primary challenge in the next election.

To further illustrate the strategies employed by the ADA and ACU, Figure 2.1 plots the estimated legislators' positions on the liberal - conservative scale from each of the tests for the year 2016. The ideal test estimates are located on the x-axis and the interest group ideal point estimates on the y-axis for both groups. In both panels of the figure, the interest groups distinguish among the two parties much more finely than the ideal test, showing that the Lumping Hypothesis is only partially true in the modern Congress. Interest groups are willing to highlight small distinctions among legislators that do not appear when the most informative votes are considered in order to highlight intra-party differences in both their preferred and opposed party. This also has the effect of decreasing the levels of polarization between the party if only the interest group ratings are considered, unlike what we would see if the Artificial Extremism Hypothesis still held true today.

[11] This is measured by taking the difference between the average Republican and the average Democrat in each year.

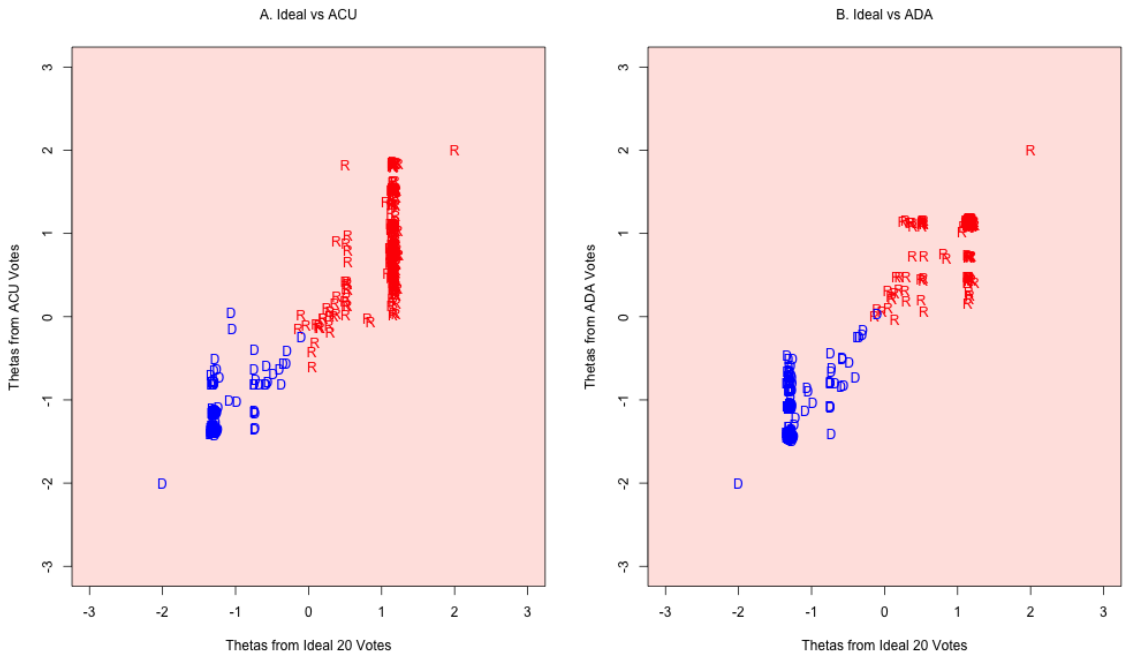

Figure 2.1: Theta Estimates for 2016

Even still, the ideal test provides more information than either interest group about the legislators ideological position using the same number of votes.

2.7. Conclusion

In this chapter, I have applied a Bayesian IRT framework to the analysis of interest group ratings in order to turn their ratings into a mirror that shows the motivations and behaviors of the groups themselves. By systematically testing the motivations and behaviors these groups use when constructing their ratings and comparing them to a mathematically constructed ideal test, I have shown that several established theories of how interest groups behave when constructing their ratings, including lumping and artificial extremism, should be revised to incorporate the role of increased party power in the modern Congress.

This IRT framework has only been applied to two interest groups in this analysis, both of which claim to create ratings that reflect the broad ideological divisions in Congress. Upon closer examination, however, other motivations have revealed themselves, including distinguishing among

members of the groups' preferred party and creating ratings that place more responsibility for political outcomes squarely on individual legislators' shoulders. The techniques used to analyze the ratings of the ACU and ADA can be usefully applied to other interest groups with both broad and narrow focuses. Additionally, the ideal test itself provides a new backdrop by which any groups' ratings can be compared using a similar number of votes.

The ideal test is created as a better alternative to the project interest groups say they are attempting to accomplish. Rather than use subjective judgments about which votes best inform voters about their legislators' behavior, the ideal test is constructed using a mathematical standard for providing the most information across the ideological spectrum given a limited number of votes. This test reveals that on average between the years 1980 - 2017 legislators in both parties vote as a group. Of course, the most informative test is the one that uses the full roll call record. But it is remarkable that even with the severe constraint of selecting only 20 informative votes, the analysis revealed significant differences in the average levels of information, intra-party variation, and inter-party distance in the interest group ratings relative to the ideally constructed test.

From these results, it can clearly be seen that the interest group ratings are crafted strategically, with an eye to distinguishing among members of the groups' preferred party, and making individual House members look more powerful than they really are in the face of party politics. Rather than presenting a complete picture of ideology using a few key votes, as they claim to do, the ADA and ACU choose votes that produce artificial divisions among their preferred party. These divisions can be used to pressure members to vote more or less conservatively or face potential electoral consequences.

In today's Congress, which the ideal tests show to be characterized by strong party unity and party power, the ADA and ACU would have the public believe that there are vast individual differences among House members within and between parties. The creation of this perception that individual politicians still retain strong political power in the face of weak party leadership and control allows interest groups to target these individuals in elections or on specific roll call votes and thereby exert their influence on political outcomes.

CHAPTER 3: ISSUE-SPECIFIC SCORES — THE CASE OF THE NATIONAL EDUCATION ASSOCIATION[12]

[12] Kisaalita, Alice E. To be submitted to *Journal of Politics*.

3.1. Abstract

This chapter examines several subsets of roll call votes in order to identify the dimension by which legislators decide educational policies. While several scholars have argued that this field is a special issue area comprising new, unexpected, and different divisions among legislators than we typically use to characterize the members of Congress (Clausen 1973, Wilcox and Clausen 1991, and Anderson 2012), my analyses do not support this conclusion. Educational voting usually falls along the same division lines as those we see on the traditional liberal to conservative dimension, and the two dimensions correlate at or above 95% on average. Given that there is evidence both for and against the existence of an education dimension, this question should not be taken for granted among political scientists until we as a field come to an agreement about the best way to measure and answer it.

3.2. Introduction

In a representative democracy, legislators are expected to vote on legislation according to their constituents' preferences. Much of the time, however, there is no simple criterion by which voters can clearly evaluate their representatives' performance in this respect. From the time a bill is introduced to the time it becomes law, there are ample opportunities for legislators to cast votes on it—in committee, on amendments to it, on rules pertaining to its passage through Congress, on the same bill after it has been changed by the other chamber of Congress, on disagreements between chambers, and potentially even on passage over the president's veto.

At each of these stages, legislators may vote multiple times on the bill and these votes are all recorded in the chamber's journal for public record. For any given bill, legislators may vote anywhere from zero to hundreds of times on it. Couple this with the large number of bills introduced in each two-year span of Congress and it becomes clear that legislators usually vote thousands of times per year on legislation in one form or another. Thus, the problem of determining whether legislators have done well representing their constituents' preferences becomes rather intractable

in the face of the the sheer quantity of votes by which we are to evaluate them. In addition, in recent decades, legislation has often been passed in omnibus form or with multiple riders which are only loosely related to the original purpose of the bill. What exactly does a "yes" or a "no" vote on this kind of bill signify in terms of a legislator's representative skills?

While people of all types have grappled with this question, be it theoretically, historically, philosophically, or practically, several interest groups have done so for the past forty years and more in a consistently measurable way by choosing among the thousands of available votes and picking out the important ones. The goals and motivations of these groups differ which leads to wide variation in the composition of their legislator ratings. For example, the ADA and ACU choose twenty votes per year to display the liberal to conservative differences among the members of Congress. The NRA, on the other hand, combines some of the votes taken in Congress with other factors like endorsements and donations when ratings legislators strictly on their gun-related preferences. But overall these ratings ostensibly provide a way for the public to gain knowledge about the distinctions among members of Congress without having to sift through the minutia and nuance comprising the Congressional Record.

This chapter will examine one such interest group, the National Education Association (NEA), in order to clarify the insights that such issue-specific ratings can provide to researchers and the public. Their mission statement is to "advocate for education professionals and to unite our members and the nation to fulfill the promise of public education to prepare every student to succeed in a diverse and interdependent world" (National Education Association). They grade members based on their support of this mission, with a focus on their "overall support for public education and educators" (National Education Association).

This specialized focus on education provides a foil to the usual way of evaluating the ideological distinctions among members of Congress. While the liberal to conservative ideological dimension often does an impeccable job of explaining or predicting legislators' voting behavior at large, many have argued that this general, bird's-eye view can cover up interesting nuances and ideological differences when it comes to specific issue areas like agriculture, education, business,

gun-related or foreign-policy related preferences, among others (see Wilcox and Clausen 1991, Anderson 2012, Clausen 1973). This chapter will take up this premise and test it with the NEA's legislator ratings to find out the usefulness of issue-specific interest group scores in the field of education.

3.3. Literature Review

3.3.1. The Education Dimension

Divisions like those that attend every aspect of educational policy-making ensure that this issue will remain a highly salient one that will continue to be debated and changed regardless of who attains power in Congress. While most people would agree that a graduate degree should not be a requirement for successfully participating in society, there remains a conflict among parents, teachers, citizens, and legislators about the basic level of education a US citizen ought to receive and what exactly it should contain. Who ought to be responsible for the consequences of citizens not knowing enough to fully participate in their own democracy?

For example, a debate about who is responsible for sex education has been raging in American politics for the last few decades. While many people take a moral or religious stance on this issue and desire to leave it exclusively to the parents to decide a child's level of sex education, others claim it is an issue that should be taught in public schools at the taxpayers' expense, pointing to the preponderance of teenage pregnancy and sexually transmitted diseases which impose both a social and financial cost on society at large. Is the subject of sex education one on which the US Congress should take action?

This question, and others like it, help to explain why education is consistently a salient issue for Americans, but also not one on which there is a consensus. Defining the content of the minimum educational standard for American citizens is a question that crosses many other traditional divisions (such as liberal versus conservative) and which can make allies out of those who otherwise would be a political opponent. After all, "[a] greater proportion of citizens may approve of teaching American history than sex education in schools...but *how* schools teach sex education

and American history matters more to most citizens than *whether* schools teach these subjects, and there is no consensus on how either American history or sex education should be taught. There is, in this crucial sense, no consensus on teaching even the 'basics"'(Gutmann 1999, pp. 4-5, emphasis in original). Given the plethora of perspectives about how the government should relate to education in American society, it is no surprise that many scholars point to this issue as one that goes beyond a simple liberal versus conservative dynamic.

As much as certain sections of American society would like the government to have no place in the educational process (e.g., libertarians[13]), most people acknowledge a role for the government to play on this matter. In her book entitled Democratic Education, Amy Gutmann poses a particularly compelling argument for why the manifestation of education in any democracy should continually be deliberated in the public policy-making spaces of that society.

> "Deliberative democracy underscores the importance of publicly supported education that develops the capacity to deliberate among all children as future free and equal citizens. The most justifiable way of making mutually binding decisions in a representative democracy—including decision not to deliberate about some matters—is by deliberative decision making, where the decision makers are accountable to the people who are most affected by their decisions. Deliberative decision making and accountability presuppose a citizenry whose education prepares them to deliberate, and to evaluate the results of the deliberations of their representatives. A primary aim of publicly mandated schooling is therefore to cultivate the skills and virtues of deliberation"(Gutmann 1999, pp. xii-xiii).

Here we see why it is not only important to understand the role of government in education, but also the role of education in government. Citizens in a democracy must be educated to such a point that they are capable of evaluating the policies and deliberations of their law-making representatives.

[13]From the 2018 Party Platforms of the Libertarian Party: "Parents, or other guardians, have the right to raise their children according to their own standards and beliefs, provided that the rights of the children to be free from abuse and neglect are also protected"(Libertarian Party 2018 Platform, Section 1.6).

In order to evaluate the activities of legislators, as is required for a healthy democracy, it is necessary to know what it is that they do. Although the US Constitution mandates that both the House and Senate keep a journal to track their activities, this resource is difficult to sift through, and the legislative process has become a veritable quagmire from which even experts have a difficult time gathering meaningful data and conclusions as a result.

For decades, scholars and the public alike have turned to shortcuts and heuristics to provide the most important and relevant information about how members actually behave when they make our laws (Fortunato and Stevenson 2016). Interest group scores are one of the prime examples of such a heuristic, and they have been used in a variety of contexts to elucidate the legislative process (see, among others, Kau and Rubin 1979, 1982; Chappell 1981, 1982; Peltzman 1984; Coughlin 1985; MacArthur and Marks 1988; Dougan and Munger 1989; Richardson and Munger 1990, Tien 2001, Poole and Rosenthal 2007, Anderson and Habel 2009, Berry et al. 2010, Chand and Schreckhise 2013, and Chand et al. 2014). These scores are created by groups who sift through the roll call voting record of the two chambers of Congress and hand-pick the votes that are most relevant to the group's stated purpose. The groups then grade the members of Congress based on the number of votes on which the legislator and the group agree.

The scores provide information on a variety of subjects, both general and issue-specific, depending on which group creates them. Groups like the American Conservative Union (ACU) and Americans for Democratic Action (ADA) create scores that focus on delivering information about the main liberal to conservative dimension that describes the main economic and social divisions likely to appear in any given legislative session. Others like the National Rifle Association (NRA) or the Sierra Club develop scores that focus on specific issue areas only, like gun rights or the environment. The National Education Association (NEA) is one of these issue-specific groups that focus on providing legislator ratings based solely on their preferences over the subject of education.

In the 1970s and 1980s, these scores were used in political science because there were very few alternative data sets that rated the relative beliefs and actions of members of Congress available to scholars on the same scale, and their use continued well into the 1990s among political scholars.

While these scores are still created and used by the public today, their use in the field of political science diminished after the development of NOMINATE scores. This method eschewed the subjective part of creating scores from only the votes deemed most relevant by human actors in favor of using all of the available votes and providing the most informative scores by which we can distinguish our legislators. However, although their use in political science has declined substantially, some scholars still argue for their use in augmenting our understanding of the US Congress.

Clausen (1973) and Wilcox and Clausen (1991) argue that even though many different issue areas may include the same groups of people voting with one another, that does not mean they are using the same reasons and logic to come to the same conclusion as their compatriots. Thus, if the different reasons for voting overlap, causing those who prioritize their rural constituents to vote the same way as those who prioritize their conservative constituents, this does not mean that both legislators are voting "conservatively". Interesting differences can underpin members of Congress who seemingly vote the same way, and it is presumptuous of scholars to attribute all voting behavior to a single liberal to conservative explanation.

Building off of this previous work, Anderson (2012) attempts to access these underlying differences by looking at subsets of votes which deal specifically with these issues. In other words, she advocates for the use of interest group scores, which are based on a small sample of twenty or so votes and which are chosen by groups dedicated to advancing those issues in particular.

