



# Development of the Attitudes Towards Police Legitimacy Scale Short Form: A Rasch Analysis

Joshua J. Reynolds<sup>1</sup>

Accepted: 7 May 2023

© The Author(s), under exclusive licence to Society for Police and Criminal Psychology 2023

## Abstract

The objective of this research was to develop a shorter version of the Attitudes Towards Police Legitimacy Scale (APLS) using Rasch analysis. Two studies are conducted here to reduce the 34-item APLS to the 11-item APLS Short Form (APLS-SF). It was hypothesized that the APLS-SF would also predict right-wing authoritarianism, social dominance, support for police in a hypothetical donation task, and justification of an officer involved shooting. In study 1 ( $N=300$ ) and study 2 ( $N=600$ ), participants were sampled from Prolific Academic. The APLS-SF fit the Rasch model and predicted each variable as hypothesized.

**Keywords** APLS · Police legitimacy · Scale development · Short form

Legitimacy, particularly police legitimacy, has been an important concept in fields such as sociology, criminology, and psychology. Police legitimacy predicts compliance and cooperation with the law as well as motivating law-related behavior and community engagement (Papachristos et al. 2012; Paternoster et al. 1997; Tyler and Fagan 2008; Tyler and Jackson 2014). Tyler (1990) has been credited with bringing attention to the importance of legitimacy. In terms of how legitimacy is defined, there is some inconsistency, specifically concerning obligation to obey authority (see also Trinkner 2019). Originally, Tyler (1990) discussed obligation to obey authority as a *source* of legitimacy. Tyler (2009), however, says that “Legitimacy is a feeling of obligation to obey the law” (p. 313, emphasis added). On the other hand, Tyler (2006) does not use obligation in the definition of legitimacy but rather argues that legitimacy can *cause* an obligation to obey. Tyler and Jackson (2014) used a three-dimensional view of legitimacy: obligation (internalization that one should obey the police), trust and confidence (authorities are sincere, concerned about citizen welfare, and can be trusted to act in ways that benefit citizens), and normative alignment (legal authorities have values and goals aligned with their own). Research by Tyler

and Jackson (2014; see also Reisig et al. 2007) and Cao and colleagues (Cao et al. 1996; Cao and Wu 2019) have viewed trust and confidence as an important source of legitimacy.

As opposed to the empirical psychological approach focusing more on the legitimization process as presented by Tyler and colleagues, Tankebe and colleagues (Bottoms and Tankebe 2012; Tankebe 2013) use a more political science-oriented normative approach (see Trinkner 2019) and examine what legitimate power means. Their multidimensional view of legitimacy includes distributive fairness (people receive fair decisions), procedural fairness (quality of decision making and treatment), effectiveness (ability to obtain effective results), and lawfulness (power that has been acquired and exercised through established rules; see also Tankebe et al. 2016). Thus, to an extent they are examining different aspects of legitimacy.

Initiating some major debate in the literature, Sun and colleagues use the same data from China to model both the Tyler and colleagues’ model (Sun et al. 2017) and the Tankebe and colleagues’ model (Sun et al. 2018). Critiquing this view, Jackson and Branford (2019) argue that there is a conceptual distinction between the empirical legitimacy focus of the Tyler model and the normative legitimacy focus of the Tankebe model (see also Jackson et al. 2018; Trinkner 2019). Empirical legitimacy is how much individuals believe that police officers, for example, are legitimate. Empirical legitimacy thus refers to one’s subjective experience. On the other hand, normative legitimacy proscribes an objective set of criteria

✉ Joshua J. Reynolds  
Joshua.Reynolds@scranton.edu

<sup>1</sup> University of Scranton, Scranton, PA, USA

that can be used to determine, from an outside observer perspective, if some authority is legitimate. Jackson and Branford (2019) argue that using data to determine which of the models is correct is inappropriate, given that each model concerns a different conceptualization of legitimacy. This contention is further argued by Trinkner (2019).

