

Problems with Open-Ended ANES Questions Measuring Factual Knowledge about Politics

James Gibson

First I would like to congratulate Jon and Skip for just the incredible effort they are about to embark on. But also for the openness and the way they have reacted to the criticisms that have been leveled at some of the ANES coding procedures. It would have been quite easy for them to hunker down but from the very moment that I and others have raised these kinds of issues, they have been encouraging, responsive and have really sought to shine all possible light on the data. I appreciate that. It is a very welcome development.

We are not really here to cast blame it seems to me, we are here to fix the problem. But to the extent that there is blame, it is not these two who are to blame by any stretch of the imagination. But I really think we ought to emphasize and I hope my remarks aren't taken in the wrong way because my objective is not to blame anyone but rather to identify the problem. I should also say that I am probably not very good at fixing the problem. My specialty is definitely not the coding of open-ended information so I'm hoping that some of what I say up here today, you'll fix and come up with solutions to. It is true, however, that my area of interest is in judicial politics. 15 years ago I was the director of a very large project that coded the decisions of the Supreme Court over a 20 year period so I did have some experience with this sort of content analysis.

The material I am going to talk about today is from a paper that is forthcoming in the *Journal of Politics* but also the big version of that paper, in fact a substantially larger version of that paper is in a book that Greg Caldera and I are publishing next year on the nomination of Sam Alito to the US Supreme Court. We are interested in all of our work in the degree to which ordinary people have knowledge of it, and understand the activities and functions of the US Supreme Court. It is a substantive issue for me. I am not in this for methodical lessons; I am in it for the substance. As you might know, this is a topic that receives a lot of attention in the legal community. For instance, people often trot out survey results like this – 77% of the American people can name 2 of Snow White's dwarfs but only 24% of the American people can name 2 Justices on the Supreme Court. Now this, I think, is Zogby poll; it is not the kind of data that we would take very seriously but you know, I was at a conference at the Annenberg Policy Center at Penn last week with Sandra Day O'Connor, former Justice of the US Supreme Court. Justice O'Connor trotted out this kind of finding to justify minimizing the impact of the mass public over the selection of judges in the US. And the ANES findings on public knowledge toward the courts are routinely cited by legal scholars and policy makers as justification for minimizing the role of the public in selecting judges. As most of you probably don't know, there is a big movement among the United States elites to try to do away with the election of judges in the American states. My point here is that this is an incredibly important substantive issue and ANES have been front and center in the debate for literally 2 decades now.

Let me stop for a minute and review the ANES approach to the measurement of knowledge. I forget if either Jon or Skip said it, but one of the points that one of them made is that the conceptualizations that led to the formation of the questions in the first place may in fact be out dated and, at a minimum, need to be made explicit and perhaps revisited and reconstructed. But knowledge for at least a portion of ANES means holding political information. Indeed, it means knowing highly technical definitions of political personalities and institutions. There is a lot of focus on personalities, on the occupants of positions, a lot of focus on national and international figures, and very little on state and local. There is an assumption that it is more important to know who Gordon Brown is than Kwame Kilpatrick, who I believe is the mayor of Detroit who is going to jail. He is in jail. He is the only mayor who is in jail. I am not necessarily quarreling with this conceptualization, it may be a good one, but there are a lot of assumptions built into the measures. I think some of them need reconsideration.

My interest is in the Supreme Court and the judiciary. I am going to focus on the operationalization of this conceptualization in the ANES. We have a lot of foreign guests today. You all might not know who William Rehnquist is. You are not ill informed because you don't. William Rehnquist is dead. He is the former chief justice of the highest court in this land, the US Supreme Court. Rehnquist is useful because he was

on the bench for almost 30 years and therefore questions were asked over a very long period of time about Rehnquist. John Roberts is the current head of the Supreme Court. Here is a question – someone put it on the screen earlier – it is the open-ended question that asked about a variety of political leaders. Rehnquist; Dennis Hastor who is the leader of the American Congress; Dick Chaney is of course the soon-to-be ex vice president of the United States; and Tony Blair. The first point that I would make is that ANES practices seem to have changed over time in how the open-ended information is recorded.

Sometimes, and I think generally, the interviewers recorded the verbatim responses of the respondents in open-ended questions. But there have been instances of on-the-spot coding in which the interviewers make a judgment as to whether the answers are correct or not correct. I'm not exactly certain about all of this. This isn't a deep dark secret but it is a little bit difficult to assemble all of the material in an accessible way to know exactly what is going on with any given year and the coding of it. There is not much documentation. I think all of this is soon to be relatively mechanical and not discretionary coding. There is no discussion of inter-coder reliability or any of the things we might see in coding open-ended text. As a function of this, I'm afraid that temporal variability in the responses, and analysis of change over time, is made considerably more difficult because the coding procedures are not necessarily consistent.

