

ORIGINAL ARTICLE

The Validity of the Enns and Koch, and Berry et al. Measures of State Policy Mood: Continuing the Debate

William D. Berry¹ , Richard C. Fording², Russell L. Hanson³ and Justin K. Crofoot¹

¹Department of Political Science, Florida State University, Tallahassee, FL, USA

²Department of Political Science, University of Alabama, Tuscaloosa, AL, USA

³Department of Political Science, Indiana University, Bloomington, IN, USA

Corresponding author: William D. Berry; Email: wberry@fsu.edu

(Received 25 November 2019; Revised 30 March 2020; Accepted 14 May 2020; published online 18 October 2021)

Abstract

Enns and Koch question the validity of the Berry, Ringquist, Fording, and Hanson measure of state policy mood and defend the validity of the Enns and Koch measure on two grounds. First, they claim policy mood has become more conservative in the South over time; we present empirical evidence to the contrary: policy mood became more liberal in the South between 1980 and 2010. Second, Enns and Koch argue that an indicator's lack of face validity in cross-sectional comparisons is irrelevant when judging the measure's suitability in the most common form of pooled cross-sectional time-series analysis. We show their argument is logically flawed, except under highly improbable circumstances. We also demonstrate, by replicating several published studies, that statistical results about the effect of state policy mood can vary dramatically depending on which of the two mood measures is used, making clear that a researcher's measurement choice can be highly consequential.

Keywords: ideology; public opinion; measurement; policy mood; state policy mood; replication analysis; MRP

Since the publication of Enns and Koch's (2013) paper introducing a measure of state policy mood constructed using multilevel regression and poststratification (MRP) and survey aggregation, we have debated with Enns and Koch (hereafter E&K) about the validity of their measure of policy mood and of our own measure (Berry et al. 1998 [hereafter BRFH]). The debate has unfolded in two articles: Berry et al. (2015), followed by Enns and Koch (2015). Our first purpose is to continue the debate. However, since much has already been written by both sets of authors, we limit our response to Enns and Koch (2015) to assessing two arguments they raise that they did not make in their 2013 piece, and that we believe are flawed. Our second purpose is to report replication analysis that sheds light on the extent to which the substantive findings of research vary

© The Author(s), 2021. Published by Cambridge University Press on behalf of the American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

depending on one's choice about which measure of state policy mood to use; we demonstrate that this measurement choice can be highly consequential, with substantive conclusions changing dramatically with a shift in the measure used.

Contrary to E&K's Claim, Public Opinion Data Indicate that Policy Mood has been Liberalizing in the South

Enns and Koch (2015) consider the longitudinal characteristics of the two state policy mood measures in the South. Their figure 2a shows that the BRFH measure indicates that policy mood has become more liberal in southern states in recent decades, and that the E&K measure shows mood in the South has become more conservative. E&K claim that their measure has greater face validity than the BRFH measure because of two trends that they believe indicate that policy mood in the South has become more conservative: in this region, (i) partisan affiliation has become "increasingly Republican" (p. 441) and (ii) welfare benefits have declined (their figure 2b). The fact that political scientists lack a measure of state policy mood known *with certainty* to be valid means that we cannot know for sure how *true* policy mood has changed in the South, and thus, that we cannot definitively evaluate E&K's claim. However, there are two important reasons why we should question E&K's claim.

First, the methodology employed by E&K to construct their measure—MRP—is not well-suited for estimating change in public opinion over time. MRP was introduced by Gelman and Little (1997) and has become increasingly popular among state politics scholars due to the dearth of reliable state-level public opinion data. The method was originally designed to estimate public opinion for a single period and its validity as a cross-sectional technique has been supported by several studies (Buttice and Highton 2013; Lax and Phillips 2009; Park, Gelman, and Bafumi 2004; Warshaw and Rodden 2012). Although many scholars (in addition to E&K) have used the MRP approach to create longitudinal estimates of state public opinion, Gelman et al. (2018, 2) question the validity of such measures due to the fact that the original MRP approach "fails to make use of all the available data and employs arbitrary assumptions as to how much change occurs over time."

Second, E&K's claim that policy mood in the South has become more conservative rests on the assumption that partisanship and welfare benefits are strong proxies for policy mood. At best, this assumption is dubious. Although partisanship and ideological self-identification (i.e., symbolic ideology) have followed similar trends over the last several decades, this is not the case for partisanship and policy mood. Indeed, upon comparing trends in macropartisanship and policy mood, Erikson (2012, 42) concluded that "the two time series are virtually uncorrelated." And although welfare benefits certainly have steadily declined in the South since the early 1970s, several studies have shown that this has been the case in all states for reasons that have little to do with policy mood (Berry, Fording, and Hanson 2003; Peterson and Rom 1990; Soss, Fording, and Schram 2011). A much better way to assess the plausibility of the BRFH measure-derived finding of a liberalizing trend in policy mood in the South is to observe how public opinion about ideologically relevant issues has changed over time in the South, thereby relying on information about public opinion to directly assess the public's policy mood.