Accordingly, Anderson offers a list of domain-specific issues whose nuances are not particularly well-captured by the more general liberal to conservative dimension most prominently displayed in NOMINATE scores. She identifies several such topics including agriculture, business interests, environmental protections, and, notably, education. She claims that "issue areas that exhibit intraparty division, such as agriculture, are not well captured by the NOMINATE scores, and interest group scores should be used in these areas" (Anderson 2012, p1188). Among the groups whose scores she selects as useful foils to NOMINATE scores in domain-specific areas, Anderson includes three agricultural interest groups, three business groups, two conservation groups, and only one for education—the National Education Association.

The National Education Association is an organization "committed to advancing the cause of public education" (About NEA). As can be seen by the variety, breadth, and changeability of the federal legislation that has affected American education across US history, the cause the NEA takes up is by no means a simple liberal to conservative one. This nuance is further illustrated in the NEA's legislator ratings, which they have been producing for each Congress since 1969. Anderson (2012) shows that the NEA scores' average correlation with NOMINATE scores hovers around the mid-70% to low-80% range, and other important variables, like the location of the median on the education dimension, correlate with NOMINATE's median member only 63% of the time on average. There is, she insists, an "education dimension" that explains the voting behavior of legislators better than the traditional liberal to conservative dimension that we usually rely upon. Instead of using the full record, we should use the more narrowly focused NEA legislator ratings to gain insight into the relevant distinctions among legislators along this dimension.

Before wholeheartedly accepting this recommendation, a look into the history of the NEA seems in order. Although many people look to this group for guidance and expertise on education policy in America, it is important to clarify who they are, how and why they make decisions about which education-related votes to include in their ratings, and why we should look to them, as Anderson (2012) does, to define a more accurate "education dimension" in the first place.

3.3.2. A Brief Look at the National Education Association

The NEA has been a leader in American education policymaking since it was established in 1857. Their organization has been instrumental in both crafting and advocating for legislation in support of public education throughout its history. In addition to implementing a political agenda in Congress, the group stresses the importance of learning about the legislative process and involving its members therein. Their Legislative Action Center encourages its members and the public to "Learn about the issues and policies that impact public education, then take action to support students" (National Education Association). For this reason, the NEA began rating members of Congress on their support for public education beginning in 1991.

Informing the public about their legislators preferences over education policies became more pressing as the topic of education became more salient on the national stage in the 1980s and 1990s. An incendiary report from the Department of Education pushed the issue into the forefront of American policymaking in the 1980s, noting that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people" (National Commission on Excellences in Education 1983). This report accompanied a slew of recommendations for shifts in the curricula geared towards improving our standing relative to other countries.

Congress passed several laws to improve education in the 1990s, including the *Goals 2000: Educate America Act* of 1994, the *National Skills Standards Act* of 1994, the *School-to-Work Opportunities Act* of 1994, the *Charter Schools Act* of 1998, and the 1998 *Reauthorization of Higher Education Act*. However, since 1970, "per-pupil spending has more than doubled (taking inflation into account) without bringing any tangible results. Students are learning and achieving at about the same level as in the 1970s" (Yanushevsky 2011, p62). This finding prompted critics of the government's previous efforts to try out alternative methods for improving schools in the US.

The *No Child Left Behind Act* of 2001 (NCLB) was one such response, which tied federal funding for the first time to achieved educational progress. This act "required all states to test K-12 students regularly in core subjects and to evaluate schools based on whether their students were making adequate progress toward achievement benchmarks, with the goal of 100 percent proficiency by 2014" (Deming and Figlio 2016). While this act was initially very well received, its popularity declined as it became more and more obvious that states would fail to achieve these standards within the prescribed time.

Several problems with NCLB arose during its implementation, the chief among them being that "President Bush forced states to follow the law but never provided money for the effort" (Jannetta 2009). The impact of this lack of overall funding was only compounded by the financial penalties that incurred when schools underperformed. Others criticized the law for its emphasis on the testing requirements by which schools, teachers, and students were evaluated. "The requirement

is bad public policy because it takes a 'tough love' approach that, like the 'War on Drugs,' results in increased sanctions but does little to remedy the underlying problem. Additionally, it is bad public policy to evaluate schools and students on a possibly culturally biased test and to force school districts to adopt instructional strategies aimed at 'teaching to the test"' (Williams 2008, p580-581).

These problems prompted a major focus on education policy reform during the 2008 presidential campaign. President Obama promised to

> "reform NCLB, which starts by funding the law. Obama and Biden believe teachers should not be forced to spend the academic year preparing students to fill in bubbles on standardized tests. He will improve the assessments used to track student progress to measure readiness for college and the workplace and improve student learning in a timely, individualized manner. Obama and Biden will also improve NCLB's accountability system so that we are supporting schools that need improvement, rather than punishing them." (Jacobson 2011)

Politifact.com, a website that has fact-checked and tracked the veracity of politicians' promises since 2007, marks this particular campaign promise as one that President Obama kept. He did this by implementing a program called Race to the Top (RTTT), which was "intended to encourage and reward states that are 'creating conditions for innovation and reform"'(Weiss 2014). Essentially, Obama waived many of the penalties for states that were clearly not going to meet the standards set by the NCLB Act, and provided other avenues for them by which they could receive federal financial aid.

Still, even the alternative provided by the Race to the Top program was accompanied by its own set of problems. A report evaluating the program after its first three years detailed three main conclusions about the program and its effectiveness. First, states made unrealistic and impossible promises. Second, RTTT policies fell short on teacher improvement and failed to address core drivers of opportunity gaps. Finally, RTTT shortcomings have spurred state-district and union-management conflicts that hinder progress (Weiss 2014).

Throughout the developments and implementations of each of these policies, the NEA has taken a stance and had a hand in how the public receives and interprets their significance. While they initially supported the *No Child Left Behind Act*, the organization opposed many aspects of its implementation. Today, the NEA advocates for continuing changes to the legislation in Congress, focusing its attention on leveling the playing field for poor children and improving the evaluation criteria for teachers (Layton 2015). And, while it praised the goals and motivations of President Obama's proposed Race to the Top program as "admirable", the NEA's response to the details of the program was mostly critical (National Education Association). They claim that the RTTT program ultimately "misses the mark" and list several suggestions for its improvement in their official response to the proposed legislation.

Many point to the NEA's consistent focus on education as the reason to trust their guidance on how well legislators uphold and promote these goals. Their positions about the federal government's involvement in public education are not dependent on which party holds political power, as is shown by their willingness to criticize and improve both NCLB and RTTT. This bipartisan approach gives the ratings a kind of credibility and legitimacy for many in the public, and the NEA is thus seen as an appropriate group to identify and relay the important education-related political news for its members and the public alike.

For the last three decades, the NEA has created legislator ratings that help measure education-related preferences and give the public a guide by which they can evaluate their representatives on this issue. Furthermore, these ratings are shown to produce distinctions among the legislators that are more nuanced than simple liberal to conservative divisions (Anderson 2012). This evidence presents a striking case for augmenting any analysis of education-related preferences in Congress with issue-specific interest group scores, as Anderson (2012) suggests. However, it is important to remember that her evidence for the extra-dimensionality of education-based preferences is based only on this one group's scores, whereas her claims about other domain-specific preferences were backed up with multiple and unrelated groups. Thus, even though this group's scores provide preliminary evidence of an extra-dimension, especially when coupled with the anecdotal evidence

described above regarding Congress's historical educational legislation, Anderson's claims about the education dimension are in need of a robustness check.

In the rest of this chapter, I give these claims just such a robustness check. Using item response theory, I proceed to test the usefulness of the NEA's scores in delimiting an education-only dimension that is distinct from the liberal to conservative one found in NOMINATE scores.

3.4. Theory

Throughout all the historical developments in federal education policy described above, the National Education Association has attempted to sort out the nuances in these divisions on behalf of the public in order to give the American citizen a basis by which to evaluate the educational policies that Congress implements. While these divisions have persisted over time, it is not necessarily the case that they split along the traditional liberal to conservative lines that typically characterize disagreements in Congress.

Indeed, many have argued that the subject of education is a special kind of issue that should not be lumped together with other economic or social issues when understanding the positions that members of Congress take regarding it. If preferences over education really differ from the liberal to conservative dimension to a significant degree, then these differences should appear consistently when we test education-related votes using a variety of methods.

Anderson (2012) calculated the raw agreement scores between legislators and the NEA's preferred positions on the bills in each year's rating. She then correlated the resulting legislators' ideal points with their NOMINATE counterparts, and found that they correlated rather poorly on average relative to other issue areas and interest group ratings like those from the ADA and ACU.

I take up a different method of analyzing the votes that the NEA included in their legislator ratings, the Item Response Theory (IRT) model, and calculate similar correlations to see if the education dimension remains visible from this new perspective. The IRT method is a good alternative to Anderson's tests because it allows for different subsets of votes and legislators to be placed on

the same scales. In this way, I can test the full roll call record[14] against the much smaller subset of votes chosen for inclusion in the NEA's ratings to find out if the resulting ideal points for the legislators really do indicate separate dimensions. Given that Anderson (2012) points to the NEA as an organization that provides domain-specific insights above and beyond the general explanatory powers of NOMINATE scores, the ideal points estimated from the full set of roll calls should be significantly different than the ideal points estimated using only education- or NEA-related roll calls.

It is important to note that, while the NEA prioritizes education when calculating their legislator ratings, they place additional emphasis on labor and union related preferences when grading these politicians. "We believe it important to hold lawmakers accountable at a time when both public and private sector unions and the right to bargain collectively are under severe attack" (NEA Legislative Report Card). Thus, the votes they choose to include in their ratings are not strictly education-only votes.

For example, the 7th roll call vote included in the ratings of the 113th Congress was for a Worker's Rights amendment to the Military Construction and Veterans Affairs and Related Agencies Appropriations Act for fiscal 2014 proposed by Rep. King (R-IA). "The proposed amendment would have prohibited the use of funds made available by H.R. 2216 to uphold the Davis-Bacon Act and its prevailing wage requirement...A 'no' vote supported the NEA position" (National Education Association). This vote is clearly centered around the union-related preferences of legislators, rather than anything to do with education. In fact, a look at the summary names[15] of the votes they include for the 113th House shows that only about half of the votes they include have any kind of direct link to federal education policy, while a full nine out of twenty-two votes are about

[14]When it uses the full roll call record, the IRT model estimates ideal points for the legislators that correlate with NOMINATE scores at a minimum of 95% on average.

[15]*House:* 1. Domestic Spending Cuts, 2. Women's Rights, 3. Worker's Rights, 4. Fiscal 2014 Budget, 5. Fiscal 2013 Sequestration, 6. College Affordability, 7. Worker's Rights, 8. K-12 Education Policy, 9. Child Nutrition, 10. Fiscal 2014 Funding, 11. Government Shutdown, 12. Fiscal 2014 Funding, 13. Fiscal 2015 Budget Alternative, 14. Fiscal 2015 Budget Alternative, 15. Fiscal 2015 Budget, 16. Charter Schools, 17. Charter Schools, 18. Charter Schools, 19. Background Checks on Gun Purchases, 20. Workforce Development, 21. Tax Fairness, 22. Immigration (National Education Association)

the budget. This is by no means a unique set of choices for the other Congresses and chambers in my analysis.

The votes they choose to include illuminate the actual priorities that the NEA gives to the many topics they grade legislators by. This has important implications for Anderson's argument about using the NEA as the education-dimension experts. If they are not using a strict focus on education alone to inform their ratings, then, even if their ratings *are* distinct from NOMINATE scores, it does not necessarily indicate an *education* dimension. They would instead be measuring and reporting some other dimension of potential import, but this one would be completely subjective and unreliable for scientific purposes in measuring legislative preferences about educational issues.

I therefore examine both the NEA-related votes and then the much smaller subset of the education-only votes in my analysis. In order to isolate the strictly educational votes, I use the Comparative Agendas Project data, which is compiled by political experts who examine each roll call vote and categorize them by topic. "Using proven measures of policy-making activity occurrences allows researchers to assess trends in policy-making activities across time and among nations. CAP's consistent system uses 20 major topics and 200+ subtopics to code those activities" (Comparative Agendas Project). These data provide a consistent and reliable measure of which roll calls are actually related to education, and they allow me to identify if the education dimension does in fact exist without relying on the subjective evaluations of an interest group whose motivations are unclear.

These tests of the different subsets of education-related votes provide a robust answer to the question of whether the education dimension really exists, and additionally speak to the propriety of using the NEA's ratings to establish such a dimension as in Anderson (2012). Given that they potentially use other topics as criteria by which they grade legislators, the relative priority that they ascribe to education is not necessarily clear.

3.5. Data and Methods

I examine the votes chosen by the National Education Association from the 110th to the 113th Congresses[16], both House and Senate, in order to answer these questions. I estimate ideal points for various sets of education-related votes and correlate them with the ideal points that result from using the votes chosen for inclusion in the NEA's report cards.

In order to start this process, I merge the NEA's scorecards with the PIPC and the Comparative Agendas Project data sets which describe the subject-matter of the roll call votes taken in each Congress. By merging these data sets, I identify the subjects by which the NEA rates members of Congress. I call these the "NEA-related" subjects, and I use them to create a pool of NEA-related roll calls from which the NEA could have chosen within each Congress. Given the NEA's claim to focus their attention on education, labor, and unions, it would be reasonable to expect their choices of what to include to be centered around those two topics. In practice, however, the NEA chooses votes that span the length and breadth of the possible topics outlined in the Comparative Agendas Project subject set.

Table 3.3 shows the breakdown of major topics by which the NEA rated legislators across the 110th-113th Congresses, the number of votes in each subject that were included in their ratings, and the number of possible votes with those subjects that they could have chosen from during these Congresses. Over the four Congresses in my analysis, the NEA chose votes from fifteen out of twenty-one possible categories[17], resulting in a pool of NEA-related votes from which they could have chosen that is in fact relatively large.

After compiling these data together, I first estimate ideal points for each member of Congress using the full set of roll calls available over the two-year span. For this and all the other estimations

[16] Although the NEA has been producing and publishing legislator ratings since 1969, they only make the votes they choose for inclusion in these ratings available starting from the 110th Congress.

[17] Excluded categories include: Agriculture, Energy, Environment, Housing, International Affairs, and Culture.

Table 3.3: Vote Counts for NEA-Related Topics, 110th - 113th Congresses

	House		Senate	
Topics	NEA-Chosen	NEA-Related	NEA-Chosen	NEA-Related
Civil Rights	4	111	4	45
Defense	2	463	2	173
Domestic Commerce	1	238	0	80
Education	14	245	5	36
Foreign Trade	1	30	0	13
Government Operations	9	1121	10	598
Health	9	169	5	116
Immigration	0	46	2	46
Labor	11	117	6	83
Law and Crime	3	33	5	24
Macroeconomics	11	202	8	202
Public Lands	0	369	1	41
Social Welfare	3	9	1	6
Technology	1	32	0	9
Transportation	1	158	3	77
Total	**70**	**3343**	**52**	**1549**

in this chapter, I employ a Bayesian 1-dimensional item response theory (IRT) model[18]. The results from this first estimation mimic NOMINATE scores because it uses all non-unanimous votes and is the most informative about the relative differences among the legislators in general. This step is important in order to establish whether or not education is a topic that differs significantly from the regularly understood liberal to conservative dimension when explaining the voting behavior of members of Congress. I use the results of this first set of estimations as a backdrop by which to compare the results from the votes chosen by the NEA.