Cao and Graham (2019), on the other hand, view some of the criticisms of the Sun et al. (2017, 2018) studies as overstated. Indeed, while the two models ostensibly concern different aspects of legitimacy, the Police Legitimacy Scale (PLS) developed by Tankebe et al. (2016) has been used to examine individual differences in people's police legitimacy, although their model is using the normative approach. For example, Ewanation et al. (2019) study the PLS as a measure of people's perception of police legitimacy in a Canadian sample. Furthermore, in Korva et al. (2022), perceptions of police use of force were examined, and the PLS was used to assess individual differences in police legitimacy.

Another legitimacy instrument, the Attitudes Towards Police Legitimacy Scale (APLS), was developed by Reynolds et al. (2018). In their scale development, they make reference to both the Tyler and Tankebe models and initially generated 73 items, some of which relate more to components like trust and confidence, discussed in the Tyler model, and some about fairness, as discussed in the Tankebe model. Reynolds et al. (2018) viewed legitimacy as an individual difference in approval of police authority, similar to Cao and Wu (2019), and explored components including bias, quality of interpersonal treatment, trustworthiness, motivation, quality and organizational integrity, being part of the community, and normative alignment. Thus, rather than start with a particular model of legitimacy, such as Tyler and Jackson (2014), they used an exploratory approach and built from the ground up. Specifically, Reynolds et al. (2018) focused on police actions and how those actions might relate to the perception of police legitimacy. In multiple studies, consistent evidence was found of a one-factor structure, and the scale, as hypothesized, predicted self-reported criminal behavior, right-wing authoritarianism, social dominance orientation, donating to a hypothetical police charity, and justification of an officer-involved shooting (Reynolds et al. 2018).

Some of the same criticisms that Jackson and Branford (2019) described of the Sun et al. (2017, 2018) research could be applicable to the APLS. For example, using items about trust and confidence (Tyler and Jackson 2014) and items about fairness (Tankebe 2013) is inappropriate, because the models are about different aspects of legitimacy (empirical versus normative). However, given the original goal of the APLS, these issues may not be a concern. The goal of the APLS was to assess people's perception of police legitimacy (meaning empirical rather than normative legitimacy). It is an empirical question as to what items assess legitimacy. Other researchers, for example, also view legitimacy as a public

perception of police activities, and therefore focus more on trust and confidence, as opposed to obligation and normative alignment (Cao et al. 1996; Cao and Wu 2019). Given that the Tyler model does not have unanimous support, and that what defines people's perception of police legitimacy is an empirical question, it is reasonable from a scale development perspective to have items that potentially tap into different aspects of legitimacy, even though the Tankebe model concerns normative legitimacy.

In three studies, Reynolds et al. (2018) consistently find a one-factor structure with items relating to areas such as fairness, trust and confidence, and normative alignment. The APLS has demonstrated strong reliability and factor structure. Subsequent research has made use of the APLS (April et al. 2022; Green and Evans 2021; Pomerantz et al. 2021). It may be particularly useful in studying the role of race and legitimacy, as it has items concerning fairness and trust and confidence, which have been argued to be key in understanding these effects (Cao and Wu 2019).

While the 34-item APLS has evidenced strong psychometric properties, one of the disadvantages is the length of the scale. If a scale is long, participants may not answer with the same attention and thoughtfulness on the last items as the first items or the response rate may be affected (Marcus et al. 2007). Furthermore, the time to complete the study is increased with longer scales. Some researchers have used a reduced number of the APLS items (Provenza 2021; Sturges et al. 2022). This suggests there is interest in the scale, but the length of the scale is prohibiting including all items. Removing items can be problematic without clear psychometric or theoretical justification, and often lowers scale reliability. Therefore, it would be beneficial to develop a shorter version of the APLS of high psychometric quality.