We obtained General Social Survey (GSS) data from 1973 to 2010 on a set of eight items asking respondents if the government is "spending too much money," "too little money," or "about the right amount" across a diverse set of program areas: welfare,

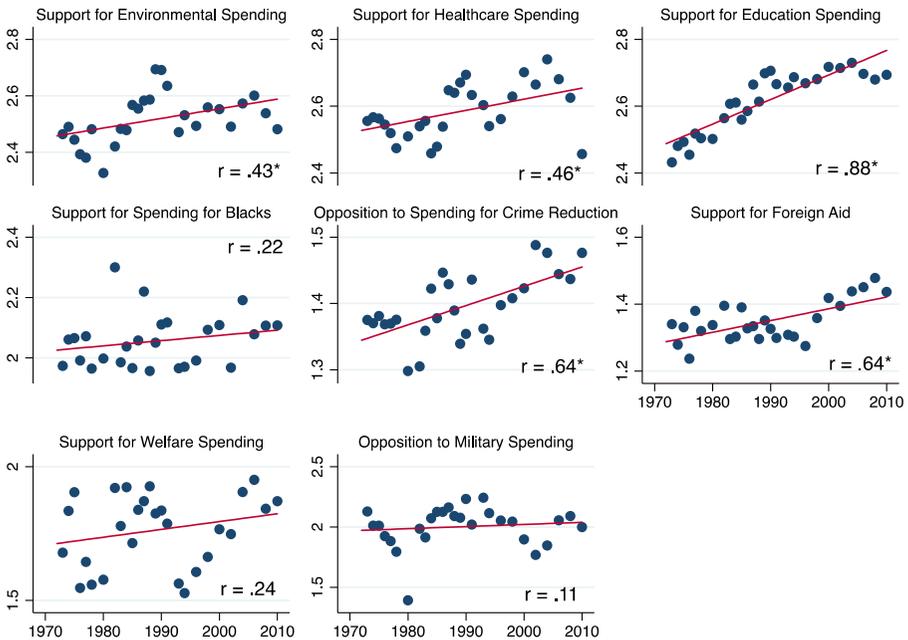


Figure 1. Trends in General Social Survey spending items in the South, 1973–2010. *Note.* To make individual plots easy to read, the scale for the vertical axis varies across plots. Thus, one cannot draw conclusions about the relative slopes of regression lines by simple visual comparison across plots.

* $p < 0.05$ (two-tailed test).

healthcare, education, improving the conditions of Blacks, environmental protection, crime, defense, and foreign aid.¹ We also secured 10 GSS items concerning issues generally thought to be related to ideology: abortion, gay rights, gun control, aid for blacks, the treatment of criminals, government redistribution, tax policy, and healthcare for the poor.² For each of the 18 items, the scores for responses were linearly transformed to range between 1 and 3, with higher scores indicating greater liberalism. In a set of figures, we analyze trends in the South in each item, as well as in four policy mood indexes we construct primarily from these items.³

¹(a) Document S1 in the Supplementary Material contains the specific wording of all GSS items used in our analyses.

(b) We chose 1973 as the first year for analysis because it is the earliest year for which GSS data are available; we end analysis in 2010 to conform to Enns and Koch's (2015) period of analysis.

²These 18 GSS items constitute each question in the GSS cumulative data file that (i) we believe would be widely viewed as reflecting a respondent's operational ideology and (ii) was asked of respondents for the first time no later than 1980, and regularly thereafter through at least 2010.

³In a regional analysis in Berry et al. (2015), we define the "South" as the 11 states of the Confederacy. To maintain consistency, we would prefer to stick with this definition. However, the GSS codes respondents' residential location using the Census definition of the South (as including Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia), and so we must conform to this alternative conception.