Once I have the ideal points for the full record ready to go, I estimated ideal points using the NEA-related votes described above from which the group could have chosen for their ratings. In addition, I estimate legislators' ideal points using only the votes that were actually selected by the NEA for that Congress. Finally, I take 100 random samples of twenty NEA-related votes and use those to estimate ideal points for the average NEA-related set of votes that use a comparably small

[18]For the House, Representative Broun (R GA-10) is set equal to 1 to set the polarity of the estimations. In the Senate, the right pole is set to Coburn (R OK). These do not change for any of the Congresses in my analysis.

number of votes as the ratings themselves do. This procedure results in six sets of correlations for each House and Senate over the four Congresses in my study: (1) NEA-related to NEA-chosen, (2) NEA-chosen to NEA-Random 20, (3) NEA-related to NEA-random 20, (4) NEA-related to all votes, (5) NEA-chosen to all votes, and (6) NEA-Random to all votes.

I then focus on the pool of votes that the Comparative Agendas Project data set identifies as "education-related". These votes focus strictly on education, and exclude other topics that the NEA uses to grade legislators, such as union and labor-related preferences, among others. I repeat the process described above to get the same kinds of correlated ideal points, but this time with a strict focus on the education-only votes. This set of correlations speak directly to the NEA's "commit[ment] to advancing the cause of public education" (Anderson 2012, p1188). The results of all these analyses are presented in the next section.

3.6. Results

Table 3.4 shows the results of my initial analysis of all NEA-related votes. These are all the votes that have to do with a subject on which the NEA graded legislators. The NEA could have chosen any number of these votes to include in their ratings, so it is important to understand whether they are choosing votes that are representative of the population[19]. I also compare their choice of votes to the full record to see if the education dimension as measured by the NEA appears different than the liberal to conservative dimension.

Table 3.4 reports the correlations first between all NEA-related votes from which the group could have chosen to those they actually chose to include in their ratings. As seen from the first column, for both the House and the Senate these correlations are extremely high on average with all Congresses correlating at either 95% or 96%. The second and third columns relate each set of ideal points estimated from the NEA-related votes and their subsets to the ideal points resulting

[19]When correlating ideal points such as these, it is important to consider the sizes of the subsets of roll calls by which they are estimated. For this reason, I perform a robustness check on each of the subsets analyzed in Table 3.4 in which I compare each set to the average ideal points estimated from 100 random samples of 20 NEA-related votes. These results are reported in Appendix B and do not result in different conclusions from the ones drawn here.

Table 3.4: Correlations for All, NEA-Related, and NEA-Chosen Votes

Congress	Related to Chosen	All to Related	All to Chosen
House			
110	0.95	1.00	0.95
111	0.96	1.00	0.96
112	0.95	1.00	0.95
113	0.95	0.99	0.95
Senate			
110	0.96	1.00	0.96
111	0.96	1.00	0.95
112	0.95	1.00	0.95
113	0.96	0.99	0.96

from all non-unanimous votes taken in the Congress, whether or not they are NEA-related. Again, these correlations are each extremely high for all three subsets in both the House and the Senate. The correlations between all to NEA-related votes are almost uniformly 1:1, while those between the votes that the NEA actually chose to all the votes are the lowest across all Congresses and chambers in my analysis, but they still average at 95% or above.

Using the subsets and tests from Table 3.4, I do not find evidence for an education dimension that is separate from the traditional liberal to conservative dimension. The subsets of NEA-related votes correlate extremely well with the full roll call record from which NOMINATE scores are estimated, indicating that these votes divide along the same traditional liberal to conservative lines that separate most other legislative subjects throughout American history. This could be because the NEA-related set focuses on many more topics beyond education or labor, and includes almost all the other major topics that the Comparative Agendas Project identifies in American law-making. From Table 3.3 it is clearly seen that the NEA does not simply use strictly education-related roll calls when grading legislators. In fact, the subjects by which they rate members of Congress span almost every aspect of government.

Since it is somewhat less than discriminatory in the subjects by which it grades members of Congress, the NEA may not be the best group by which the education dimension is measured.

Table 3.5: Correlations for All, Education-Related, and NEA-Chosen Votes

Congress	Related to Chosen	All to Related	All to Chosen
House			
110	0.94	0.96	0.95
111	0.92	0.95	0.96
112	0.91	0.96	0.95
113	0.88	0.93	0.95
Senate			
110	0.91	0.95	0.96
111	0.95	0.96	0.95
112	NA	NA	0.95
113	NA	NA	0.96

Notes. The 112th and 113th Senates do not have enough non-unanimous education-only votes to estimate ideal points, so their correlations cannot be reported here.

Even if Anderson (2012) is right to point to education as a distinct dimension with nuances that go beyond strict liberal to conservative divisions, the set of NEA-related votes does not reveal them well. But this does not necessarily mean that there is no education dimension. Rather, it just means that the NEA may not be the best resource to use when identifying the relevant educational votes.

For this reason, I perform a secondary analysis of the education dimension to see if education appears distinctly when examining votes that are strictly education-only, as defined by the Comparative Agendas Project. Table 3.5 reports the results of this next analysis.

The results from Table 3.5 are very similar to those from Table 3.4, showing high levels of correlations between the education dimension and the liberal to conservative one as measured by all votes. Although the average level of correlations is lower for the strictly educational vote set than for the broader NEA-related set, the difference is only really between the lower and upper halves of the ninetieth decile across the board.

The correlation of the ideal points from all the education-only votes to the ones the NEA chose, reported in the first column, are the lowest of the all the sets of correlations in the table. And yet

they are still fairly high, especially when compared with the much lower correlations of the NEA and NOMINATE scores reported in Anderson (2012). The relation between education-only votes and the full roll call record is well into the nineties as well, enough to dispel any remaining doubts that they should be measured on separate dimensions.

Even with a strict focus on only education related votes, a distinct education dimension fails to appear. The subject of education in the legislative process divides legislators along the exact same pivotal points as any other liberal or conservative issue. Anderson (2012) may be right to point to other dimensions like agriculture, foreign policy, business, and the environment as interesting sub-dimensions only accessible by looking at specific subsets of votes. But my analysis shows that education should not be included in this list at least for modern Congresses.

3.7. Conclusion

"There are two human inventions which may be considered more difficult than any others—the art of government, and the art of education; and people still contend as to their very meaning"(Kant 1900, p.12). Perhaps the reason these two arts are so difficult to master is that they require similar sorts of considerations. This at least seems to be the case when American legislators create laws about both governing in general and education policy at large.

Although the educational policies that have been implemented by the federal government over the last century have been far-reaching, affecting the funding, organization, and curricula of American schools at every level, these seemingly specific considerations overlap with the concerns of other completely unrelated legislative topics. The similarity of education to governing in general is evident to the point where an almost ordering of legislators is found whether you look at who agrees with whom in educational policy-making or in policy-making at large.

In this chapter I examined several subsets of roll call votes in order to identify the dimension by which legislators decide educational policies. While several scholars have argued that this field is a special issue area comprising new, unexpected, and different divisions among legislators than we typically use to characterize the members of Congress (Clausen 1973, Wilcox and Clausen 1991,

and Anderson 2012), my analyses do not support this conclusion. Educational voting usually falls along the same division lines as those we see on the traditional liberal to conservative dimension, and the two dimensions correlate at or above 95% on average.

The Item Response Theory model provides a different way to evaluate these older claims, offering a robustness check to the usefulness of issue-specific interest group scores in political science and for understanding the legislative process in general. The model allows for the interest group's ratings to be placed on the same scale as other subsets of the roll call record as well as the full record itself. In this way, a different comparison can be made from the kind found in Anderson (2012), who correlated the full record ideal points with the raw scores reported by the interest groups in her study.

That the two methods reveal different conclusions about the relative uniqueness of education-related votes is an important point. The field of political science can only really be counted as a kind of science if repetition, replication, and robustness checks like the one found in this chapter are performed routinely on previously "settled" questions. Given that there is now evidence both for and against the existence of an education dimension, this question should be continue to be reworked using these and other methods until we as a field come to an agreement about the best way to measure and answer it.

The work in this chapter can go further by taking on the other examples of issue-specific ratings that Anderson (2012) identifies to see whether the IRT method supports their existence or not. The other issue-areas she pointed to in her article were each supported by evidence from the analysis of multiple groups, so I do not anticipate that the IRT method will contradict them. The education issue-area was the only one supported by only one interest group's scores, which was why my robustness check began there. Still, as good scientists, replication and repetition are the only way to be sure our conclusions about the political world are trustworthy and reliable, so these other issues should also be put to the test using the IRT model and other methods.

In addition, this work can be extended by an examination of a larger timeframe than the one I used in this chapter. My work here was limited by the availability of the votes used in the NEA's

ratings, which were only available beginning in the 110th Congress. Other interest groups who make their ratings publicly available for longer periods of time may provide insight into the historical nature of issue-specific policy dimensions. It is possible that over a larger timeframe even the education dimension would exhibit some distinct nuances relative to the liberal and conservative dimension that it mimics in modern Congresses.

CHAPTER 4: ABBREVIATED IDEOLOGY[20]

[20] Kisaalita, Alice E. To be submitted to *Journal of Politics*.

4.1. Abstract

This chapter tests the limits of the w-NOMINATE method for estimating ideal points for small subsets of roll call votes and small numbers of legislators. I establish that the recommendations of Poole and Rosenthal for the minimum number of votes and voters to include can be lowered from 100 votes and 50 voters to as few as 11 votes and 4 voters, while still producing ideal points consistent with those estimated from the full roll call record. Additionally, I test whether certain kinds of votes exert undue influence over estimations of ideology, and find that most, but not all, vote types produce consistent results. This study shows that the w-NOMINATE method can be used to analyze legislatures, subsets of votes, and voting bodies which are much smaller than those to which it is currently applied, such as bureaucracies, committees, and interest group ratings.

4.2. Introduction

It is often necessary to consider legislative bodies with small numbers of votes and even fewer voters when thinking about important questions for understanding the legislative process. For example, many of the earliest Congresses in US history comprised as few as 18 members voting on less than 50 votes in a two year period. The small legislature sizes reflected the difficulties of travel in the 18th and 19th centuries coupled with the fact that there were only 13 states to govern. Still, political scientists today use the same techniques to analyze these early legislatures as they do for Congresses that contain hundreds of legislators who take thousands of votes each year.

There are many examples in which others desire to analyze small numbers of voters and votes, including congressional committees, bureaucracies, judiciaries, and issue-specific votes among many others. For example, several interest groups have stepped up in order to aid interested viewers in understanding this complicated legislative process. Over the last four decades and more, interest groups, like the Americans for Democratic Action (ADA) and the American Conservative Union (ACU), have based their legislator ratings on the premise that the full record may not be ideal for all circumstances. Both groups claim to pick out the twenty or so most important liberal or

conservative votes in each chamber of Congress that are taken in a calendar year, and then use only those twenty votes to calculate their member ideology scores.

Even the analysis of large modern day legislatures suffers from some methodological problems when the same technique is applied across the board without regard to the nuances within each Congress. For example, over the course of his presidency, President Obama dealt with a string of Republican-led attempts to repeal, replace or alter his signature piece of legislation, The Affordable Care Act. Republicans in the House of Representatives voted more than fifty times on the record to repeal the bill, all of which were unsuccessful in the Senate. Each of these votes are identical to one another, and contain exactly the same information about where each Representative stands on the issue of Obamacare. And yet, the Republican-led House members decided to include over fifty of these identical votes in order to make sure that their constituents were aware of their position on the matter.

This sort of repetitive signaling is a commonly used political tactic that allows them to say that they tried their best, even though there was never any possibility that they could pass the Democratically-led Senate, let alone over a Presidential veto. As Mayhew famously claimed, "We can all point to a good many instances in which congressmen seem to have gotten into trouble by being on the *wrong* side in a roll call vote, but who can think of one where a member got into trouble by being on the *losing* side?"(Mayhew 2004, p118). However, a consequence of this tactic is that the roll call record by which voters, interest groups, political scientists, and many others learn about the legislative process becomes inundated with redundant information. For these discerning viewers, a picture of the relevant distinctions among legislators might include only one of those fifty votes.

This chapter begins with the premise that it is sometimes necessary to examine smaller subsets of roll calls than is typically found in the most common techniques used to estimate legislators' ideology. The most widely used technique among political scientists, NOMINATE scores, is not

recommended for legislatures with fewer than 50 legislators and 100 votes[21]. Yet it is still used to estimate ideal points for those early legislators, which are then used to inform and make conclusions about important political questions for that time. Whether because the voting body itself is smaller or because specific subsets of votes are selected for study, examining smaller legislatures is a key part of understanding the legislative process.

I will use the this chapter to consider more fully the idea of a mathematically derived standard for choosing and analyzing a smaller set of roll calls for use in ideological calculations of the US Congress. Of course, the most informative method of calculating ideology will be the one that includes the entire roll call record, as in NOMINATE scores. But there is a trade-off in terms of including repetitive, erroneous, miscellaneous, and other votes unrelated to legislators' ideology that comes with the full set of roll calls versus including just enough of those roll calls needed to give a good picture of ideology for the measure to be both complete and efficient.

The main focus of this chapter will be to find out if smaller subsets of votes result in a new ordering of legislators than the one found in NOMINATE. If so, are these orderings better, worse, or functionally equivalent? Do they solve the problem of repetitive or extraneous roll calls, which can muddy the ideological waters of the full roll call record, so to speak? If the answer to any of these questions is "yes", then this chapter will have contributed to the project of measuring politician's ideologies by providing a mathematical standard for distinguishing important or useful votes from the plethora of available options in the roll call record each Congress. The insights for dealing with subsets of roll calls can then be applied to other contexts, in which small numbers of positions are more common.

---

[21]Poole, on the accuracy of the parameters estimated using NOMINATE: "Howard Rosenthal and I performed a large number of simulation studies of our NOMINATE model and found that it performed very well (Poole and Rosenthal, 1991; 1997). Our approach was to create artificial data that met all the assumptions of the model and then test how accurate the method was in reproducing the true legislator ideal points and roll call parameters. We found that when the number of legislators was 50 or more with 100 or more roll calls, the recovery of the true parameters by NOMINATE was very good. We concluded that bias was not a serious problem with reasonable-size roll call matrices." Keith T. Poole. Spatial Models of Parliamentary Voting (Analytical Methods for Social Research) (Kindle Locations 1603-1606). Kindle Edition.

4.3. Literature Review

Roll calls have historically been used to reveal the ideology of the legislators who voted on them. Methods of estimating ideology, both simple and complex, rest on the assumption that roll calls provide enough information about legislators' revealed ideological preferences to create a realistic, practical, and useful ordering of legislators from liberal to conservative. While roll call votes (also known as recorded votes) are by no means the most common type of voting found in Congress (see Lynch & Madonna 2013), they have become more broadly used in recent decades[22]. These are the only kind of votes that Congress members take to which they attach their own individual names[23]. As such, they are the only data we have available to analyze the individual voting behavior of our elected representatives. For this reason, roll calls have been used to calculate legislators' ideologies in a variety of ways.