## Current Research

The purpose of this research is to develop a shorter version of the APLS and to psychometrically improve the APLS. One of the limitations of Reynolds et al. (2018) is that the scale was developed solely with classical test theory, with the data being analyzed via exploratory and confirmatory factor analysis (CFA). While factor analysis is useful for examining if a set of items assesses the same latent construct, limitations of the classical test theory approach include that estimates are sample dependent, more items are required, the data must be complete (i.e., no missing data), and there is no metric for difficulty/endorsibility of the items (Bailes and Nandakumar 2020; Bond et al. 2021). Rasch modeling, on the other hand, has none of these limitations and provides a particularly rich method to examine the strengths and weaknesses of an instrument. While the 34-item APLS evidenced high Cronbach's  $\alpha$  ( $> 0.98$ ), Cronbach's  $\alpha$  is an overestimate

of the consistency of a scale, and not a strong indication of a scale's reliability (Anselmi et al. 2019; Linacre 1997). Rasch analyses on the other hand, can examine if an instrument discriminates the sample into enough levels (person reliability) and if the sample is big enough to precisely locate the items on the latent trait (item reliability; Linacre 2022c).

One major difference with the Rasch model is that the items are not assumed to be of the same endorsibility, nor the step measures between categories equidistant. Different items are likely to have different endorsibility, and therefore by including items that capture different levels of the latent trait, the assessment instrument is improved. Rasch analyses allow both items and persons to be evaluated on a common metric (based in logits). Rasch analyses also excel at using a few well-chosen items to assess some latent trait, and therefore is highly advantageous in creating a short form (Bond et al. 2021). One of the most important differences concerns constructing linear estimates.

On a scale designed through classical test theory, like the 34-item APLS or the PLS, responses to the Likert items (1–5, 1–7, etc.) are typically summed or averaged. Let us imagine three individuals took the 34-item APLS and had the following summed scores: person A = 34, person B = 68, and person C = 136. Person C would be assumed to have twice the police legitimacy as person B, and person B would be assumed to have twice the police legitimacy as person A. However, this is unlikely to be the case as the scores cannot be assumed to be linear and the most that could be claimed is that person C ranks higher than B, and B ranks higher than A; to what degree they are different is not clear. This is partly because such scales are technically counts of qualitatively ordered observations (Linacre and Wright 1993).

On the other hand, it has been argued that if the data fit the Rasch model and quantity is present in the data, it is possible to construct linear estimates from these counts, and persons can be more meaningfully compared, although the subject of measures and the Rasch model is debated (Bond et al. 2021; Michell 2014). In the above example, if those scores were Rasch person scores, then it could be possible to make inferences about their relative levels of the latent trait, beyond their rank. This would be like having a meter stick with equally spaced intervals (Rasch person scores), as opposed to a meter stick with different-sized intervals (counts of qualitatively ordered observations; Boone et al. 2014). While this is a hotly debated topic, at the very least, if the data fit the Rasch model, person measures on the latent trait are more appropriate for parametric analyses, than the counts that would come from raw scores.

The specific Rasch model used in this research is the Rating Scale Model (RSM), designed for Likert-type items when the response categories are the same across all items. The equation  $\text{logit} = B_n - D_i - F_k$  expresses the log odds of a person ( $n$ ) choosing some category ( $k$ ) of item ( $i$ ), where  $B_n$

is a person's ability level,  $D_i$  is the agreeability/endorsibility level of the item, and  $F_k$  is the threshold level of category  $k$  of the item (Andrich 1978; Bond et al. 2021; Wright and Masters 1982). Given some of the advantages of the Rasch model, in particular that it excels in creating short forms, this research focuses on using Rasch analyses. Nonetheless, it is often informative to use both Rasch analyses and classical test theory approaches like factor analysis, so the results of the short form are reported for each type of analysis. This dual approach has been used previously in scale development and assessment research (Bailes and Nandakumar 2020; Hart et al. 2013; West et al. 2018; Yan 2020).