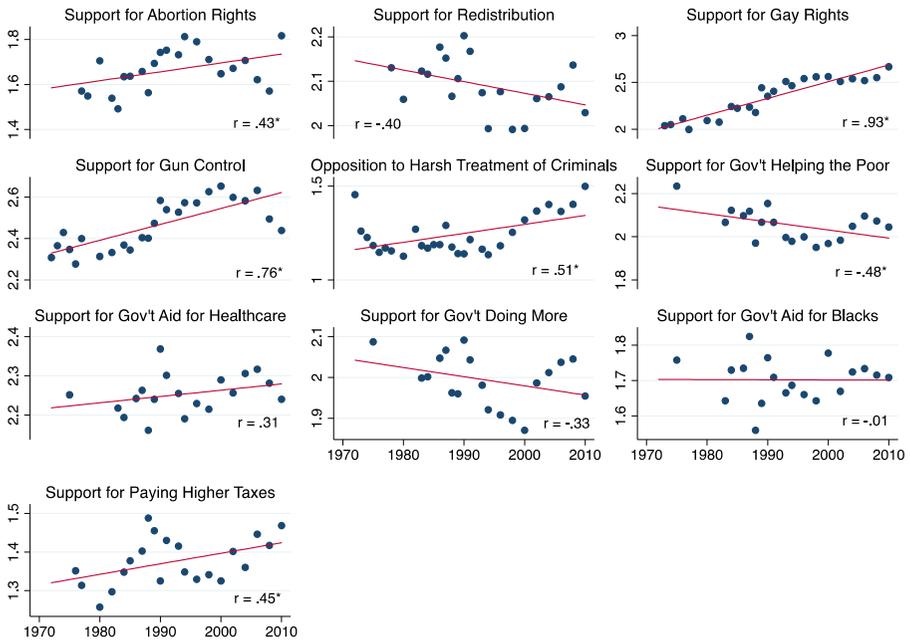


Figure 2. Trends in General Social Survey non-spending items in the South, 1973–2010. *Note.* The year of the first observation varies (1973–1978). To make individual plots easy to read, the scale for the vertical axis varies across plots. Thus, one cannot draw conclusions about the relative slopes of regression lines by simple visual comparison across plots.

* $p < 0.05$ (two-tailed test).

For each of the eight items based on government spending preferences, [Figure 1](#) presents a plot of the average score for the item among respondents from the South against the year of observation—overlaid with the ordinary least squares (OLS) bivariate regression line, and the correlation between the average item score and a year-count variable. A positive correlation indicates that the South is liberalizing over the period of analysis, and a negative correlation implies that the South is becoming more conservative. [Figure 2](#) presents similar plots based on the 10 additional GSS items that are not about spending.

The trends displayed in [Figures 1](#) and [2](#) vary substantially in strength across the 18 GSS items, but for the great majority (14) of the 18 plots, the overall trend is increasingly liberal policy mood over time, as reflected in a positive correlation between the opinion item and the year of observation. Moreover, the positive correlation is statistically significant (0.05 level, two-tailed test) in 10 of the 14 plots. In contrast, for only one item—support for “government helping the poor” is the correlation negative and statistically significant.

We also constructed four indexes of policy mood based on alternative conceptualizations of ideology. These indexes were based on (i) all 18 items in [Figures 1](#) and [2](#) (“Policy Liberalism Index”), (ii) the 8 spending items presented in [Figure 1](#) (“Support for Spending Index”), (iii) the 10 nonspending items in [Figure 2](#) (“NonSpending

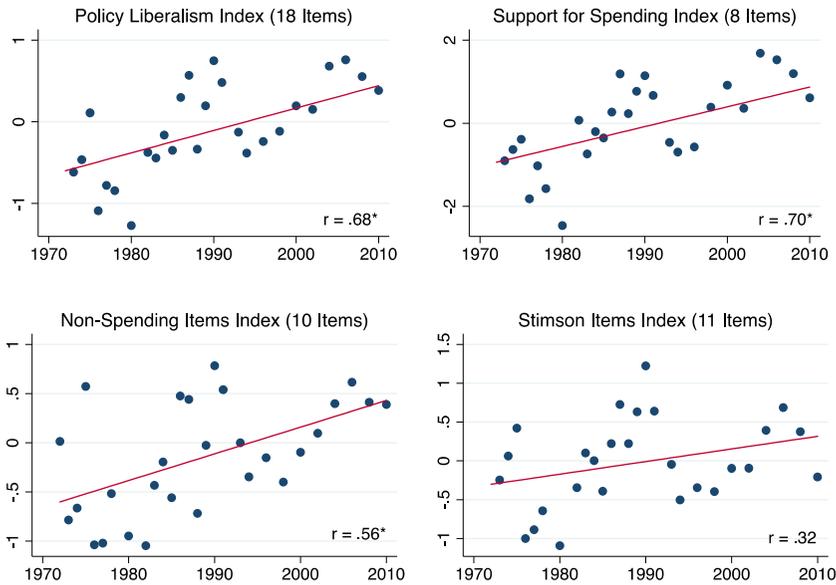


Figure 3. Trends in Alternative Indexes of policy mood in the South, 1973–2010
Note. To make individual plots easy to read, the scale for the vertical axis varies across plots. Thus, one cannot draw conclusions about the relative slopes of regression lines by simple visual comparison across plots.
 * $p < 0.05$ (two-tailed test).