In 2004, these were the instructions for coding the question about Bill Rehnquist, a judge. What is important here, if you read this, you can see that only Chief Justice is correct. Justice alone is definitely not acceptable. Court alone is not acceptable. Those are fairly detailed instructions for how to go about coding the responses. And the same is true for the other people mentioned in the survey, like Tony Blair.

I have 2 observations. First, this conceptualization of knowledge is a very, very technical one. Let me give you an example – which is correct – I'll put some money on the table that you all will get this wrong. William Rehnquist, John Roberts is the Chief Justice of the United States Supreme Court or the Chief Justice of the United States? Virtually everyone would select # 1 but of course we all know that the official title is Chief Justice of the United States, not of the Supreme Court. There is a similar sort of deal with Tony Blair. I think that if you say Blair or Gordon Brown is the Prime Minister of the British Isles, yes, that is probably incorrect but it is definitely scored as incorrect. You have to say these sorts of magic words in order to get to be a correct answer.

Now I understand why that is the case. I think it is obviously an effort to limit the discretion of the coders, to increase reliability. Of course I appreciate that. But I'm afraid what has been done is that validity has been compromised as a result of the effort to get greater reliability.

So, as has been told, I was able to get access to the verbatim information, verbatim codes for a variety of years. I signed a pledge. What I am about to show you, I can't put quotation marks around, because I am strictly and legally forbidden from revealing any of the responses.. You can see there is a gigantic difference across the 4 different offices. Virtually everyone can identify Chaney. But the important figure in here is 28, which does bounce around a fair amount from year to year. The important figure is 28 and 62 for Tony Blair. So there is a lot of variability across these and presumably that is their purpose. At some level you can't care about the individuals. You want an easy test, you want a more difficult test, and you want a very difficult test, and an extremely difficult test. From the point of view of frequency distribution for a knowledge scale that is actually not necessarily a bad set of items.

Here are incorrect identifications of Rehnquist. For those of you not from the United States, Rehnquist is indeed the head honcho, the main guy, he did indeed – I would have never gotten this right – he did indeed run the Clinton impeachment hearings in his capacity of Chief Justice. I put up there – no one of course put governor of West Virginia because we all know he is not – but I put up there that answer because it emphasizes the point that all these answers are scored the same way. It is the classic definition of measurement error. The observed score is the same. The true score varies enormously across the respondents. Indeed I cant' resist mentioning this – when Richard Nixon was about to nominate Rehnquist to the US Supreme Court, around 1972, Nixon got his name wrong all the time. We know from his memoirs he called him Renschberg, not Rehnquist. According to ANES, Richard Nixon is not very well informed.

I would put in a little footnote the sort of irritation one feels about the confidentiality stuff. It has been a sledge hammer in the past. By that I mean there has been no serious effort to examine the open-ended content to see if it is worthy of confidentiality concerns. It is a hammer approach in which the same rules apply across the board. You may say to me it is too expensive to do each individual request and I might accept that but it really has been a hammer approach and therefore, as I say, this material, according to the contract, cannot be put forth in public.

Well I did in fact get the data and I did my own recoding. I did not accept head honcho as a correct answer but I did accept a variety of answers that said he was a member of the Supreme Court, he is the chief justice of the court – not the Supreme Court because that would be counted as wrong, and with pretty strict coding. Not head honcho but still pretty strict coding; I determined that 72% of the responses judged to be incorrect are at a minimum partially correct and in my opinion, almost perfectly correct. That is a lot of measurement error, it seems to me. The fact that you can't see that very well – it is 28%; 28% are in fact truly incorrect. Don't let me mislead you because there still is a big category of don't know respondents. I worry a lot about those don't know respondents but these are people who offered a response and was judged by the coder to be incorrect.

This "don't know" issue is important. I don't believe it is the practice and I could be wrong but I don't believe it is a practice to probe don't knows on the knowledge questions. I have seen no data to indicate the results of such probing. I should have emphasized this – it is a part of my coding – of the 349 incorrect responses – 91 identified Rehnquist as the Supreme Court Justice, 54 as a Supreme Court Judge, and 61 instances the only recorded verbatim was the Supreme Court. In fact I treated all of those as correct. It is hard to code these items by the way. Just this morning I pulled out some of the verbatim text from the files I received from ANES. Here you see the main judge. I did this to sort of to make everyone aware of the nature; at least in this case, of the verbatim responses because if you all are thinking about machine coding data, I think there are a number of problems in here. Surely colloquialisms, the misspelling and all, that we can clean up, but the colloquialisms and implicit meaning, I think, and I will defer to your judgment, but I think these could be pretty difficult items to subject to some sort of machine coding.