4.3.1. The Full Roll Call Record

Most prominently, the NOMINATE method, developed by Keith Poole and Howard Rosenthal in the 1980s, uses the full set of recorded votes in a Congress to explain the most variation in roll call voting with the fewest dimensions possible[24], and they find evidence for a stable liberal-conservative first dimension that explains most of the variation in voting behavior (see Poole and Rosenthal 1985; Poole and Rosenthal 1991a; Poole and Rosenthal 1991b; and Poole and Rosenthal 1993, among others). When coupled with a second dimension, an average of 89% of the votes in a given Congress can be correctly classified based on the estimated ideological location produced by the NOMINATE model (Bonica 2013, p300). For its incredible explanatory power, coupled with

[22]Technological improvements and the reforms of the 1970s which allowed for roll call votes in the Committee of the Whole have contributed to the increased use of recorded votes. Oleszek 2014 describes the ease with which they can now be obtained, saying, "Recorded votes...usually can be obtained easily; only a sufficient second – one fifth of [those] present – is needed, with a minimum of 11 as required by the Constitution" (Oleszek 2014, p300).

[23]Other vote types, including voice votes, division votes, and teller votes, all produce legislative outcomes without the associated members being responsible for their individual contributions to that vote because only the outcome of the vote is reported in the Congressional Record.

[24]Usually two, sometimes even one, suffice to explain most of the observed voting behavior in Congress.

the fact that Poole and Rosenthal made their scores available to the public for free, the NOMINATE model has become the standard way by which political scientists measure legislators' ideological preferences.

But scholars have largely used these scores without fully understanding how they are developed. For example, since the legislators do not vote on the same set of questions in each Congress, NOMINATE scores need to be analyzed carefully if they are to be compared across time. While DW-NOMINATE connects the scores through time by linking legislators who served in more than one Congress, changes in the content of the roll call votes are not accounted for by the process. So, a "liberal" in the 1950s could believe in very different policies than a "liberal" in the 2000s, even though they may have the same NOMINATE score on a -1 to 1 scale (Bateman et al 2017).

In addition, the history of the roll call record shows an increase in the number of roll call votes taken in each Congress, but this does not necessarily mean that legislators are voting ideologically more often than in older congresses. Any explanation of the impact of the increased number of votes on ideology needs to be conditioned on the convenience and ease with which roll calls are taken in modern congresses, given improvements in modern day technology. This provides another reason to be cautious when comparing older congresses to newer ones.

The picture of ideology produced by NOMINATE scores may also be skewed by the legislators themselves, if, for example, they stack the roll call record full of miscellaneous, repetitive, or spurious votes that are specifically designed to exaggerate the similarity or differences between members of Congress. For instance, majority party leaders will often include roll call votes on the approval of the journal for the specific purpose of letting party members vote against their leadership on a vote that does not actually produce any meaningful or substantive policy changes (Crespin et al 2011). This tactic decreases a rank-and-file member's party unity score which can be used against them in elections if it is too high (see Canes-Wrone et al 2002 and Carson et al 2010). These journal votes are included in the regular NOMINATE scores and contribute to a member's placement on the ideological scale, even though they do not necessarily reflect that member's liberal or conservative beliefs in any way.

Finally, NOMINATE scores may be imperfect measures of ideology due to the selection bias that lies inherently within the roll call record upon which the scores are based. Legislators do not put their name on every vote that they take. The votes which get a roll call may be strategically chosen according to a legislator's own motivations, revealing only a portion of his or her ideological preferences. The point is that because legislators have a role to play in deciding which votes they attach their names to and to which they do not, the votes themselves may not always be perfect proxy measures of the legislators' true ideological preferences about a given topic. Other factors besides their dedication to conservative or liberal values may influence how they vote, like party pressure, constituency pressure, interest group and lobbying pressures, and many others.

One important consequence of these complications is that the full roll call record may not always be the best set of votes to use in order to fully understand the nuances inherent in legislators' ideologies. NOMINATE scores produce a single summary statistic for a legislator's behavior over a two year period in which the legislator may have voted hundreds if not thousands of times. While research has shown that this single statistic has extraordinary predictive power, it does not apply in all cases or perfectly explain a legislator's voting behavior.

4.3.2. Subsetting the Roll Call Record

Although the NOMINATE model still predicts the votes of legislators across almost all issue areas relatively well, a potential problem can still arise in discerning the underlying motivation for those votes. This is because voting behavior is not necessarily always motivated by the liberal to conservative dimension, even if that same dimension can predict the vote fairly well. Thus, there may be some merit to the idea that a smaller set of roll call votes can better explain specific votes or issue positions than the full roll call record on which NOMINATE scores are based.

Poole discusses the possibility of subsetting the roll call record and its consequences for the NOMINATE method in his book, Spatial Models of Parliamentary Voting. When developing the method, he and Rosenthal performed several tests with subsets of the roll call record and with simulated roll calls in order to test the validity of their process and the limitations of the model.

"Our approach was to create artificial data that met all the assumptions of the model and then test how accurate the method was in reproducing the true legislator ideal points and roll call parameters. We found that when the number of legislators was 50 or more with 100 or more roll calls, the recovery of the true parameters by NOMINATE was very good"(Poole 2005, ch4). While it is not clear from this quote exactly which measures Poole and Rosenthal used to establish the meaning of the "very good" recovery of parameters by their method, still the use of the NOMINATE model to analyze subsets of the record is an appropriate way to proceed as long as these minimum parameters are met. However, most applications of the NOMINATE model apply strictly to the full roll call record.

The question remains as to when and why we should ever subset the record in the first place. "Returning to the question..why we should scale *subsets* of roll calls at all?...I think it is a useful tool to uncover what is going on when *structural change* is occurring, either in the electorate that is selecting the legislature or in the legislature itself" (Poole 2005). Poole follows this claim up with the example of the pre- and post-Civil War congresses, wherein votes on the issue of slavery switched from second dimensional ones to votes which mapped onto the first dimension. This switch accompanied the disintegration of the Whig Party and the electoral realignment that produced the Republican party, resulting in a huge change in the structure of the congressional voting. Thus, when the structure of voting itself changes, rather than the minute changes from Congress to Congress on exactly what it means to be a liberal or a conservative, then an analysis of a smaller set of roll calls is in order to track such changes. This is the only case identified in Poole's book as an appropriate time to subset the record.

The analysis of smaller historical Congresses is thus the textbook example of when to subset the record. Whether because of structural change in the ideological make-up of its members, as in Poole's example above, or because there are simply not enough votes and voters, as in the earliest Senate chambers, gaining insight from historical Congresses provides an important reason to test how well the analysis of small legislatures holds up within the NOMINATE process. Without this

more thorough investigation, comparisons of early Congresses to later ones may contain erroneous conclusions even though the same methods are applied to the estimation of each set.

In addition, others have identified further reasons to analyze less than the full set of available votes. Wilcox and Clausen (1991), for example, discuss how legislators bring different reasons to bear on their voting behavior for different issue areas, even if they end up voting with the same set of people in both cases. The reason a legislator may vote one way or another on issues related governmental involvement in the economy, for instance, might very well be due to his or her beliefs as they map on to a liberal to conservative policy dimension. But, that does not necessarily mean that the same legislator uses liberal or conservative values as the basis for *all* of their votes across *all* issue areas. The same legislator may vote completely differently on agricultural issues based on the urban or rural composition of his or her constituency. Clausen (1973) sorted through and classified all roll call votes from the 1950s and 1960s into issue-specific categories and found that legislators line up in different rank orders depending on which issue area was being voted upon.

Interest groups like Americans for Democratic Action (ADA), in addition, have used a small selection of roll calls to provide brief, summarized accounts of legislators' ideologies since the 1940s and they continue to do so today. At the time of this writing, Project Vote Smart, a group who tracks interest groups that rate members of Congress, counted fifty-eight groups dedicated solely to producing measures of legislators' overall conservative or liberal ideology[25] from a smaller number of votes[26] than the full record (Project Vote Smart).

Given that as few as twenty votes are often analyzed in our modern day political context, it is important to identify the limitations of subsetting the record more clearly. Can twenty votes ever produce a picture of ideology that is consistent with the more informative full record? If not, how few are needed for consistent estimates? And are there any specific kinds of votes that do a better (or worse) job of producing consistent ideal point estimates? I plan to investigate these questions in the rest of this chapter.

[25] As opposed to groups who concern themselves with measuring legislators' ideological beliefs about a special issue area, like the environment, abortion, or gun rights, for example.

[26] Usually between twenty to thirty.

4.4. Theory

Given that Poole and Rosenthal claim that 100 votes are the minimum required to produce "very good" estimates of legislators' ideal points, I do not expect that any fewer votes will sufficiently reflect the true ideological spectrum. However, many important political variables are measured with as few as twenty or thirty votes. Therefore, I will begin by reproducing the analysis of fewer than 100 votes that Poole began in his book to test the limitations of the model, and I will describe the consequences of this analysis in terms of ideal point correlations to the full record estimates. This will not only provide some insight into the consequences of analyzing subsets of even as few as twenty-five votes to rate members of Congress, but will also produce meaningful and reproducible measures by which we can more exactly define the limitations of the NOMINATE model.

I posit that with a random sample of 100 votes from a given Congress the ideal points should highly correlate with the full record. In this way, I plan to reaffirm Poole's conclusions about the minimum number of votes needed for consistent estimates.

From the analysis of 100 votes, I move to a subset with only twenty-five votes. If a random sample of twenty-five votes gives noticeably uncorrelated ideal points on average, then we can be sure that twenty-five votes are just too few to analyze the ideology of members of the US Congress, just as Poole and Rosenthal claim. If, on the other hand, the average random sample of twenty-five votes produces the same results as the full roll call record, then any significant deviations found in the interest groups' ratings, committees, bureaucracies, state legislatures, historical Congresses, etc. that use twenty-five votes are disingenuous misrepresentations of the legislature.

Furthermore, certain kinds of votes may influence measures of consistency. Only including straight party-line votes, for example, could result in estimates that exaggerate levels of polarization. After establishing the minimum number of votes that are necessary to produce consistent estimates, I take advantage of the PIPC dataset produced by Crespin and Rohde which categorizes all roll calls in a given Congress into vote types, such as final passage, amendment, and procedural

motions. I can subset the roll call record into each of these categories as long as they have the minimum number of votes for consistency. In this way, I can see if a specific subset of votes differs significantly from both the minimum consistent average random sample and from the full record.

I posit that procedural votes[27] in the House will be heavily monitored and controlled by party politics relative to substantive votes, leading to ideal point estimates that are more polarized than the full record and that correlate poorly with the full record. This analysis will illuminate the link between different kinds of votes and their effect on overall estimations of ideology.

Finally, I test the other half of Poole's statement, that at least 50 legislators are needed to produce consistent estimations. If what he claims is true, then any analysis of the early Senates, which had 26[28] members many of whom were not always perfectly in attendance, should be impossible. Yet it is still done, and the ideal points of these senators are publicly available for use in the scientific study of these early Congresses. I will test how consistently the NOMINATE model reproduces the ideal points of legislators when there are less than 50 available. This is an important step in delimiting the usefulness of this technique for smaller voting bodies.

4.5. Data and Methods

I analyze the roll call record of the US Congress from the 96th to the 115th Congress. For each congress, I take one hundred random samples of one hundred votes and run them through the w-NOMINATE model in one dimension. In each Congress, I preprocessed the votes to remove unanimous and near-unanimous votes before sampling. In this way, I ensure that my samples would only include scalable votes and they could be estimated without further loss. I then compare each sample to the results from the full record in terms of correlation of ideal points. Finally, I repeat this process for one hundred samples of twenty-five votes.

[27] I include in this category such things as motions to recommit, passage of rules, approving the journal, previous questions on special rules, etc. Basically anything that is not an amendment or final passage vote I call procedural.

[28] In fact, the fewest number of legislators estimated in the earliest Congresses is 29. This is due to the election of New Jersey's William Paterson to Governor on November 13, 1790, and to the death of Virginia's William Grayson on March 12, 1790, who was replaced via appointment by John Walker. John Walker was then replaced via election by James Monroe on November 9, 1790.

After establishing the minimum number of votes needed for consistent estimates, I take advantage of the PIPC data, which codes each roll call for its type. I subset the votes in each congress into each category and estimate ideal points using only the votes within each category[29]. I then compare the correlation of the ideal point estimates from each category to the full record to see if any category in particular deviates significantly from the full record.

Finally, I repeat the process used to establish the minimum number of votes and apply it to the analysis of legislators. I keep the number of votes consistent and simulate the ideal point estimation procedure for different numbers of legislators from a minimum of 3 to the recommended minimum of 50 for each Congress. I correlate the ideal points from each subset of legislators to their estimates from the full set to pinpoint the minimum number that works with this method.

4.6. Results

4.6.1. How Many Votes Are Needed for Consistency?

In this section, I find that voting records with as few as 11 votes can produce consistent and reliable estimated ideal points. This is good news for anyone interested in evaluating interest group ratings, small committee-level voting behavior, and other legislatures that vote on a limited set of issues. Before I exhibit these results in full, however, a short digression is warranted.

In order to find out how many votes are needed to produce consistent estimations, I delved very deeply into built-in assumptions of the w-NOMINATE model. One of these assumptions is that there must be 20 scalable votes for the w-NOMINATE program to estimate ideal points. It is important to note that this is simply the default setting, and there is no theoretical underpinning for why it is set at 20 and not some other value. Indeed, given that Poole recommends at least 100 votes (Poole 2005), it would make more sense for this number to be set to 100 instead of 20. When using the w-NOMINATE function in R, this assumption is programmed into the default settings as the "minvote" parameter, and the user can call w-NOMINATE function without ever knowing that

[29]Categories are estimated if they have more than 11 votes per congress after filtering out unanimous and near-unanimous votes.

this parameter exists, let alone changing it. But it can also be changed and manipulated like any other variable, and it turns out that consistent estimations rely heavily upon this particular choice.

I focus so closely on the minvote parameter due to an important discovery in my first attempts at finding the basic, most fundamental conditions needed for consistent estimations. I initially ran random samples of twenty and one hundred votes through the w-NOMINATE package in R and set the default minimum number of votes equal to five. This quantity, the minimum vote parameter, is the number of votes that a legislator has to take a yea or nay position on in order to be included in the estimated ideal points. The default for this parameter set into the w-NOMINATE package in R is set to 20 votes, but I changed it to 5 for my initial estimation of 20-vote samples. The choice to set this parameter to 5 was an arbitrary choice motivated by the desire to include as many of the twenty votes in each sample, but this choice also proved to be very illuminating.

When the minimum number of votes was set to five, the correlation of ideal points to the full record estimates was very low on average across all the congresses in my data. The Pearson's correlation coefficient was 0.05 on average for one hundred random samples of 20 votes per congress and 0.22 on average for one hundred random samples of 100 votes per congress. This prompted me to call into question Poole's claim that even 100 votes were sufficient to produce consistent ideal point estimations, since the correlation to the full record for the average 100 vote sample was barely reaching over 20%.