Items Index”), and (iv) 11 items measuring attitudes about the scope of government used by Stimson (1991) to construct his measure of policy mood (“Stimson Items Index”).⁴ Each of the indexes was constructed based on the unweighted average of its component GSS items, each of which were standardized (linearly transformed to have a mean of 0 and a variance of 1) prior to creating the index.⁵

Plots of the four policy mood indexes over time are presented in Figure 3. In every case, the policy mood index is positively correlated with time, indicating increasing southern liberalism over time. Across the four plots, the year-mood correlations range from a low of 0.32 (Stimson Items Index) to 0.70 (Support for Spending Index),

⁴Although Stimson uses more than 11 GSS items to construct his version of policy mood, we restrict our analysis to the items that were asked of respondents for the first time no later than 1980, and regularly thereafter through at least 2010 (the same rule we applied when choosing the items analyzed in Figures 1 and 2). The 11 items include 4 of the spending items in Figure 1 (support for spending on welfare, healthcare, education, and the environment), 5 of the nonspending items in Figure 2 (support for redistribution, helping the poor, government aid for health care, the government doing more, and paying higher taxes), as well as 2 items that we did not include among those analyzed in Figures 1 or 2 because we believed they did not unambiguously reflect the liberal-conservative ideological divide: support for spending on drug addiction programs, and support for spending on issues facing big cities. We included these two items in the Stimson Items Index despite our concern about their face validity because our goal is to construct an index of policy mood as close as possible to Stimson’s mood measure given the data constraints we face.

⁵Because different items have missing values in different years, each index’s value is an average over a set of items that varies from one year to the next.

and are statistically significant at the 0.05 level in all but one case (Stimson Items Index).⁶ These results are consistent with the southern trend in BRFH's policy mood measure, but inconsistent with the trend in the South in E&K's measure of mood.⁷ Thus, our empirical evidence leads us to reject E&K's assertion that BRFH scores for southern states lack face validity; to the contrary BRFH scores conform nicely to available evidence about changes in the South in public opinion, while E&K scores do not.

Contrary to E&K's Claim, Cross-Sectional Performance of the State Policy Mood Measures is Relevant

In our 2015 *SPPQ* paper, we argue that Enns and Koch's (2013) state policy mood scores lack face validity in cross-sectional comparisons. In their reply, Enns and Koch (2015) do not challenge this claim. Rather, they question the relevance of the cross-sectional performance of their measure based on an argument that the "standard approach of including state fixed effects in cross-sectional time-series models... means that most analyses focus explicitly on *over-time* (within state) relationships" [emphasis added] (Enns and Koch 2015, 440). We believe this argument misses the point: it is inappropriate to dismiss cross-sectional characteristics of the E&K measure that lack face validity as irrelevant *even if* one uses the E&K measure solely for pooled cross-sectional time-series analyses specifying state fixed effects (so that the only relevant variation is longitudinal). Our contention is premised on the fact that we see no reason to believe that E&K's methodology could yield measurement error sufficient to invalidate cross-sectional comparisons without simultaneously invalidating longitudinal comparisons.

Consider the following thought experiment. Assume that we know the *true* value of policy mood in each state in each year over a long period. Denote this true value in state s in year t ($t = 1, 2, \dots, T$) by $TrueMood_{s,t}$. We create an imperfect measure of policy mood, $ObservedMood_{s,t}$, by introducing systematic (i.e., nonrandom) measurement error. Specifically, for each state-year, we adjust the true score (some up, some down) by an amount sufficient to substantially distort cross-sectional comparisons of policy mood in each year. Denote the amount of the adjustment to the value of true mood in state s in year t —that is, the amount of error in observing $TrueMood_{s,t}$ —by $Error_{s,t}$. Given this notation,

$$ObservedMood_{s,t} = TrueMood_{s,t} + Error_{s,t}.$$

Consider the case in which for each state, the amount of error is *stable across years*, that is, for each state, s , $Error_{s,1} = Error_{s,2} = \dots = Error_{s,T}$. In this special case, the measurement error introduced would not distort longitudinal comparisons in any state. This is because for any state, s , and any two years, $t1$ and $t2$,

⁶The p -value for the correlation based on the Stimson Items Index is 0.11.