There is also the problem of don't know. As I said, the don't knows are not probed but there are some coding instructions that ANES uses that worries me a lot. The DK key is not allowed for this question. That is a telling – I think the data entry people but maybe the interviewer, I'm not exactly certain on this – the DK is not allowed for this question but in fact a large number of people recorded it as making no attempt to guess. Here is what I see as the problem – people are learning, every respondent, despite the fact that DK is not an option, they are learning at some point in the instrument that they can say don't know and get away with it. At that point forward, don't know is an implicit part of the response set. Furthermore, I don't think people are randomly discovering the availability of DK. I think it is a systematic finding. Maybe there are instances where I am wrong but the knowledge items have not typically been rotated and so the response set – the availability of don't know is itself a variable of the items that are presented to the respondents to measure knowledge.

I don't want to go too far down the road I am about to go because I just want to show you some alternative results of knowledge measures that are based on a different approach. These are closed-ended questions that I have asked in my own surveys in the last decade or so. They are not focused on personalities; rather they are focused on the attributes of the Supreme Court itself. And sort of the conceptualization here is that it is probably more politically significant, in my view, to know that a Supreme Court Judge cannot be held accountable in any way whatsoever for his or her policies. It is more significant to know that they have a life term than to know who the occupants of the position are at any given moment in time. These are not great figures but they pale in comparison to the 28% that you saw for the Rehnquist figure earlier on. We have to be careful from a substantive point of view that people know a lot about the US Supreme Court even if they may get the occupants wrong. Now you are saying they are guessing, right? In fact, in that battery of three questions, the joint probability of getting all 3 correct is only 8.25. In fact, we find that 44.4% of nationally representative samples get those answers correct. Obviously there is the closed-ended problem of guessing and lots of other problems that you all will probably point out. But the substantive conclusions are just dramatically different.

Indeed, in a recent survey, I did a very simple experiment – splitting the ballot between open-ended and closed-ended. First you will see that Rehnquist is very much better known than Chief Justice Roberts. Rehnquist has been around for 30 years, Roberts for 3, so that difference is important. But look at this difference between the open-ended and the closed-ended – 15% could volunteer Robert's name, 46% could recognize Robert's name. 42% could volunteer Rehnquist's name, 71% could recognize the name. It is fair to say that I don't ask a whole lot of open-ended questions in my own surveys because they are just too damn expensive. They are too expensive from the upfront costs, they are too expensive from the coding and rear end costs. But we know for sure that the question form has a dramatic impact on the nature of the responses.

Let me draw a few conclusions and points that we might want to talk about over the course of the next couple of days. I'm sure that some of you are there thinking that doing point estimates from survey data is a silly enterprise in the first place. We can fight about that probably. But the real truth is that people are doing it all the time. They are drawing substantive conclusions about levels of this and levels of that and they are doing it on the basis of point estimates. We have got to be sensitive to the incredibly substantive implications of responses to our items. I think we ought to rethink the conceptualization of political knowledge. I'm sure that Skip and Jon are going to say in a minute that they are doing so, especially Skip, who is writing some really, really extraordinarily interesting new thinking about political knowledge. It is not so much what people are able to go upstairs and bring down and spit out of their mouth but rather it may be the political knowledge is the ability to solve problems. It is the ability to gather information when it becomes important and salient for gathering that information. I promise you I literally could not tell you the name of the governor of Missouri. I could and no one would know if it was right or not so I could get away with it but I couldn't recall the name of the governor -- but I could find it out if it became in any sense useful for me to know. I think that re-conceptualizing, in some ways, is even more important than the operational sorts of problems. As I have already mentioned, I question whether the personalities are really the best – in my own work I care about institutions, I care about things like legitimacy institutions to convince people to accept decisions with which they disagree. Those are not matters of personality. They are matters of procedures, procedural justice, legitimacy and so on.

I think everyone in this room will agree that consistency in coding practices is just about as important as consistency in question wording. To be absolutely honest, I mucked about in the open-ended verbatims for awhile. In the end, I tried to draw minimum conclusions and bail out. I just wasn't convinced that there was enough information in there to really do a very systematic kind of analysis. Let me be clear, this is not a blanket indictment of ANES. This is the examination of 4 knowledge items. But boy in the end I just wasn't convinced that there was enough there to put the effort into it. Anyone who is looking at this question over time really should be careful. I think strict instructions are not necessarily the best key to reliability. As I say, we coded the substantive content of every Supreme Court opinion and there are – you know better than I – there are methods by which highly subjective material can in fact be coded. And then finally, I would like to see further research on the question of the difference between those who get stuff partially correct and those who get it fully correct. You know what my hypothesis is. I think we need to do a lot more work on the substantive and the theoretical implications of these different approaches of measuring political knowledge. Thank you very much for listening.