I proceeded to investigate whether this problem persisted when the samples of twenty and one hundred votes were re-estimated in two dimensions. I kept every other parameter the same, including the minimum vote requirement of only five, and I only switched the dimensions from 1 to 2. At this point, the w-NOMINATE program in R had trouble converging. Once I switched the minimum vote requirement from 5 to 15, the program ran smoothly for estimating 100 vote samples in two dimensions. This change resulted in a correlation coefficient of 0.98 on average—a huge leap from the 1-dimensional estimates!

Suspecting that the low minimum vote requirement might be the culprit for the poor correlations in one dimension, I re-estimated both the twenty and one hundred vote samples again in one

dimension with a 15 vote minimum. This time, the correlation coefficients were almost as good as in two-dimensional estimations, even with only twenty votes. The one-dimensional correlation coefficient for 100 samples of 20 votes averaged 0.89 after changing the minvote parameter to 15, and an average of 0.98 for 100 samples of 100 votes. If a minimum of 15 votes could produce such high correlations, then clearly the original default value for this parameter of 20 would be sufficient to produce consistent correlations.

But this left several questions still unanswered. What happened between the minimum of 5 votes and the minimum of 15? Why was there such a large leap in the average correlations of 100 sampled subsets to the full record when that parameter changed? At what point does this change take place? Is the default of 20 votes so much better than the correlations at 15 votes, which reached almost 99% in these samples?

In order to answer these questions, I performed several tests specifically designed to test the minvotes parameter. For each Congress from the 96th - 106th, I set the minimum vote parameter equal to 5, estimated ideal points from the full record, sampled 25 votes 100 times, and estimated ideal points for each sample. I then measured the correlation for each sample to the full record ideal points and recorded the average correlation for all 100 samples. Next, I repeated this sampling and estimation process setting the minimum vote parameter to 6, 7, etc. up to the original default value of 20. So, I ended up with average correlations from 100 samples of 25[30] votes for each of the 16 minimum parameter values for the 96th - 106th congresses. Figure 4.2 displays the average correlations from each of these simulations.

From Figure 4.2 it is clear that when the minvotes parameter is set anywhere from 5 to 10, the average correlation of 25 random votes to the full record is very low[31]. But once the minvotes parameter is set to 11, it the average correlations of samples even as small as 25 votes to the full

[30] I had to change the sample size from 20 to 25 in order to test the default minvotes setting of 20. When the minvotes parameter approaches the sample size, then all legislators who do not meet the minimum vote level are omitted from the analysis. So, if everyone has at least 1 absence in a sample of 20 votes and a minvote setting of 20, then no one can be estimated. Increasing the sample size to 25 solved this problem while still maintaining comparability to an average interest group rating size.

[31] For Congresses 96 - 106, the average correlation for the range of 5 - 10 minvotes is 0.11.

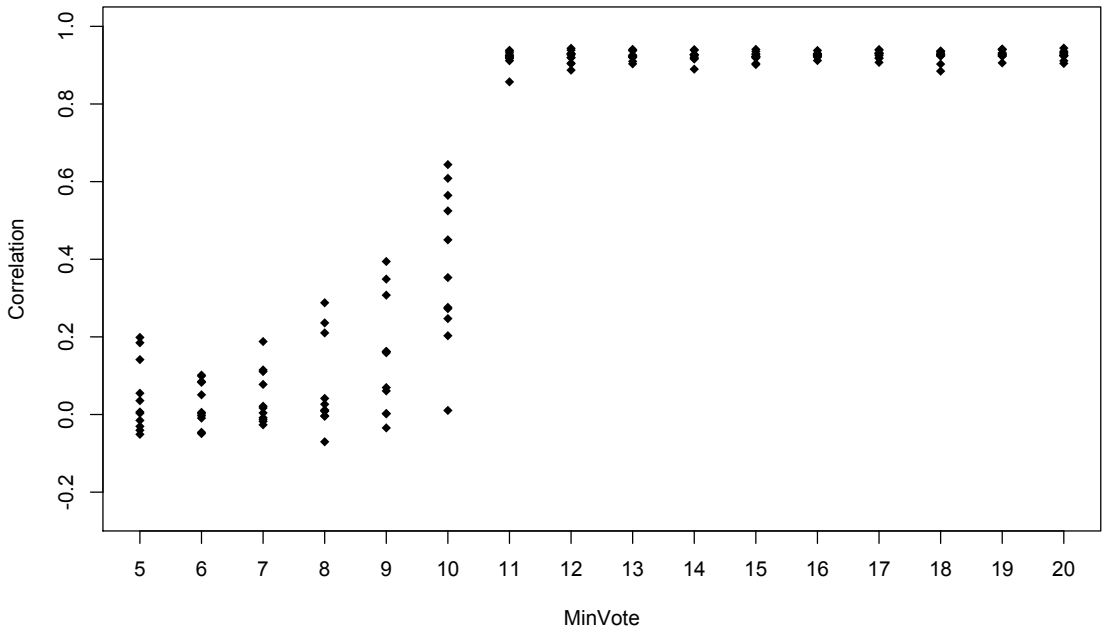

Figure 4.2: Average Correlation to Full Record for Increasing MinVotes

record jumps to almost 92%. The average correlation at 11 minvotes (0.918) is not statistically different from the average correlation at 20 minvotes (0.925) or any other minvotes setting in between.

From the analysis so far we can take away several important conclusions. While 100 votes are certainly sufficient for producing consistent estimates, as Poole and Rosenthal claimed from their initial simulations, samples of roll call votes with as few as 25 can consistently reach correlations above 90% in one-dimensional w-NOMINATE estimations as long as the minimum vote parameter is set to at least 11. There is no significant difference in average correlations to the full record when the minimum vote parameter is set anywhere from 11 to 20, so the default value set into the program can be lowered to 11 safely.

A random set of 25 votes is shown to be consistent with the full record on average as long as the appropriate parameter settings are in place. Because of this consistency, it is not inherently problematic that interest groups, for example, analyze small numbers of votes when they create

their legislator ratings. But any time that an interest group selects votes that produce significant differences in ideal points from the full record, then the legislators' positions that they present to the public would be misleading.

4.6.2. Do Specific Kinds of Votes Make a Difference?

While a random selection of votes usually produces consistency, it remains to be seen if there are specific types of votes that have undue influence over ideal point correlations in the US Congress. To test this, I subset the roll call record non-randomly by using the categories provided by the PIPC data. I estimate ideal points and correlations in each of the 96th - 115th congresses for every category in which the most conservative legislator[32] voted on at least 11 scalable votes.

I discuss the results of these analyses by vote type: first by final passage votes, then amendment votes, and finally procedural votes. The first two categories of votes (final passage and amendments) can be classified as substantive votes, in which legislators express their preferences on the actual content of the proposed legislation. This is opposed to procedural votes, which are roll calls designed to move legislation through the law-making process without necessarily altering the content of the legislation in any way. As I posited earlier, one might expect that the liberal to conservative dimension asserts itself most prominently in the substantive, content-altering kinds of votes, because we usually conceive of ideological differences as meaningful, policy-oriented disagreements. But in fact I do not find any significant difference in these two categories. If anything, in fact, there is more ideological variation in substantive votes on average than in procedural votes.

The results of the analysis for final passage votes are presented in Table 4.6. The average correlations for final passage votes to the full record are relatively high, ranging from 82% to 97% over the course of 40 years. The highest correlation appears in the final passage of regular bills (97%), followed by the passage of regular bills under the suspension of rules (93%). These high numbers are not surprising, given the number of votes taken in these two categories. They have a

[32]This member is required to set the polarity of the estimations and keep this polarity consistent for every estimation in that Congress. I identified the most conservative legislator in each Congress using their DW-NOMINATE scores provided by www.voteview.com.

Table 4.6: Correlation to Full Record by Vote Types – Final Passage Votes

Vote Type	**Avg**	96	97	98	99	100	101	102	103	104	105
Passage - Bill	**0.97**	0.96	0.96	0.96	0.96	0.95	0.95	0.96	0.97	0.98	0.98
		(153)	(102)	(108)	(90)	(105)	(105)	(107)	(101)	(122)	(115)
Passage - Conf. Report	**0.89**	0.96	0.85	0.85	0.82	0.88	0.87	0.79	0.86	0.95	0.88
		(68)	(29)	(41)	(33)	(39)	(44)	(48)	(42)	(49)	(35)
Passage - Res.	**0.87**	0.88	–	0.84	–	–	–	0.85	–	–	0.91
		(68)	(14)	(20)	(15)	(17)	(10)	(22)	(11)	(15)	(22)
Passage - Joint Res.	**0.82**	0.82	0.90	0.72	0.84	–	0.73	–	–	–	–
		(16)	(24)	(13)	(26)	(11)	(12)	(10)	(6)	(8)	(9)
Passage - Bill Susp. Rules	**0.93**	0.94	0.95	0.95	0.92	0.93	0.89	0.94	0.92	0.93	0.97
		(101)	(72)	(104)	(59)	(113)	(66)	(65)	(88)	(76)	(113)
Passage - Con. Res.	**0.87**	–	–	–	–	–	–	–	–	–	–
		(15)	(12)	(13)	(10)	(6)	(5)	(7)	(3)	(8)	(11)
Passage - Res. Susp. Rules	**0.87**	–	–	–	–	–	–	–	–	–	–
		(3)	(3)	(2)	(4)	(15)	(7)	(5)	(5)	(7)	(24)

Vote Type	**Avg**	106	107	108	109	110	111	112	113	114	115
Passage - Bill	**0.97**	0.97	0.98	0.97	0.97	0.99	0.97	0.97	0.94	0.99	0.95
		(138)	(95)	(121)	(114)	(146)	(101)	(118)	(124)	(152)	(153)
Passage - Conf. Report	**0.89**	0.93	0.87	0.94	0.93	0.95	0.90	–	–	–	–
		(44)	(30)	(33)	(27)	(15)	(12)	(7)	(1)	(8)	(6)
Passage - Res.	**0.87**	–	–	–	–	–	–	–	–	–	–
		(6)	(9)	(11)	(10)	(11)	(5)	(10)	(6)	(6)	(5)
Passage - Joint Res.	**0.82**	–	–	–	–	–	–	–	0.84	–	0.91
		(28)	(16)	(8)	(5)	(3)	(3)	(8)	(15)	(7)	(19)
Passage - Bill Susp. Rules	**0.93**	0.97	0.95	0.85	0.97	0.96	0.93	0.97	0.94	0.89	0.83
		(171)	(147)	(163)	(149)	(280)	(303)	(178)	(188)	(195)	(241)
Passage - Con. Res.	**0.89**	–	–	–	–	0.94	0.87	0.79	–	–	–
		(8)	(4)	(9)	(8)	(13)	(18)	(12)	(4)	(5)	(3)
Passage - Res. Susp. Rules	**0.87**	–	–	–	0.83	–	0.91	–	–	–	–
		(34)	(42)	(93)	(108)	(178)	(343)	(9)	(9)	(13)	(18)

Notes. This table shows the estimated correlation of the full roll call record to final passage votes in each Congress on the polarity member participated in at least 11 scalable votes. Numbers in parentheses are the total number of votes taken in that category during that Congress.

relatively high proportion of the total votes taken in a given Congress, and so it makes sense that they also correlate highly with the full record.

Still, of the seven categories of final passage included in the PIPC data five of them correlate below 90%. Given the findings above, showing that a random sample with as few as 11 votes from

Table 4.7: Correlation to Full Record by Vote Types – Amendment Votes

Vote Type	**Avg**	96	97	98	99	100	101	102	103	104	105
Straight Amdts.	**0.99**	0.99	0.98	0.99	0.98	0.98	0.98	0.98	0.99	0.99	0.99
		(376)	(201)	(259)	(255)	(298)	(280)	(276)	(431)	(544)	(362)
Amdts. to Amdts.	**0.89**	0.92	0.95	0.92	0.91	–	–	0.87	0.84	0.89	0.83
		(31)	(43)	(38)	(35)	(12)	(9)	(16)	(14)	(20)	(12)
Substitute to an Amdt.	**0.84**	0.95	0.72	0.81	0.93	–	0.83	0.79	0.90	0.81	–
		(36)	(14)	(14)	(34)	(10)	(14)	(11)	(19)	(16)	(8)
Amdt. to Substitute	**0.87**	0.79	–	0.90	0.89	–	–	–	–	–	0.90
		(15)	(2)	(13)	(15)	(0)	(1)	(2)	(5)	(8)	(32)

Vote Type	**Avg**	106	107	108	109	110	111	112	113	114	115
Straight Amdts.	**0.99**	0.98	0.98	0.98	0.99	0.98	0.99	1.00	0.99	0.99	0.99
		(372)	(224)	(326)	(413)	(459)	(357)	(840)	(452)	(515)	(295)
Amdts. to Amdts.	**0.89**	–	–	–	–	–	–	–	–	–	–
		(10)	(2)	(0)	(2)	(3)	(3)	(0)	(6)	(0)	(0)
Substitute to an Amdt.	**0.84**	–	–	–	–	–	–	–	–	–	–
		(4)	(4)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Amdt. to Substitute	**0.87**	–	–	–	–	–	–	–	–	–	–
		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes. This table shows the estimated correlation of the full roll call record to amendment vote types in each Congress on the polarity member participated in at least 11 scalable votes. Numbers in parentheses are the total number of votes taken in that category during that Congress.

a given Congress correlates at 90% or higher, the finding that many of the substantive final passage votes consistently correlate below this level shows a degree of variation in legislators' ideological positions on these kinds of votes. This finding shows that the substantive content of legislation in its final form can cause legislators to express their actual preferences about the substance of the bill, rather than show a strict adherence to liberal and conservative party pressures in their voting behavior for this kind of roll call vote. Thus, when including these kinds of vote types in a subset of votes for the purposes of estimating ideology, the resulting estimations may include more variation away from the liberal to conservative line than a random draw of votes would produce.

Table 4.7 displays the results of this analysis for amendment vote types. Amendments are another kind of substantive vote in which legislators have the opportunity to express their preferences over the substantive content of legislation. These kinds of votes include straight amendments,

which directly change the text of a bill, and other kinds of amendments, which amend the text of amendments. Straight Amendments correlate at an average of 99% with the ideal points estimated from the full record. These votes make up a large proportion of the total voting record in a given Congress, so it is again not surprising to see a high level of correlation overall in this category to the full roll call record.

The other kinds of amendments correlate less well, from 84% to 89% on average. While there are fewer votes for the other kinds of amendments relative to straight amendments, and almost none to speak of in the later Congresses in my analysis, the ideal points for these amendments are still based on a high enough number of votes to produce consistency. Yet, they correlate slightly less well on average than the random draws of a similar sample size, which averaged above 90%. So, here again we can see some evidence for ideological variation in substantive vote types beyond the average for a random draw from the roll call record.

Finally, Table 4.8 shows the correlations for procedural vote types to the full roll call record. Almost all of these vote types correlate at or above 90%, consistent with a random draw from the full roll call record, and higher, on average, than their substantive counterparts. Only three vote types correlate at less than 90%: Motions to Adjourn (83%), Motions to Recede and Concur (86%), and Motions to Approve the House Journal (21%).

Among all of the vote types I have considered, both substantive and procedural, votes to approve the House journal stand out for their significantly worse correlation to the full roll call record than any other type. The House of Representatives is constitutionally required to keep a journal of their proceedings (US Constitution, Article 1, Section 5). But the way that this requirement has been implemented has changed over time, becoming more strategic over the last fifty years.