Ultimately, the goal of this research is to reduce the original 34-item APLS to 25–50% of its original items, or approximately 8–17 items. In the first analysis, study 3 data from Reynolds et al. (2018) is reanalyzed via Rasch analyses (“Initial Item Reduction”). The goal is to identify potentially redundant and misfitting items and other areas of possible scale improvement before new data is collected. Next, in study 1, 300 participants are sampled, and the reduced APLS items are examined via Rasch analysis and CFA to determine if items can be removed or added, or are problematic. Finally, in study 2, 600 participants are sampled, a Rasch analysis and CFA are conducted, and the relationship between the APLS Short Form (APLS-SF) right-wing authoritarianism, social dominance orientation, support for police in a hypothetical donation task, and justification of a police shooting are examined.

## Initial Item Reduction

Before collecting new data, it was informative to examine previous data on the APLS from the Rasch model perspective. Thus, a Rasch analysis was performed on study 3 data ( $N = 669$ ) from Reynolds et al. (2018). Item selection was guided by the importance of the content in assessing police legitimacy, having good fit to the Rasch model (as indicated by fit statistics,  $> 0.60$  and  $< 1.40$ ), and covering a range of the latent trait (as plotted in the Wright map for example). Data was analyzed using WINSTEPS 5.2.3.0 (Linacre 2022a). Wright maps, Infit and Outfit statistics, a dimensionality map, a pathway map, and others were examined, but for brevity are not shown here (see Supplementary Information for those tables and figures: <https://osf.io/jf2qn/>).

The Wright map indicated that the scale tended to capture individuals around average  $\pm 1$  logit; however, it did not perform well at capturing those higher or lower on trait police legitimacy and thus the scale could be improved in this respect. It also indicated that there were many redundant items. In examining the Infit and Outfit statistics, four items indicated some misfit, such as the item, “Police

officers are respected by the communities they serve” (Infit MNSQ = 1.70; Outfit MNSQ = 2.48; listed as C61 in the analysis), and were removed. As items were removed, the model was refit and the scale properties were re-examined.

In total, 21 items were removed, mostly due to having similar difficulty/endorsability and therefore being redundant. To capture those individuals lower than average and higher than average on trait police legitimacy, some items were retained but rewritten. For example, the item, “Police officers communicate well with people,” was modified to “Police officers are excellent at communicating with people.” Therefore, higher levels of police legitimacy would be required to endorse this item. In total, seven items were rewritten (items 7–13 in Table 1).

The Andrich thresholds and response category probability curves were examined to determine whether the original scale response range fit the Rasch model (1–7 scale). It appeared that category responses 2 (Andrich threshold = -1.79) and 3 (Andrich threshold = -1.27), and to a lesser extent 4 (Andrich threshold = -1.00) and 5 (Andrich threshold = 0.13), might not be well-defined and mutually exclusive. The difference in Andrich thresholds, for example, between response 2 and 3, was less than 1.2 (recommendations are > 1.2 and < 5.0 logits; Linacre 2022c). Thus, improvements could be made by collapsing categories.

## Study 1

Based on the initial item reduction analysis, the 34 items were reduced to 13. The scaling was also changed from the original Likert-type 1 = *strongly disagree* to 7 = *strongly*

*agree*, to a 1 = *strongly disagree* to 5 = *strongly agree*. A Rasch analysis and CFA on the 13-item APLS was used to evaluate the psychometric qualities of the scale. Item selection was based on the aforementioned Rasch characteristics.

## Method

### Participants

Participants were 300 individuals (46% males; 54% females) recruited through Prolific Academic (see Peer et al. 2017 for advantages of this research platform). The average age for participants was 36.19 ( $SD = 12.48$ ). Participants were White (non-Hispanic; 73%), Mixed (~9%), Black (~7%), Asian (~6%), Hispanic (~5%), and Pacific Islander, Native American, Middle Eastern, and other (less than 1%). The most common education background for participants was a bachelor’s degree (~40%), followed by some college (~21%) and a master’s degree (14%). The most common income level was \$30,001 to \$40,000 (~15%) and less than \$10,000 (~14%).

### Materials

#### Demographics

Participants were given a brief demographics form including questions about their age, sex, income, education level, and race/ethnicity.