⁷In contrast, as shown in Figure S1 in the Supplementary Material, the GSS's 7-point measure of ideological self-placement (i.e., symbolic ideology) shows the South becoming more conservative over the years, reinforcing the frequently-made observation that symbolic ideology is a concept distinct from policy mood (Berry 2007; Ellis and Stimson 2009; Stimson 1991). The figure also shows that the GSS's 7-point indicator of party identification exhibits a clear trend of increasing Republicanism in the South.

$$\text{ObservedMood}_{s,t2} - \text{ObservedMood}_{s,t1} = \text{TrueMood}_{s,t2} - \text{TrueMood}_{s,t1}.$$

However, with *any other pattern of measurement error*—that is, with any departure from error that is stable across years—the measurement error would distort not only within-year cross-sectional comparisons of mood, but also within-state longitudinal comparisons.

Without knowing the exact nature of the measurement error in E&K mood scores that produces what we have contended are distorted cross-sectional comparisons (a contention E&K have not disputed), one cannot know with certainty whether this measurement error would also distort longitudinal comparisons. However, it seems implausible to us that the error in E&K's measure would be stable across the more than 50 years E&K have observed—which we have shown is the only condition under which longitudinal comparisons of E&K scores would be shielded from distortion.⁸ This would imply that even if one cares only about longitudinal variation in state policy mood, one cannot dismiss evidence of poor cross-sectional performance of the E&K measure because its poor cross-sectional performance signals the presence of measurement error that is likely to undermine longitudinal comparisons as well.⁹

Conclusion about Enns and Koch's and Berry et al.'s Measures of Policy Mood

Nothing in the response by Enns and Koch (2015) to our *SPPQ* paper (Berry et al. 2015) leads us to retract any of our arguments; we continue to stand behind the claims in our paper. We remain doubtful that E&K's measure is valid, largely because its characterization of mood in the states departs substantially from conventional wisdom and current scholarship; and we continue to believe that the Berry et al. (1998) indicator is a reasonable proxy for policy mood that fares well on a variety of reliability, face validity and construct validity tests described in this article and in previous papers (Berry et al. 1998; 2007; 2015). As a consequence, we think the BRFH measure can serve the needs of state politics scholars until a superior measure based on public opinion surveys is developed.¹⁰

The Implications of the Choice about what Measure of State Policy Mood to Use on Research Results

Across all state-years in which both the BRFH and E&K measures of state policy mood are available (observations for each year during the period 1960–2010), the

⁸At a minimum, since “stability in error across years” is a very strong assumption, it should not be made in the absence of an affirmative argument that it is plausible.

⁹A second reason we find it puzzling that E&K question the relevance of the cross-sectional performance of their mood measure is that in the article in which Enns and Koch (2013) introduce their mood measure (and other MRP-based measures), they rely on cross-sectional evidence to validate their indicators of state partisanship and state symbolic ideology (see the authors' tables 1 and 2). If it is appropriate for E&K to use cross-sectional tests of convergent validity to assess measures of state partisanship and state symbolic ideology, we see no reason E&K should challenge our reliance on cross-sectional analysis to evaluate measures of state policy mood.

¹⁰In contrast, the Berry et al. (1998) indicator estimates state policy mood from interest group ratings of members of Congress and the distribution of votes for candidates in congressional elections, based on the assumption that voters choose the candidate whose ideology is closest to their own.

correlation between the two measures is just 0.10 (Berry et al. 2015, 2).¹¹ This suggests a strong possibility that a researcher estimating a model including one of the two measures as an independent variable would often derive substantially different results about the effect of policy mood if she used the other measure instead. Also, to the extent that other independent variables in a model are correlated with at least one of the measures of policy mood, the estimated effects of these other variables may also be sensitive to the choice about which measure of policy mood to use.

In this section, we report the results of replications of several published studies to empirically assess the extent to which results from models including state policy mood as an independent variable vary depending on the measure of mood employed.¹² To identify a sample of studies to replicate, we used the ISI Web of Knowledge search mechanism to identify each article (i) published between 2013¹³ and 2019 in a “political science” journal with a 2017 JCR Impact Factor of at least 1.0¹⁴ and (ii) that cites one of the papers introducing the two policy mood measures: Berry et al. (1998) or Enns and Koch (2013). This search yielded 99 articles. One of us visually scanned each of these articles to identify the subset that report empirical analysis in which one of the two measures of state policy mood is used as an independent variable in an econometric model.¹⁵ On practicality grounds, we restricted our analysis to articles for which replication data were publicly available, and for which executing author-provided code allowed us to reproduce published results.¹⁶ This winnowed our sample to seven articles. To avoid artificially inflating the number of distinct models we replicate by including some models that are minor “tweaks” of another, when an article estimated multiple models including state policy mood, we randomly chose one model for replication.¹⁷

¹¹Given the frequency of studies doing pooled cross-sectional time-series analysis specifying state fixed effects, it is relevant to determine the similarity of the two measures in a fixed-effects context. To do so, we calculate the pooled bivariate correlation between the measures *after each was de-meaned* (i.e., after scores on each variable were transformed to deviations from their within-state mean). After de-meaning, the correlation between the two mood measures over the pooled observations remains nearly zero: -0.06 .