> "Since the passage of the Legislative Reorganization Act of 1970,...the Speaker has had the authority to announce his approval of the *Journal* in lieu of it being read. In order to allow for corrections to the *Journal*, any member may request a vote on the Speaker's approval: if such a vote is decided in the negative, then the *Journal* is read and possibly amended (Kravitz 1990, p380). By 1970 the *Journal* and its place in the

Table 4.8: Correlation to Full Record by Vote Types – Procedural Votes

Vote Type	**Avg**	96	97	98	99	100	101	102	103	104	105
Miscellaneous	**0.92**	0.92	–	–	–	–	–	–	–	0.92	–
		(24)	(15)	(16)	(2)	(10)	(4)	(7)	(7)	(22)	(7)
Rise from COTW	**0.91**	–	–	–	–	–	–	–	–	0.93	0.87
		(2)	(1)	(9)	(8)	(4)	(3)	(6)	(8)	(14)	(16)
Passage of Rules	**0.94**	0.90	0.92	0.90	0.95	0.94	0.92	0.94	0.96	0.97	0.97
		(98)	(51)	(74)	(99)	(84)	(74)	(84)	(95)	(103)	(86)
Approve House Journal	**0.21**	-0.01	-0.21	-0.53	0.81	0.83	0.76	0.81	0.85	-0.71	0.43
		(45)	(78)	(71)	(89)	(86)	(74)	(45)	(77)	(36)	(40)
Motion to Adjourn	**0.83**	–	–	–	–	–	–	0.85	0.75	0.84	0.88
		(3)	(4)	(4)	(9)	(10)	(5)	(13)	(14)	(17)	(54)
Motion to Recommit	**0.93**	0.92	0.86	0.88	0.89	–	0.92	0.94	0.94	0.95	0.88
		(26)	(17)	(25)	(15)	(10)	(18)	(26)	(35)	(51)	(30)
Instruct Conferees	**0.91**	–	–	–	–	–	0.89	–	0.88	0.96	0.86
		(7)	(3)	(3)	(6)	(11)	(22)	(14)	(27)	(22)	(22)
Motion to Table	**0.91**	0.85	–	–	–	0.83	–	0.91	0.93	0.83	0.94
		(20)	(6)	(7)	(9)	(21)	(4)	(26)	(20)	(37)	(39)
Recede and Concur	**0.86**	–	0.88	0.86	0.71	–	0.91	0.85	0.79	–	–
		(10)	(11)	(17)	(14)	(11)	(27)	(19)	(17)	(7)	(2)
Prev. Ques. on Special Rule	**0.91**	0.85	–	–	0.86	–	0.83	0.91	0.91	0.95	0.91
		(16)	(9)	(8)	(14)	(10)	(18)	(28)	(30)	(54)	(39)

Notes. This table shows the estimated correlation of the full roll call record to procedural vote types in each Congress on the polarity member participated in at least 11 scalable votes. Numbers in parentheses are the total number of votes taken in that category during that Congress. *Continued on next page.*

> ordinary business of the House represented at best a quaint tip of the hat to history or, at worst, a prime opportunity for obstruction and delay by the minority party. Put bluntly, a recorded vote on the *Journal* represents nothing more than a vote on the 'minutes' for the previous day's session. There is no policy impact of the outcome of the vote, *per se*" (Patty 2010, pp121 - 122).

Journal votes have been shown to be used strategically by both the majority and minority party to produce delay and obstruction, without indicating any substantive policy preferences at all (Patty 2010). There is thus a substantive reason for political scientists to exclude them from ideological estimations and analyses. In addition, given that they are the only kind of vote that produce statistically significant differences in correlation to the full record, I have now added a methodological reason not to include them in ideal point estimations. When the roll call record is subsetted, a

Table 4.3 *(Continued)*: Correlation to Full Record by Vote Types – Procedural Votes

Vote Type	**Avg**	106	107	108	109	110	111	112	113	114	115
Miscellaneous	**0.92**	–	–	–	–	–	–	–	–	–	–
		(10)	(5)	(7)	(6)	(5)	(13)	(3)	(1)	(7)	(7)
Rise from COTW	**0.91**	0.92	0.89	–	–	0.93	–	–	–	–	–
		(11)	(18)	(9)	(2)	(18)	(1)	(1)	(0)	(0)	(0)
Passage of Rules	**0.94**	0.92	0.92	0.96	0.96	0.96	0.94	0.96	0.95	0.96	0.96
		(87)	(88)	(73)	(94)	(159)	(132)	(114)	(102)	(125)	(134)
Approve House Journal	**0.21**	-0.43	-0.60	-0.40	–	0.89	0.87	0.37	-0.10	0.12	0.23
		(55)	(66)	(20)	(8)	(31)	(13)	(30)	(34)	(18)	(43)
Motion to Adjourn	**0.83**	0.81	0.91	0.81	–	0.91	0.74	–	–	–	–
		(13)	(24)	(18)	(4)	(62)	(19)	(5)	(10)	(9)	(9)
Motion to Recommit	**0.93**	0.90	0.93	0.95	0.95	0.95	0.92	0.95	0.96	0.98	0.98
		(35)	(44)	(49)	(53)	(118)	(63)	(110)	(103)	(90)	(77)
Instruct Conferees	**0.91**	0.93	0.89	0.94	0.95	0.88	–	0.89	–	–	–
		(37)	(24)	(63)	(26)	(21)	(11)	(18)	(7)	(6)	(6)
Motion to Table	**0.91**	0.86	0.89	0.97	0.95	0.90	0.95	–	0.97	–	0.98
		(14)	(13)	(19)	(17)	(58)	(35)	(7)	(19)	(10)	(22)
Recede and Concur	**0.86**	–	–	–	–	0.88	0.94	–	–	0.92	0.81
		(5)	(5)	(1)	(2)	(25)	(37)	(8)	(10)	(16)	(14)
Prev. Ques. on Special Rule	**0.91**	0.84	0.85	0.86	0.85	0.94	0.96	0.95	0.97	0.97	0.98
		(29)	(34)	(75)	(71)	(153)	(68)	(82)	(85)	(101)	(42)

Notes. This table shows the estimated correlation of the full roll call record to procedural vote types in each Congress on the polarity member participated in at least 11 scalable votes. Numbers in parentheses are the total number of votes taken in that category during that Congress.

journal vote could exert undue influence on the resulting ideal points and significantly alter the conclusions drawn from such an analysis.

Almost all the procedural votes correlated higher with the full record than substantive final passage votes. This shows that procedural votes are exerting more of an influence on the overall w-NOMINATE estimations than the final passage votes when they are estimated in one dimension. This might be due to the fact that w-NOMINATE attempts to maximize explained variance in a small number of dimensions. If the procedural votes all vary together along party lines, then they would dominate the first dimension ideal point coordinates as long as there is a large enough quantity of them to rival the substantive votes in a given Congress.

The relevant distinctions among legislators within a party would not be exposed by these procedural votes on this first dimension because they would be drowned out by the increasing party

unity that drives the legislative process in modern congresses. A second dimension that captures the actual preferences of legislators when it comes to final passage votes, which are less able to be fully dominated by party power, would probably correlate much better to these kinds of votes than the first dimension does.

This kind of analysis is important for anyone who wants to subset the record for particular reasons, because it specifies how specific kinds of votes influence the order of legislators produced by the w-NOMINATE estimation technique. Interest groups, such as the ACU and ADA who say they want to produce a picture of overall ideology from liberal to conservative from a small set of votes, may want to avoid using Journal Approval votes in their ratings, since these do a very poor job of correlating to the full record. Scholars interested in the role of party power in defining what it means to be liberal or conservative might focus specifically on procedural votes, which explain so much of the variance in the first dimension. Others, who may be interested in understanding what other factors besides a legislator's liberal or conservative leanings affect his or her vote, can take a closer look into final passage votes and amendment votes specifically to better understand the meaning of these alternative dimensions.

4.6.3. How Many Legislators are Required for Consistency?

Finally, I investigate the minimum number of legislators needed to produce consistent estimates of ideology. To answer this question, I performed a series of ideal point estimations on the 96th to the 115th Congresses using varying numbers of randomly sampled legislators. First, I identified and removed any legislators who did not have at least 11 votes. I also removed any unanimous or extremely lopsided votes from the roll call record, so that the population of votes and legislators from which I sampled were all able to be included in the estimations.

Next, for each Congress I estimated ideal points for the entire population of votes and legislators. I then sampled 20 votes at random and estimated ideal points for the 20-vote subset. I used these same 20 votes to estimate the ideal points of the varying numbers of legislators. Thus,

as I changed the number of legislators, I was able to compare the resulting ideal points to both a random small sample and to the full record.

The next step in the process was to sample different numbers of legislators randomly and compare the average ideal points to the full record and the 20-vote subset. To illustrate, I set the number of legislators to 49, then I took a random sample of 49 legislators[33] and estimated ideal points using the same 20 randomly sampled votes described above. Next, I correlated the resulting ideal points to the full roll call record and to the 20-vote subset that included all the legislators. I then sampled another 49 legislators and repeated this process 100 times at the 49-legislator level. This produced a series of 100 correlations of 49 legislators to the full roll call record with all the legislators. Additionally, it also produced 100 correlations of 49 legislators to the subsetted roll call record with all the legislators. I recorded the average of both of these series, and then I switched the number of legislators to 48 and repeated the process.

Finally, I repeated this process in each Congress from the 96th to the 115th, creating 100 samples each of 49 to 3 legislators and recording their correlations to the estimations of all legislators. This procedure allowed me to leverage these repeated simulations to produce highly informative averages for each number of sampled legislators. As a result, I found that consistent estimations can be produced using fewer than 50 legislators.

Figure 4.3 displays the correlation of ideal points estimated with varying numbers of legislators to the ideal points estimated from the full record from the 96th to 115th Congresses. While the standard error varies greatly for legislatures with fewer than 20 voters, the average correlation of ideal points to those from the full record is 0.88 across all the simulations and never dips below 80% correlation. A similar result holds for the simulations that examine correlations of small legislatures to smaller subsets of 20 votes, as shown in Figure 4.4. Across all permutations of legislator numbers, the estimated ideal points correlate at 0.90 on average, and again, never correlate at less than 0.80.

[33]I did have to make one non-random choice to include the polarity-member in each sample of legislators. Thus, in this example I would have 48 randomly-selected legislators and the one polarity-member.

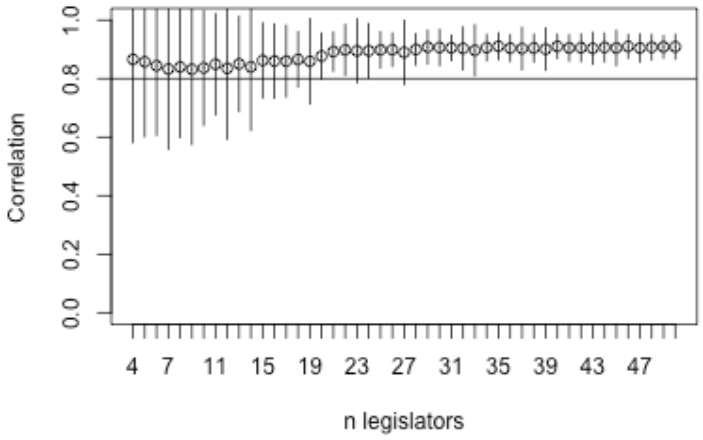

Figure 4.3: Average Correlation of 100 Samples of nLegs to the Full Record

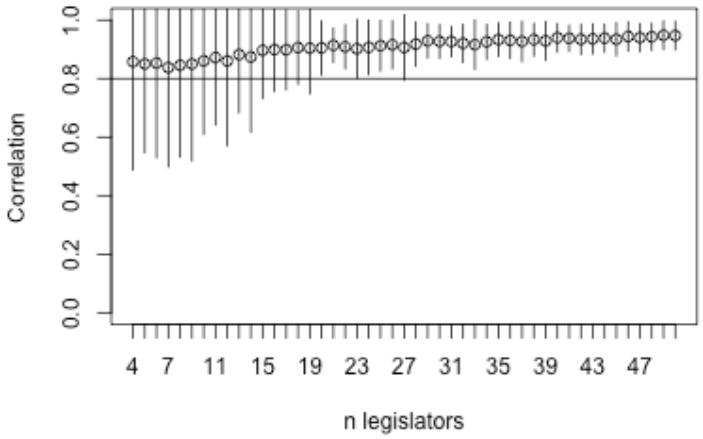

Figure 4.4: Average Correlation of 100 Samples of nLegs to a Random 20 Votes

These findings greatly expand the current possible uses for the NOMINATE method by increasing the number of voting bodies to which the technique can apply. Legislatures with as few as 4 voters can easily be analyzed with this method and produce consistent results. Poole and Rosenthal recommend at least 50 legislators for NOMINATE scores, but this chapter demonstrates that much smaller voting bodies can now be incorporated using this method in order to better understand the legislative process.

4.7. Conclusion

In this paper, I have verified the initial claims of Poole and Rosenthal, who said that 100 votes would serve to provide consistent estimates of legislators' ideological positions. Additionally, I have established that fewer than 100 votes can be used as long as the minimum vote requirement for a legislator to be included in the estimation is set to at least 11 votes. This setting is lower than the default one of 20 that is programmed into the w-NOMINATE package in R, and it performs equally as well as the default value in recovering estimates that correlate with the full record. With at least 11 scalable votes, samples of roll calls with even as few as 25 can produce consistent estimates in one dimension.

This is good news for many political actors and scholars who are interested in using or studying small sets of roll calls. Analyses of interest group ratings, executive or judicial ideology, state legislatures, historical legislatures like the much smaller Constitutional Convention, bureaucracies, committees within Congress, issue-specific ideology questions, and many more examples can all benefit from these findings.

Using hundreds of random samples of both 25 and 100 sample sizes, I have established that a random subset of the modern day roll call record will be sufficient to recover ideal points consistent with those from the full record. This finding can be used to produce a test of the face validity of groups such as interest groups, who create ratings of legislators with the promise to deliver a clear picture of ideology with a small set of important votes. If an analysis of their votes delivers ratings

that differ significantly from the full record, then their ratings can be shown to be misleading and disingenuous misrepresentations of the Representatives' legislative behavior.

After finding the smallest minimum vote parameter that suffices for consistent estimates, I turned to an investigation of the kinds involved in the modern day roll call record to find out how each category influences the overall ordering of legislators. Procedural votes and votes on straight amendments correlate with the full record best, while final passage votes and journal votes correlate worst.

This kind of information is useful for a variety of purposes. For example, since journal votes produce totally inconsistent ideal points relative to the full record, these votes in particular should not be used to calculate ideology in any abbreviated subset like those that interest groups produce. But, journal votes might be the best tool for someone who is interested in making a party seem less unified, such as a member of Congress who does not want to be targeted for being a rubber stamp for his or her party.

Finally, I have demonstrated that the minimum number of legislators needed to produce consistent estimates is also lower than initially claimed by Poole and Rosenthal. Instead of the 50 voters previously presumed necessary, voting bodies with as few as 4 people can now be analyzed and used to uncover more insights into the legislative process.