**Table 1** Item statistics for study 1: 13-item APLS

Item	Standardized factor loading	Item difficulty/endorsability	SE	Infit MNSQ	Outfit MNSQ
1. Most police officers care about the communities they work in	0.893	-0.62	0.09	0.79	0.83
2. The presence of police makes me feel safe	0.888	0.00	0.09	1.04	1.14
3. If I have a problem, I feel confident that the police can help me solve it	0.885	0.54	0.09	0.93	0.89
4. People should trust the police to help	0.906	-0.17	0.09	0.78	0.78
5. I feel that police officers are willing to listen to me when I come into contact with them	0.869	-0.32	0.09	1.01	0.97
6. Most police officers define right and wrong the same way that I do	0.898	0.70	0.09	0.84	0.79
7. Police officers usually have a valid reason when they arrest people	0.872	-0.28	0.09	0.91	1.05
8. Police officers always treat people with respect	0.835	2.20	0.10	1.15	1.16
9. At least some police are trustworthy	0.818	-2.52	0.10	1.40	1.35
10. Police officers are excellent at communicating with people	0.868	1.50	0.10	0.93	0.91
11. Some police officers uphold values that are important to me	0.875	-1.59	0.10	1.08	1.04
12. The explanations that police officers give for a stop are always fair	0.790	1.71	0.10	1.30	1.47
13. If I call the police, they will try to help me	0.900	-1.16	0.10	0.80	0.79

### 13-Item APLS

Participants completed a 13-item version of the APLS using a Likert-type response from 1 = *strongly disagree* to 5 = *strongly agree*. The order of the items was randomized.

### Attention Check

Embedded in the APLS items was the item, “This is an attention check question; for this question select response 2.”

### Procedure

Participants signed up for the study and were redirected to an external survey delivered through Qualtrics. Participants completed the informed consent, answered the demographic questions, and completed the 13-item APLS and attention check item.

## Results

### Attention Check

Two participants (#178 and #265) failed the attention check item and were removed from subsequent analyses ( $N = 298$ ).

### Rasch Analysis

The Rasch model was fit in WINSTEPS 5.2.3.0 and converged with no issues (Linacre 2022a). For additional information on all analyses, see Supplementary Information.

### Reliability

The person separation index (284 non-extreme) was 4.21 and person reliability 0.95, indicating that the test discriminates around 4 levels of ability, an excellent level of separation (Linacre 2022c). The item separation index was 12.95 and item reliability 0.99, indicating excellent item separation.

### Dimensionality

The raw variance explained by measures was high (Eigenvalue = 35.21; 73%) and the unexplained variance in the first contrast (Eigenvalue = 1.86; 3.9%) and subsequent contrasts was low, indicating unidimensionality.

### Category Function

For each response category, the observed average was similar to the sample expectations. Infit MNSQ values were between 0.86 and 1.24 and Outfit MNSQ values were between 0.93

and 1.20 (see Table 2). In examining the Andrich thresholds, the differences in adjacent categories were all larger than 1.20 but smaller than 5 logits. As can be seen in Fig. 1 of the response category probability curves, the categories do appear to function well; for example, each response category is the most likely choice across some region of the person scores. This suggests that the 1–5 response categories were well-defined and mutually exclusive.

### Item-Person Map

Examining the item-person or Wright map (see Fig. 2), there is wide variability at the person level ( $\pm 7$  logits); however, the majority of person abilities fell within  $\pm 3$  logits. Item 8 was the most difficult item to endorse at 2.20 logits and item 9 was the easiest to endorse at  $-2.52$  logits. For the item statistics, see Table 1. The items captured a reasonable range of person abilities on the police legitimacy latent trait. Lastly, items 4, 5, and 7 appeared to capture a similar level of the person ability, and therefore indicates some redundancy among those items.