¹²We limit consideration to the BRFH and E&K measures, and do not consider Caughey and Warshaw’s (2018) measures of “mass liberalism.” This is because Caughey and Warshaw do not construct a measure of “overall” policy mood; instead they construct separate measures of two dimensions of mood: “economic” and “social.”

¹³The year 2013 was chosen because E&K introduced their measure in this year.

¹⁴The threshold of 1.0 was selected with the knowledge that it would result in the inclusion of *American Journal of Political Science*, *American Political Science Review*, *Journal of Politics*, and *State Politics and Policy Quarterly*. But this threshold also picked up *Legislative Studies Quarterly*, *Political Research Quarterly*, and several other journals.

¹⁵To restrict the analysis to models for which it would be easy to assess how findings differ when one substitutes one measure of policy mood for the other, we use models that contain only a single term involving the mood measure (thereby eliminating, e.g., both nonlinear models containing both the mood measure and the measure squared, and interactive models containing both the mood measure and the product of mood with some other variable).

¹⁶We also eliminated from consideration one article (Pacheco 2021) in which the average value of policy mood in each state over a number of years was computed, and then a state’s average policy mood was used as an independent variable in a cross-sectional model. We eliminated this article because its use of a multi-year measure of policy mood in cross-sectional analysis is highly unusual in state politics and policy research.

¹⁷This procedure led us to replicate a model from each of Boehmke and Shipan (2015), Boehmke, Osborn, and Schilling (2015), Hannah and Mallinson (2018), Hawes and McCrea (2018), Hayes and Dennis (2014),

For each of the seven models to be replicated, using the authors' data—and minimally changing their Stata or R code—we re-estimated the model twice, once using the BRFH measure of policy mood, and once using the E&K measure of mood.¹⁸ For five of the seven models, the finding about the effect of policy mood varies substantially with the measure of mood employed. As can be seen in Table 1, for one of these five models (Hayes and Dennis), the coefficient for policy mood is statistically significant at the 0.05 level in both the BRFH version of the model and the E&K version, but positive in one and negative in the other. In three other models (Boehmke and Shipan 2015; Hawes and McCrea 2018; Ojeda et al. 2019), the coefficient for mood is statistically significant at the 0.05 level in one version of the model, and far from significant in the other (with a p -value greater than 0.80).¹⁹ In a fifth model (Boehmke, Osborn, and Schilling 2015), the coefficient for mood is positive and significant at the 0.10 level in one version, and weakly negative in the other. In the remaining two models, the difference in results across versions is less stark. In the Hanna and Mallinson model, there is a positive coefficient for mood in both versions, but the p -values are nontrivially different (0.07 and 0.36); in the Taylor model, the coefficient for mood is positive for one measure of ideology and negative for the other, but neither close to statistical significance at the 0.05 level (with p -values of 0.52 and 0.59).

We can also consider whether the choice of the measure for policy mood affects the coefficient estimates for other independent variables in a model. There are 70 non-mood independent variables in the seven replicated models together.²⁰ As Table S1 in the Supplementary Material shows, in the vast majority of cases (i.e., the 61 rows that are not shaded in gray), the coefficient for a variable is either (i) statistically significant at the 0.05 level with the same sign in both versions of the model or (ii) statistically insignificant at the 0.05 level in both versions. However, in two of these 61 cases (see rows 3 and 30 of Table S1), one coefficient is statistically significant at the 0.10 level and the other is far from statistically significant (with a p -value of 0.45 or 0.90). On the other hand, there are also five cases (see rows 27, 31, 64, 66, and 68) among the nine shaded rows in which a variable's coefficient is

Ojeda et al. (2019), and Taylor, Haider-Markel, and Rogers (2019); the specific model from each article is listed in the left-most column of Table 1, and the dependent variable is listed in the next column.

¹⁸To facilitate comparison of results across models with different measures of mood, for each model, we deleted all observations falling outside the period for which both measures are available, 1960–2010. Then, for the resulting period of analysis, we linearly transformed the BRFH mood measure, the E&K mood measure (both of which are scaled such that higher values indicate greater liberalism), and the dependent variable of the model to the range between 0 and 1.

¹⁹Since the continuous dependent variable in Hawes and McRea's regression model—state welfare generosity—is linearly transformed to the range between 0 and 1, as are both measures of policy mood, the magnitudes of the two mood coefficients are comparable. In the “E&K model,” the coefficient for mood implies that an increase in liberalism across the full range of mood scores (i.e., from 0 to 1) is associated, on average, with a *substantively-consequential decrease* in state welfare generosity equivalent to 7% of the range of generosity scores. In contrast, in the “BRFH model,” the coefficient for mood implies that the same increase in liberalism is associated with a *trivial increase* in state welfare generosity—equivalent to less than 1% of range of generosity scores.