Overall, this paper has explored the limitations of the w-NOMINATE method for estimating small subsets of the roll call record in one dimension. This method is a highly versatile tool that is absolutely suited to the task of estimating both large and small numbers of votes as long as specific parameters are appropriately set. The sufficient settings for these parameters, including the minimum number of votes a member needs to take to be included and the minimum number of votes that need to be included in the set overall, are both shown to be lower than previously claimed.

Future research should address some of the lingering questions about the limitations and implications of the w-NOMINATE model. Important questions remain, such as how well certain kinds of votes correlate with the second dimension in w-NOMINATE models estimated in 2 dimensions,

and what is the substantive meaning of such higher dimensions? How often do interest groups use inappropriate votes, like journal votes, in their ratings? Have procedural votes always been so influenced by parties, or has this changed over time? Do these same tendencies in final passage versus procedural votes also hold in the Senate? This paper is just the first step in answering these and many other questions related to the use and effectiveness of the w-NOMINATE model to analyze the roll call record of the US Congress.

CHAPTER 5: CONCLUSION

The three essays in this dissertation have each approached ideal point estimation using small subsets of congressional votes from different angles. Ideal points estimated from the full roll call record are by definition the most informative about the distinctions between legislators, and the fact that they are kept updated and made freely available in the form of DW-NOMINATE scores completely explains their extensive and continued use in political science. But subsetting the voting record allows for a variety of new questions to be asked and answered, and this technique is sure to be used more and more often as technology improves and as the field of political science matures in its statistical training. It is thus important to examine the consequences of subsetting the record and provide a foundation for the best practices to use when subsetting the record and analyzing these samples. My dissertation takes some of these first steps towards this end and identifies further avenues to explore along this path.

The first essay examines the ADA's and ACU's selection of the roll call subsets they include in their legislator ratings in order to gain insight into the groups' own political motivations. I find that the interest group ratings are crafted strategically, with a focus on distinguishing among members of the groups' preferred party and making individual House members look more powerful than they really are in the face of party politics. Instead of displaying the basic liberal to conservative ideological spectrum using a few key votes, as they claim to do, the ADA and ACU choose votes that produce artificial divisions among their preferred party. By publicizing these artificial distinctions to their members and the public, the groups can put pressure on members to vote more or less conservatively or face potential electoral consequences.

In addition, this chapter shows how the behavior of these interest groups has changed over time in response to increasing party power in Congress. In the past, when party power in Congress was much less powerful, interest groups tended to choose votes that created artificial polarization and lumping in the resulting legislator ideal points. In response to the increase in party power that took place over the last few decades, the ADA and ACU have strategically changed their behavior to now select votes that artificially moderate the members of the House of Representatives and thereby put pressure on these artificial moderates to vote with the party more often.

This chapter makes several important contributions to political science, and offers many avenues for future research. I use Item Response Theory to construct a mathematical standard by which to evaluate the quality of interest group ratings. This technique can be applied to more groups than just the two in my initial study to gain insights into the motivations and behaviors of interest groups as a broad category of influential political actors. This method can also expose changes in these groups' behaviors over time, like the ADA and ACU's responses to shifts in party power in Congress. Finally, I demonstrate why and how interest group ratings themselves can be usefully employed to provide important scientific insights into the political world. These ratings comprise interesting subsets of votes that provide information about the groups, the legislators, and the votes themselves, while at the same time presenting a clear example of the need to carefully examine the consequences of subsetting the roll call record in the first place.

The second essay translates the Item Response Theory model from the analysis of the basic liberal to conservative dimension that underlies most ideological differences in Congress to an analysis of a special interest group whose subsets of roll calls are chosen based on a single issue. This chapter provides a template for using subsets of votes to establish whether or not an issue reveals distinct preferences that cannot be explained with the liberal to conservative dimension that political scientist often rely on by default. Neither the subset of education-related votes chosen by the NEA nor the one identified by political experts exposes a distinct education-related dimension relative to the liberal-conservative one, but this technique can be fruitfully used to test other issue-areas and other subsets of issue-specific votes.

The ideal point estimation technique I employ in these first two essays is borrowed from the educational testing literature, and it offers several unique features for analyzing congressional votes and legislators and evaluating the quality of others' analyses. The Item Response Theory model allows not only the legislators' ideal points and distances to be quantified using their yes and no votes, but also quantifies the quality of the votes themselves. This technique reveals how much information a vote provides, how well it helps distinguish among legislators, and at which place along the ideological scale that the vote distinguishes best. This information is not available when other ideal point estimation techniques are used, and this dissertation only provides a beginning demonstration for how the IRT model can be fruitfully applied to the analysis of the US Congress and other legislatures.

The second essay offers many noteworthy contributions. First, it provides further evidence for the importance of incorporating new techniques like the IRT model into political science to ask and answer interesting questions. It takes subsets of roll calls created by interest groups and demonstrates their relevance to modern day political science. This is an important reminder that even though these data have long been considered uninteresting ever since the development of better techniques such as NOMINATE scores for understanding and accessing ideological differences among US legislators, interest group ratings can still offer unique insights and directions for political research. Additionally, this chapter exposes a need in political science for repetition and continued investigation into questions that may be considered settled. When new evidence or techniques present themselves, they can shed new light on these old questions and provide a deeper understanding of the complexities of the political world.

My third essay makes strides towards these goals by providing a better use and understanding of currently available estimation methods, clarifying the limits and potential uses of the w-NOMINATE technique and opening it up for use in a much wider array of contexts. Although other techniques for analyzing small subsets of votes are already in use (namely, the IRT method used in the first two essays), the NOMINATE method offers several unique and important features that should not be forgotten or used only when a sufficiently large voting record is available.

By expanding the potential uses of the NOMINATE method to voting bodies comprising smaller numbers of votes and voters, researchers now have access to the unique aspects of the NOMINATE method, such as cutpoints/cutlines and built-in bootstrapped standard errors, when analyzing small voting bodies. This chapter adds breadth to an already extremely versatile tool and gives researchers access a technique that delivers value and robustness to studies of smaller legislatures and voting bodies.

This chapter also examines the consequences of subsetting the roll call record in the first place, showing that small subsets of votes can be used instructively to gain a better understanding of politics. Not only can samples of as few as 11 votes produce reliable and consistent estimations of ideology, they can do so for legislatures and bureaucracies with as few as 4 voters. This provides yet another metric by which we can judge others, such as interest groups, who create conclusions based on subsets of these sizes. The quality of their ratings can be easily evaluated relative to both random samples of the population from which they drew and the analysis of the population itself. This kind of analysis can help to understand the motivations of these kinds of groups and potentially to expose any egregious rating behavior, such as the inclusion of journal votes, for example, which are known to produce misleading ideological distinctions.

One important consequence of subsetting the voting record of any legislature is the ability to isolate and analyze specific kinds of votes or votes that deal directly with specific issues. The third essay takes up this question in regards to the specific kinds of votes that are found in the US Congress and found that most of the types of votes can be included in subsets of votes without substantially altering the resulting ideological positions. However, one kind of vote in particular, journal votes, produce wildly inconsistent estimations of legislator ideology. These votes should not be used in any subset of votes that is used to measure ideology, such as the ADA's and ACU's legislator ratings, for instance. In fact, it is questionable whether they should be included in the ideological estimations that use the full roll call record. If they are excluded from DW-NOMINATE scores, for example, would the ideal point estimations change? This could potentially affect the

measures of polarization and party power that political scientists have relied upon for years to make conclusions about the US Congress.

While these are questions that must be left to future research, they point to the importance of understanding the consequences of subsetting the voting record of the US Congress. This is the motivating idea which I have tried to build upon and begin to answer in this dissertation. I found that there are times when it is necessary to subset the record, as in the attempts to access an issue-specific voting dimension. There are times when subsetting the record is done subjectively, as by certain interest group ratings, which leads to strategic and manufactured distinctions among legislators. The resulting ideal points can also be evaluated by subsetting the record using a less subjective, mathematical standard, such as that provided by the IRT method. Finally, subsetting the record can provide insight into the effect that certain kinds of votes have on the ideal point estimations themselves.

The point is that researchers should better understand how ideal points are created and affected by the assumptions and choices embedded in the default parameter settings of the programs that perform the estimations. Sometimes it is better to use the full record to estimate ideal points, as in DW-NOMINATE scores, whereas at other times it might be better to subset the record in order to answer a specific question. Understanding the difference between the resulting ideal points and their underlying meaning is of utmost importance given the prevalence of "ideology" as a variable within political science studies. The more that researchers know about subsetting the record, especially about how it is done, best practices, and what to avoid when choosing which votes to include, the better their subsets will reflect and implement their goals. This dissertation provides an initial starting point for better understanding and evaluating these consequences and how these subsetting techniques can be fruitfully applied in the future.

REFERENCES

American Conservative Union. "ACU Ratings of Congress." http://acuratings.conservative.org/acu-federal-legislative-ratings/ (Retrieved April 14, 2018).

American Conservative Union Twitter Account. Tweet about Neil Gorsuch. https://twitter.com/ACUConservative/status/849985975044460544.

Americans for Democratic Action. "ADA Voting Records." https://adaction.org/ada-voting-records/ (Retrieved April 14, 2018).

Anderson, Sarah E. 2012. "Policy Domain-Specific Ideology: When Interest Group Scores Offer More Insight." *Politics & Policy* 40(6): 1186-1202.

Anderson, Sarah E., and Philip Habel. 2009. "Revisiting Adjusted ADA Scores for the US Congress, 1947-2007." *Political Analysis* 17: 83-88.

Ansolabehere, Stephen and Philip E. Jones. 2010. "Constituents' Responses to Congressional Roll-Call Voting." *American Journal of Political Science* 54(3): 583-597.

Bateman, David A., Joshua D. Clinton, and John S. Lapinski. 2017. "A House Divided? Roll Calls, Polarization, and Politcy Differences in the US House, 1877-2011." *American Journal of Political Science* 61(3): 698-714.

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960-93." *American Journal of Political Science* 42(1): 327-348.

Berry, William D., Richard C. Fording, Evan J. Ringquist, and Carl E. Klarner. 2010. "Measuring Citizen and Government Ideology in the US States: A Re-appraisal." *State Politics & Policy Quarterly* 10, no. 2: 119-137.

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2): 294-311.

Brunell, Thomas L., William Koetzle, John Dinardo, Bernard Grofman, and Scott L. Feld. 1999. "The R2 = .93: Where Then Do They Differ? Comparing Liberal and Conservative Interest Group Ratings." *Legislative Studies Quarterly* 24(1): 87-102.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. The American Voter. University of Chicago Press.

Candeub, Adam. 2013. "Transparency in the Administrative State." In: *Houston Law Review* 51(2): 385-416.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(March): 127-140.

Carney, Eliza Newlin. 2013. "Keeping Score: The Power of Lawmaker 'Report Cards'." *CQ Weekly*. http://public.cq.com/docs/weeklyreport/weeklyreport-000004207216.html.

Carpini, Michael X. Delli, and Scott Keeter. 1996. What Americans Know About Politics and Why It Matters. Yale University Press.

Carson, Jamie L., Gregory Koger, Matthew J. Lebo, and Everett Young. 2010. "The Electoral Costs of Roll-Call Voting." *American Journal of Political Science* 54(July): 598-616.

Chand, Daniel E., and William D. Schreckhise. 2013. "Keeping Score on Congress: Explaining Variations in Interest Group Ratings of US Senators." *Business and Politics* 15(3): 401-433.

Chand, Daniel E., William D. Schreckhise, and Janine A. Parry. 2014. "In Defense of the Lowly Scorecard: What Legislator Ratings Can Tell Use about Interest Groups." *Interest Groups & Advocacy* 3(3): 268-292.

Chappell, Henry W. 1981. "Campaign Contributions and Voting on the Cargo Preference Bill: A Comparison of Simultaneous Models." *Public Choice* 36: 301-312.

Chappell, Henry W. 1982. "Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model." *Review of Economics and Statistics* 64: 77-83.

Clausen, Aage. 1973. How Congressmen Decide: A Policy Focus. New York, NY: St. Martin's Press.

Crespin, Michael H. and David Rohde. 2019. Political Institutions and Public Choice Roll-Call Database. Retrieved from https://ou.edu/carlalbertcenter/research/pipc-votes/

Crespin, Michael H., David W. Rohde, and Ryan J. Vander Wielen. 2011. "Measuring Variations in Party Unity Voting." *Party Politics* 19(3): 432-457.

Comparative Agendas Project. "About." https://www.comparativeagendas.net/pages/About. Retrieved January 2, 2020.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." *Critical Review* 18, no. 1-3: 1-74.

Coughlin, Cletus C. 1985. "Domestic Content Legislation: House Voting and the Economic Theory of Regulation." *Economic Inquiry* 23: 437-448.

Crespin, Michael H. and David Rohde. Retrieved Dec. 9, 2019. Political Institutions and Public Choice Roll-Call Database. Retrieved from https://ou.edu/carlalbertcenter/research/pipc-votes/

Davis, Michael L., and Philip K. Porter. 1989. "A Test of Pure and Apparent Ideology in Congressional Voting." *Public Choice* 60: 101-111.

Deming, David J. and David Figlio. 2016. "Accountability in US Education: Applying Lessons from K-12 Experience to Higher Education." *Journal of Economic Perspectives* 30(3): 33-56.

Dougan, William R., and Michael C. Munger. 1989. "The Rationality of Ideology." *Journal of Law and Economics* 32: 119-142.

E. Scott Adler and John Wilkerson, Congressional Bills Project: (2007-2014), NSF 00880066 and 00880061. The views expressed are those of the authors and not the National Science Foundation.

Erikson, Robert S., and Gerald Wright. 1980. "Policy Representation of Constituency Interests." *Political Behavior* 2: 91-106.

Fenster, Mark. 2006. "The Opacity of Transparency." In: *Iowa Law Review* 91(3): 885-950.

Fortunato, David and Randolph T. Stevenson. 2016. "Heuristics in Context." *Political Science Research and Methods* 7(2):311-330.

Fowler, Linda L. 1982. "How Interest Groups Select Issues for Rating Voting Records of Members of the U.S. Congress." *Legislative Studies Quarterly* 7(3): 401-413.

GovTrack Voting Records. https://www.govtrack.us/congress/votes.

Groseclose, Tim, Steven D. Levitt, James M. Snyder, Jr. 1999. "Comparing Interest Group Scores Across Time and Chambers: Adjusted ADA Scores for the U.S. Congress." *American Political Science Review* 93(1): 33-50.

Gutmann, Amy. 1999. Democratic Education. Princeton, NJ: Princeton University Press.

Hambleton, Ronald K., H. Swaminathan, and H. Jane Rogers. 1991. Fundamentals of Item Response Theory. Newbury Park: Sage Publications.

Herron, Michael C. 1999. "Artificial Extremism in Interest Group Ratings and the Preferences versus Party Debate." *Legislative Studies Quarterly* 24(4): 525-542.

Jackson, John E., and John W. Kingdon. 1992. "Ideology, Interest Group Scores, and Legislative Votes." *American Journal of Political Science* 36(3): 805-823.

Jacobson, Gary C., and Jamie L. Carson. 2015. The politics of congressional elections. Rowman & Littlefield.

Jacobson, Louis. 2011. "Major Overhauls of 'No Child' Under Way." https://www.politifact.com/truth-o-meter/promises/obameter/promise/245/reform-no-child-left-behind/. Retrieved December 8, 2019.