### Item Fit

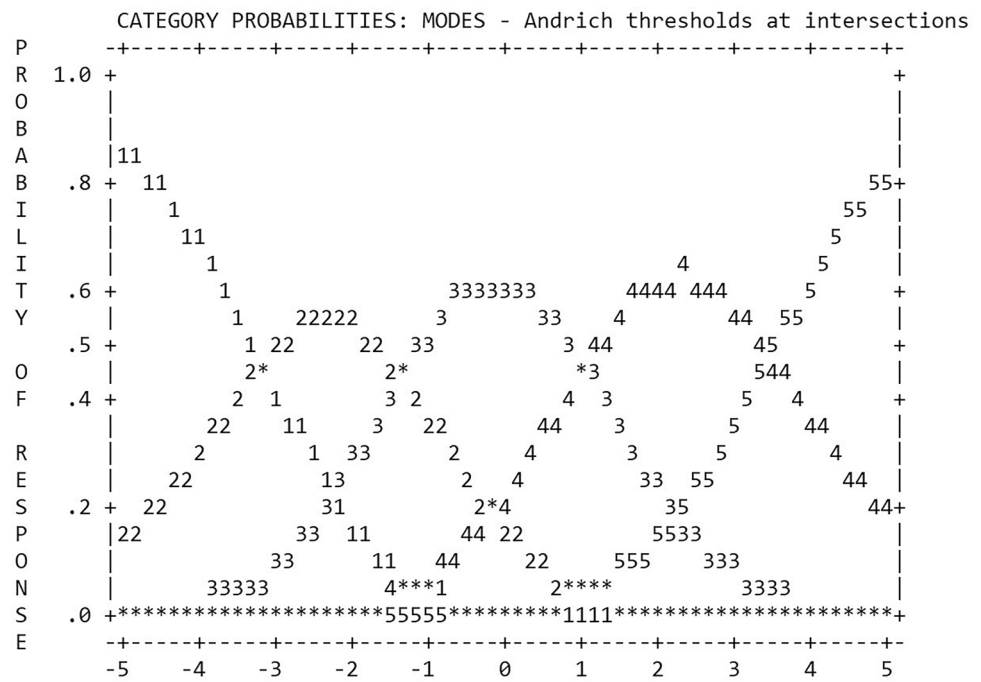
To examine how the items fit the Rasch model, Infit and Outfit statistics are available in Table 1. Recommended values for MNSQ fit statistics are between 0.6 and 1.40 (Linacre 2022c), with Outfit generally being the more important of the two. Item 12 and item 9 indicated some misfit. Odd responses by some individuals can impact the fit of items and following Boone et al. (2014); data was reanalyzed after accounting for these misfitting persons on those items. Specifically, a copy of the control file in WINSTEPS was created and the items for those persons indicating misfit ( $z$ -residual  $\geq 2$ ) were replaced by blanks. Then the model was refit and Infit and Outfit were reexamined. After accounting for the misfitting persons on those items, item 9 had an Infit MNSQ = 1.03 and Outfit MNSQ = 0.93 and item 12 had an Infit MNSQ = 0.96 and Outfit MNSQ = 0.88, indicating no misfit.

**Table 2** Category response statistics for study 1: 13-item APLS

Category label	Observed count	Infit MNSQ	Outfit MNSQ	Andrich threshold
1	634	1.05	1.05	None
2	749	0.96	1.00	-3.15
3	1155	0.95	0.97	-1.35
4	891	0.86	0.93	1.03
5	445	1.24	1.20	3.47



**Fig. 1** Response category probability curves study 1: 13-item APLS



**Confirmatory Factor Analysis**

A one-factor model with all 13 items was specified. Given that the response categories are ordered categories, a diagonally weighted least squares estimator was used. Model fit was partly based on fit indices including root mean square error of approximation (RMSEA; values less than 0.08 being adequate and less than 0.06 being close fit), standardized root mean square residual (SRMR; values less than 0.10 being adequate and less than 0.08 being close fit), and comparative fit index and Tucker–Lewis index (CFI, TLI; values greater than 0.90 being adequate and greater than 0.95 being close fit; Bentler 1990; Hu and Bentler 1999; Kline 2005). Analyses were conducted using the lavann package in R (R Core Team 2021; Rosseel 2012). The model of the hypothesized one-factor structure was significant,  $\chi^2(65, N=298)$  316.11,  $p < 0.001$ , and provided the following fit indices: RMSEA = 0.114, 90% CI [0.102, 0.127], SRMR = 0.038, CFI = 0.985, TLI = 0.982. All fit indices indicated close fit with the exception of RMSEA. The standardized factor loadings are in Table 1. In addition to the CFA, Cronbach’s  $\alpha$  was calculated at 0.965.