²⁰We exclude from consideration (i) state and time dummy variables since researchers rarely treat the coefficients for these variables as quantities warranting substantive interpretation and (ii) variables that are contained in multiple terms in the model, which would complicate the interpretation of the coefficients for the variables.

Table 1. The implications of the choice about how to measure state policy mood on the estimated impact of mood in seven published articles

Replicated article and model	Dependent variable ^a	Coefficient for BRFH measure of policy mood (higher = more liberal)	p-value for BRFH measure	Coefficient for E&K measure of policy mood (higher = more liberal)	p-value for E&K measure
Boehmke and Shipan (2015): Model 3 in table 2 (a pooled cross-sectional time-series negative binomial model that includes state and year fixed effects)	Number of deficiencies found in nursing home inspections	0.032	0.89	0.358**	<0.001
Boehmke, Osborn, and Schilling (2015): third model from the left in table 3 (a longitudinal negative binomial model that includes a year count variable and its square)	Number of citizen-sponsored initiatives (+)	1.14*	0.08	-0.468	0.42
	Whether a state adopts medical marijuana policy (+)	4.74*	0.07	3.74	0.36
Hannah and Mallinson (2018): logit event history analysis model in table 2 (that includes a year count variable)	State welfare generosity (+)	0.0082	0.80	-0.0722**	0.02
Hawes and McCrea (2018): Model 2 in table 2 (a pooled cross-sectional time-series regression that includes state and year fixed effects)	Whether a state permits deduction of fed. income tax against state income tax (-)	-3.42**	< 0.001	6.97**	< 0.001
Hayes and Dennis (2014): Cox proportional hazards model in table 2	Number of TANF work exemptions	0.008	0.91	0.109**	0.049
Ojeda et al. (2019): random effects GLS model (that includes year fixed effects) for "Work Exemptions" in table 5	Whether a state adopts a sexual orientation employment discrimination law (+)	-3.47	0.52	2.59	0.59
Taylor, Haider-Markel, and Rogers (2019): Model 4 (Cox proportional hazards) in table 4					

Abbreviations: BRFH, Berry et al.; E&K, Enns and Koch.

^aIf an article's author(s) offer an explicit hypothesis about the direction of the effect of policy mood on the dependent variable, the predicted direction (+ or -) is enclosed in parentheses after the dependent variable listed below.

**Statistically significant at the conventional threshold ($p < 0.05$) in political science research.

*Not statistically significant at the conventional level, but would be significant using a slightly higher threshold ($p < 0.10$).

statistically significant at the 0.05 level in one version of the model, and nearly significant in the other version. In three other cases (see rows 49, 52, and 53), a variable's coefficient is statistically significant at the 0.05 level in one version and not even close to significant in the other (with a p -value of 0.98, 0.99, or 0.49, respectively). The most striking difference in results is a case (see row 46) in which the coefficient for a variable is statistically significant at the 0.05 level in both versions of the model, but is positive in one and negative in the other.²¹

Since we replicate just seven studies, our sample cannot be assumed to be representative of the universe of research in which state policy mood has been used as an independent variable. On the other hand, we chose the studies using a procedure that guarantees that they were not "cherry picked" to produce results of one kind or another. We believe that it is evident from our replications that the decision about which measure of state policy mood to use when doing research should not be made casually. There is clearly a substantial risk that one's choice about which measure of policy mood to use will have a large impact on one's finding about the effect of policy mood. There is also at least a small risk that one's choice of mood measure will affect one's findings about the effects of other variables included in one's model. Thus, even if policy mood is being employed solely as a control variable that allows one to derive an unbiased estimate of the effect of some other variable, one cannot safely assume that the choice about how to measure mood is inconsequential.²²

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/spq.2021.23>.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at <https://doi.org/10.15139/S3/CQTBQU> (Berry et al. 2021).

Funding Statement. The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 2007. "The Measurement and Stability of State Citizen Ideology." *State Politics and Policy Quarterly* 7: 111–32.
- Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 2015. "Assessing the Validity of Enns and Koch's Measure of State Policy Mood." *State Politics and Policy Quarterly* 15: 425–35.
- Berry, William D., Richard C. Fording, and Russell L. Hanson. 2003. "Reassessing the 'Race to the Bottom' in State Welfare Policy." *Journal of Politics* 65: 327–49.
- 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: 327–48.