Jannetta, Ian. 2009. "Planning Underway for Reforms." https://www.politifact.com/truth-o-meter/promises/obameter/promise/245/reform-no-child-left-behind/. Retrieved December 8, 2019.

Johannes, John, and John McAdams. 1981. "The Congressional Incumbency Effect: Is It Casework, Policy Compatibility, or Something Else? An Examination of the 1978 Election." *American Journal of Political Science* 25: 512-542.

Kalt, Joseph P., and Mark A. Zupan. 1984. "Capture and Ideology in the Economic Theory of Politics." *American Economic Review* 74: 279-300.

Kant, Immanuel. 1900. Kant on Education (Ueber Padagogik), trans. Annette Churton. Boston: D.C. Heath and Co., p12.

Kau, James B., and Paul H. Rubin. 1979. "Self-Interest, Ideology, and Logrolling in Congressional Voting." *Journal of Law and Economics* 22 365-384.

Kau, James B., and Paul H. Rubin. 1982. Congressmen, Constituents and Contributors. Boston: Martinus Nijhoff.

Kisaalita, Alice E., and Michael S. Lynch. "Are they Asking the Right Questions? Assessing Interest Group Scores Using Item Response Theory." manuscript, 2020.

Krehbiel, Keith. 1990. "Are Congressional Committees Composed of Preference Outliers?" *American Political Science Review* 84: 149-163.

Kritzer, Herbert M. 1978. "Ideology and American Political Elites." *Public Opinion Quarterly* 42: 484-502.

Layton, Lyndsey. 2015. "NEA: No Child Left Behind Rewrite Doesn't Level the Playing Field." *Washington Post.* https://www.washingtonpost.com/local/education/nea-no-child-left-behind-rewrite-doesnt-level-the-playing- field/2015/04/10/.html. Retrieved January 2, 2020.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2017). Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

Libertarian 2018 Platform. https://www.lp.org/platform/. Retrieved Dec. 9, 2019.

Lynch, Michael and Anthony J. Madonna. 2013. "Viva Voce: Implications from the Disappearing Voice Vote, 1865-1996." *Social Science Quarterly* 94(2): 530-550.

MacArthur, John, and Steven V. Marks. 1988. "Constituent Interest vs. Legislator Ideology: The Role of Political Opportunity Cost." *Economic Inquiry* 26: 461-470.

Markus, Gregory B. 1974. "Electoral Coalitions and Senate Roll Call Behavior: An Ecological Analysis." *American Journal of Political Science* 18: 595-607.

Mayhew, David. 2004. Congress: The Electoral Connection. 2nd Ed. New Haven: Yale University Press.

Meltzer, Erica. 2017. "Here's why Sen. Bennet says he'll vote against Neil Gorsuch." Denverite. https://denverite.com/2017/04/06/sen-michael-bennet-votes-cloture-fellow-democrats-pass-40-vote-threshold-filibuster-gorsuch/.

National Commission on Excellence in Education. 1983. "A Nation at Risk: The Imperative for Educational Reform, 1983." Department of Education.

National Education Association. "Issues and Actions." https://educationvotes.nea.org/issues-and-actions/. Retrieved January 2, 2020.

National Education Association. "Legislative Report Card." http://www.nea.org/home/19413.htm. Retrieved December 10, 2019.

National Education Association. "NEA's Response to Race to the Top." http://www.nea.org/home/35447.htm. Retrieved January 2, 2020.

National Education Association. "NEA's Vision, Mission, and Values." http://www.nea.org/home/19583.htm. Retrieved December 6, 2019.

Oleszek, Walter J. 2014. Congressional Procedures and the Policy Process. 9th Ed. Los Angeles: Sage.

Patty, John W. 2010. "Dilatory or anticipatory? Voting on the Journal in the House of Representatives." *Public Choice* 143: 121-133.

Peltzman, Sam. 1984. "Constituent Interest and Congressional Voting." *Journal of Law and Economics* 27: 211-240.

Poole, Keith T. 1981. "Dimensions of Interest Group Evaluation of the U.S. Senate, 1969-1978." *American Journal of Political Science* 25(1): 49-67.

Poole, Keith T. 2005. Spatial Models of Parliamentary Voting. New York: Cambridge University Press.

Poole, Keith T., and R. Steven Daniels. 1985. "Ideology, Party, and Voting in the U.S. Congress, 1959-1980." *American Political Science Review* 79: 323-399.

Poole, Keith T., and Howard Rosenthal. 1984. "The Polarization of American Politics." *Journal of Politics* 46: 1061-1079.

Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29: 357-384.

Poole, Keith T., and Howard Rosenthal. 1991a. "Patterns of Congressional Voting." *American Journal of Political Science* 35: 228-278.

Poole, Keith T., and Howard Rosenthal. 1991b. "On Dimensionalizing Roll Call Votes in the U.S. Congress (Controversy with Kenneth Koford)." *American Political Science Review* 85: 955-975.

Poole, Keith T., and Howard Rosenthal. 1993. "Spatial Realignment and the Mapping of Issues in American History: The Evidence from Roll Call Voting." In Agenda Formation, edited by William H. Riker. Ann Arbor: University of Michigan Press.

Poole, Keith T., and Howard Rosenthal. 1998. "The dynamics of interest group evaluations of Congress." *Public Choice* 97: 323-361.

Poole, Keith T., and Howard Rosenthal. 2007. Ideology and Congress. New Brunswick, New Jersey: Transaction Publishers.

Project Vote Smart. "About Vote Smart." https://votesmart.org/about#vsds, (Retrieved June 2, 2019).

Ray, Bruce. 1980. "The Responsiveness of the U.S. Congressional Armed Services Committees to Their Parent Bodies." *Legislative Studies Quarterly* 5: 501-515.

Reeher, Grant. 2001. "Self-Reported Legislator Ideology versus Interest Group Ratings." *Polity* 34(2): 231-240.

Richardson, Liiliard E., and Michael C. Munger. 1990. "Shirking, Representation, and Congressional Behavior: Voting on the 1983 Amendments to the Social Security Act." *Public Choice* 67: 11-33.

Roberts, Jason, David Rohde, and Michael H. Crespin. Retrieved Dec. 9, 2019. Political Institutions and Public Choice Senate Roll-Call Database. Retrieved from https://ou.edu/carlalbert center/research/pipc-votes/.

Rohde, David. Political Institutions and Public Choice House Roll-Call Database. Duke University, Durham, NC. 2010.

Rules of the Senate. https://www.rules.senate.gov/rules-of-the-senate.

Schwartz, John E., and Barton Fenmore. 1977. "Presidential Election Results and Congressional Roll Call Behavior: The Cases of 1964, 1968, and 1972." *Legislative Studies Quarterly* 2: 409-423.

SEMrush.com. "Traffic Analytics: votesmart.org". https://www.semrush.com/analytics/traffic/overview/votesmart.org, (Retrieved June 2, 2019).

Smith, Steven S. 2007. Party Influence in Congress. Cambridge, New York: Cambridge University Press.

Snyder, James M. 1992. "Artificial Extremism in Interest Group Ratings." *Legislative Studies Quarterly* 17(3): 319-345.

"The Constitution of the United States." Article 1, Section 5.

Tien, Charles. 2001. "Representation, Voluntary Retirement, and Shirking in the Last Term." Public Choice 106: 117-130.

Weingast, Barry R., and William Marshall. 1988. "The Industrial Organization of Congress." *Journal of Political Economy* 96: 132-163.

Weiss, Elaine. 2014. "Mismatches in Race to the Top Limit Education Improvement." *Education Digest* 79(5): 60-65.

Wilcox, Clyde and Aage Clausen. 1991. "The Dimensionality of Roll-Call Voting Reconsidered." *Legislative Studies Quarterly* 16(3): 393-406.

Williams, Joshua. 2008. "The War on Education: The Negative Impact of the No Child Left Behind Act on Inner-City Public Schools, Students, and Teachers." *Journal of Gender, Race & Justice* 11(3): 573-598.

Yanushevsky, Rafael. 2011. Improving Education in the US: A Political Paradox. New York, NY: Algora Publishing.

Zaller, John R. 1992. The Nature and Origins of Mass Opinion. Cambridge University Press.

APPENDICES

Appendix A

This section is devoted to specifying the IRT model explicitly (see also Hambleton et al. 1991; and Kisaalita & Lynch, manuscript). An IRT model is estimated for all votes in a certain year of interest. The IRT model is estimated using a one-dimensional scale of ideology. I use a two-parameter item response model to estimate the latent variable of ideology (θ). In addition to estimating ideology, a two-parameter item response model estimates α and β coefficients, where α is the item difficulty parameter and β is the item discrimination parameter.

The data in the estimation model consists of n legislators voting on m different roll call votes. For each roll call $j = 1, \ldots, m$, legislator i (where $i = 1, \ldots, n$ and n is the total number of legislators) chooses between voting in a manner consistent with the stated choice of the interest group, (called "liberal" for the ADA and "conservative" for the ACU) or voting against the stated choice of the group (called "conservative" for the ADA and "liberal" for the ACU). The choice is then between a group preferred location ψ_j, and a group nonpreferred location ζ_j. Let $y_{i,j} = 1$ if the i^{th} legislator votes in a manner consistent with the group on the j^{th} roll call, and let $y_{i,j} = 0$ if the i^{th} legislator votes in a manner not consistent with the group on the j^{th} roll call. For ideal scores, $y_{i,j} = 1$ if the i^{th} legislator votes "yea", $y_{i,j} = 0$ if the i_{th} legislator votes "nay", and $y_{i,j}$ is missing if the legislator does not vote on the item. Because votes are not identified as being "liberal" or "conservative", the IRT model estimates whether the items are "liberal" or "conservative". If the β value for an item is negative, then a liberal legislator would be expected to vote "yea" on the item. If the β value for an item is positive, then a conservative legislator would be expected to vote "yea" on the item. $\mathbf{Y}$ is the $n \times m$ matrix of observed roll call votes, containing $y_{i,j}$ for all i and j. Legislators are assumed

to have quadratic utility functions over the $\Re^1$ policy space, such that $U_i(\psi_j) = -\|\theta_i - \psi_j\|^2 + \eta_{i,j}$ and $U_i(\zeta_j) = -\|\theta_i - \zeta_j\|^2 + \nu_{i,j}$, where $\theta_i \varepsilon \Re^1$ is the ideal point of the i^{th} legislator, $\eta_{i,j}$ and $\nu_{i,j}$ are the errors of utility, and $\|\|$ is the Euclidean norm. Let $y^*_{i,j} = U_i(\zeta_j) - U_i(\psi_j)$. Through utility maximization:

$$y_{i,j} = \begin{cases} 1 \Longleftrightarrow y^*_{i,j} > 0 \\ 0 \Longleftrightarrow y^*_{i,j} \leq 0. \end{cases}$$

This gives the following equation for $y^*_{i,j}$, which can be rearranged and substituted to form the second equation (see Clinton et al. 2004),

$$y^*_{i,j} = -\|\theta_i - \zeta_j\|^2 + \|\theta_i - \psi_j\|^2 + \nu_{i,j} - \eta_{i,j}$$

$$y^*_{i,j} = \beta'_j x_i - \alpha_j + \varepsilon_{i,j}$$

where $\beta_j = 2(\zeta_j - \psi_j), \alpha_j = \psi'_j\psi_j - \zeta'_j\zeta_j$, and $\varepsilon_{i,j} = \eta_{i,j} - \nu_{i,j}$. Given binary roll call data, these equations produce a probit model.

The likelihood is,

$$L(\alpha, \beta, \theta|\mathbf{Y}) = \prod_{i=1}^{n}\prod_{j=1}^{n} \phi(\theta'_j\beta_j - \alpha_j)^{y_{i,j}}(1 - \phi(\theta'_j\beta_j - \alpha_j))^{1-y_{i,j}}$$

where α is an $m \times 1$ vector of difficulty parameters (α_j is the difficulty parameter for the j^{th} roll call), $\mathbf{B}$ is an $m \times 1$ vector of discrimination parameters (β_j is the discrimination parameter for the j^{th} roll call), and θ is an $n \times 1$ vector of ideal points (θ_i is the ideal point parameter for the i^{th} legislator).

Appendix B

Table B.1: Correlations for All, NEA-Related, and NEA-Chosen Votes to Random 20

Congress	Related to Random 20	Chosen to Random 20	All to Random 20
House			
110	0.99	0.97	0.99
111	0.99	0.96	0.99
112	0.99	0.95	0.99
113	0.99	0.95	0.98
Senate			
110	0.99	0.97	0.99
111	0.99	0.96	0.98
112	0.99	0.97	0.99
113	0.99	0.97	0.99

As a robustness check, I test whether the NEA is doing a better job selecting votes to include in their ratings than an average draw of NEA-related and education-only votes. Tables B.1 displays correlations for the ideal points for the three subsets of roll calls that I have examined in this chapter in relation to ideal points estimated from 100 random samples of 20 NEA-related votes.

The first column displays the average correlations of the ideal points estimated from all NEA-related votes with the ideal points estimated from 100 random samples of 20 NEA-related votes. This set of correlations tests whether a small subset of NEA-related votes can ever provide consistent, representative ideal points relative to the larger, more informative population of votes from which the subset is drawn. It helps to understand whether the NEA's project of picking between 15 to 25 votes per Congress is ever a useful endeavor for understanding the key distinctions among legislators on this dimension. From the results, we see that the random subsets of twenty votes correlate extremely well with the full NEA-related set of votes at an average of 99% across the board.

The second column in Table B.1 shows how well the NEA's chosen votes correlate with the average random twenty NEA-related votes. While these correlations are again extremely high,

Table B.2: Correlations for All, Education-Related, and NEA-Chosen Votes to Random 20

Congress	Related to Random 20	Chosen to Random 20	All to Random 20
House			
110	0.99	0.95	0.96
111	0.99	0.94	0.95
112	0.99	0.93	0.97
113	0.99	0.90	0.95
Senate			
110	0.99	0.93	0.95
111	0.99	0.96	0.97
112	NA	NA	NA
113	NA	NA	NA

Notes. The 112th and 113th Senates do not have enough non-unanimous education-only votes to estimate ideal points, so their correlations cannot be reported here.

ranging from 95% - 97%, it is interesting to note that they correlate slightly less well on average than those from column 2. On average, twenty random NEA-related votes correlate with the full set of NEA-related votes at almost 1:1, whereas the NEA's chosen votes do slightly worse in representing this set of votes. Still, the NEA seems to be choosing votes that correlate at very high levels with the set from which they are drawn, and the difference between their ratings and the average random draw of the same size is not significant.

The last column relates each set of ideal points estimated from the NEA-related votes and their subsets to the ideal points resulting from all non-unanimous votes taken in the Congress, whether or not they are NEA-related. Again, these correlations are extremely high for this subset in both the House and the Senate, averaging at almost 1:1 for all votes to the average draw of twenty NEA-related votes.

Table B.2 shows the correlations of the education-only related votes to the average ideal points estimated from 100 random draws of 20 education-only votes, as identified by the Policy Agenda Project experts. All the subsets in the table correlate at 90% or above. These results comport

with those from every other analysis in this chapter, showing that the "education dimension" is not separate from the liberal to conservative one that explains most voting behavior in Congress.