**Discussion**

Results of the Rasch modeling indicated that the 13-item APLS has strong psychometric properties but could be improved. The scale indicated excellent person and item

separation, the 1–5 response categories functioned well, there was strong evidence of unidimensionality, the items fit the Rasch model, and the items covered a range of person abilities. Two items (9 and 12) showed some misfit; however, when correcting for misfitting persons, these items were not misfit. Item 9, “At least some police are trustworthy,” was useful in that it captured those lower on police legitimacy. But the item might be somewhat too extreme in the wording which could have contributed to some of the misfit. Modifying Item 9 to, “Some police are trustworthy,” might improve fit while still capturing those lower on the latent trait. Items 4, 5, and 7 captured a similar level of the latent trait; and therefore, there is some redundancy in those items. Those items all fit the Rasch model well and had high factor loadings. Item 4 concerns trust and item 5 concerns communication, which are covered by other items. Item 7 is the only item that mentions arrests. Given the redundancy and to keep the scale as short as possible, items 4 and 5 could be removed.

In the CFA results, the analysis was remarkably similar to study 3 findings from Reynolds et al. (2018), with the exception of the RMSEA fit index. RMSEA was the only fit index here that did not show close fit. Unfortunately, even when fit indices are all in agreement about close model fit, this does not necessarily indicate that the model is reasonable (Lai and Green 2016). RMSEA can incorrectly indicate poor fit with simple models and few degrees of freedom, particularly if the sample size is lower, or even regardless of the degrees of freedom (Kenny et al. 2015). Additionally, based on a simulation study, Shi et al. (2019) found that caution should be given