²¹We believe the sample of studies we replicate is insufficiently large to justify any claims about specific variables that are especially prone to have coefficients that are sensitive to the choice about the measure for policy mood.

²²When state policy mood is included as a control variable in a model designed to estimate the effect of some other independent variable, we do not believe that it is always appropriate to draw causal inferences about the impact of state policy mood. Keele, Stevenson, and Elwert (2020) discuss how researchers can identify which parameter estimates from a regression model warrant causal interpretation.

- Berry, William, Richard Fording, Russell Hanson, and Justin K. Crofoot. 2021. "Replication Data for: The Validity of the Enns and Koch, and Berry et al. Measures of State Policy Mood: Continuing the Debate." UNC Dataverse. Dataset. <https://doi.org/10.15139/S3/CQTBQU>.
- Boehmke, Frederick J., Tracy L. Osborn, and Emily U. Schilling. 2015. "Pivotal Politics and Initiative Use in the American States." *Political Research Quarterly* 68: 665–77.
- Boehmke, Frederick J. and Charles R. Shipan. 2015. "Oversight Capabilities in the States: Are Professionalized Legislatures Better at Getting What They Want?" *State Politics and Policy Quarterly* 15: 366–86.
- Buttice, Matthew K., and Benjamin Highton. 2013. "How Does Multilevel Regression and Poststratification Perform with Conventional National Surveys?" *Political Analysis* 21: 449–67.
- Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014." *American Political Science Review* 112: 249–66.
- Ellis, Christopher, and James A. Stimson. 2009. "Symbolic Ideology in the American Electorate." *Electoral Studies* 28: 388–402.
- Enns, Peter K., and Julianna Koch. 2015. "State Policy Mood: The Importance of Over-Time Dynamics." *State Politics and Policy Quarterly* 15: 436–46.
- Enns, Peter K., and Julianna Koch. 2013. "Public Opinion in the U.S. States: 1956–2010." *State Politics and Policy Quarterly* 13: 349–72.
- Erikson, Robert S. 2012. "Public Opinion at the Macro Level." *Daedalus* 141: 35–49.
- Gelman, Andrew, Jeffrey Lax, Justin Phillips, Jonah Gabry, and Robert Trangucci. 2018. "Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion." unpublished manuscript, August 28, 2018.
- Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23: 127–35.
- Hannah, A. Lee, and Daniel J. Mallinson. 2018. "Defiant Innovation: The Adoption of Medical Marijuana Laws in the American States." *Policy Studies Journal* 46: 402–23.
- Hawes, Daniel P., and Austin Michael McCrea. 2018. "Give Us Your Tired, Your Poor and We *Might* Buy Them Dinner: Social Capital, Immigration, and Welfare Generosity in the American States." *Political Research Quarterly* 71: 347–60.
- Hayes, Thomas J., and Christopher Dennis. 2014. "State Adoption of Tax Policy: New Data and New Insights." *American Politics Research* 42: 929–55.
- Keele, Luke, Randolph T. Stevenson, and Felix Elwert. 2020. "The Causal Interpretation of Estimated Associations in Regression Models." *Political Science Research and Methods* 8: 1–13.
- Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53: 107–21.
- Ojeda, Christopher, Anne M. Whitesell, Michael B. Berkman, and Eric Plutzer. 2019. "Federalism and the Racialization of Welfare Policy." *State Politics and Policy Quarterly* 19: 474–501.
- Pacheco, Julianna. 2021. "The Policy Consequences of Health Bias in Political Voice." *Political Research Quarterly* 74: 46–58.
- Park, David K., Andrew Gelman and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12: 375–85.
- Peterson, Paul E., and Mark C. Rom. 1990. *Welfare Magnets: A New Case for a National Standard*. Washington, DC: Brookings Institution Press.
- Soss, Joe, Richard C. Fording and Sanford F. Schram. 2011. *Disciplining the Poor: Neoliberal Paternalism and the Persistent Power of Race*. Chicago: University of Chicago Press.
- Stimson, James A. 1991. *Public Opinion in America: Moods, Cycles, and Swings*. Boulder, CO: Westview Press.
- Taylor, Jami K., Donald P. Haider-Markel, and Benjamin Rogers. 2019. "Toward a New Measure of State-Level LGBT Interest Group Strength." *State Politics and Policy Quarterly* 19: 334–50.
- Warshaw, Christopher, and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74: 203–19.

Cite this article: Berry, William D., Richard C. Fording, Russell L. Hanson, and Justin K. Crofoot. 2023. The Validity of the Enns and Koch, and Berry et al. Measures of State Policy Mood: Continuing the Debate. *State Politics & Policy Quarterly* 23 (4): 347–358, <https://doi.org/10.1017/spq.2021.23>