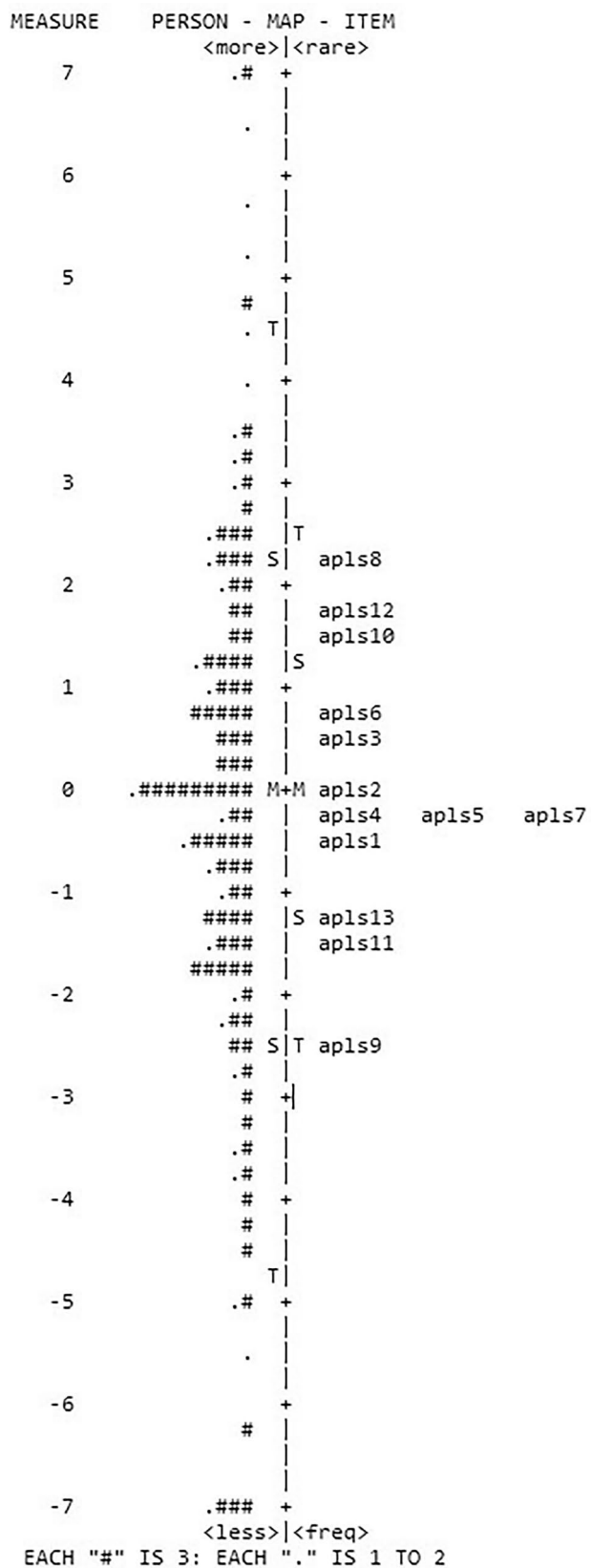


Fig. 2 Wright map for study 1: 13-item APLS

when interpreting high RMSEA values with smaller models (around 10 items) with high-quality indicators (Lambdas around 0.80), which is the situation in the present data. There were 13 items and standardized factor loadings ranged from 0.790 to 0.906. Based on the factor loadings, the majority of the fit indices in the CFA, and unidimensionality results from the Rasch model, a one-factor model appears reasonable. Lastly, the reliability was very high as assessed under the Rasch model and with Cronbach's  $\alpha$ .

### Study 2

Based on the results of study 1, two items were removed (items 4 and 5) and one item was modified (item 9). The 11 items are reordered and available in Table 3, referred to now as the APLS short form (APLS-SF). As in study 1, new data is collected and analyzed under the Rasch model and using a CFA. For validity purposes, several variables were examined that were included in Reynolds et al. (2018). Specifically, participants completed assessments of attitudes towards social dominance and attitudes concerning right-wing authoritarianism, which should both be positively related to police legitimacy. Police legitimacy should predict support for the police, so some participants were given a hypothetical donation task. In this task, there were five fictional charities, and participants allocated \$100 to one or more charities, with the prediction that higher police legitimacy scores would be associated with greater allocations to the police charity. Lastly, some participants watched a video of a real officer-involved shooting and were asked if the shooting was justified, with the prediction that higher legitimacy scores would predict greater odds of perceiving the shooting as justified. To keep the study as short as possible and given the large effect sizes observed in Reynolds et al. (2018) for the shooting task, participants were randomly assigned to complete either the hypothetical donation task or the shooting task, but otherwise completed the same materials.

### Method

#### Participants

Participants were 600 individuals (~43% males; ~57% females) recruited through Prolific Academic. Those participants who completed study 1 were excluded from participation in study 2. The average age for participants was 38.44 ( $SD=14.31$ ). Participants were White (non-Hispanic; 72%), Mixed (~9%), Black (~7%), Asian (~6%), Hispanic (~5%), and Native American, Middle Eastern, and other (less than 1%). The most common education background

**Table 3** Item statistics for study 2: APLS-SF

Item	Standardized factor loading	Item difficulty/ endorsibility	SE	Infit MNSQ	Outfit MNSQ
1. Most police officers care about the communities they work in	0.865	-0.83	0.06	0.85	0.83
2. The presence of police makes me feel safe	0.896	-0.11	0.06	0.89	0.86
3. If I have a problem, I feel confident that the police can help me solve it	0.894	0.42	0.06	0.86	0.85
4. Most police officers define right and wrong the same way that I do	0.892	0.59	0.06	0.86	0.81
5. Police officers usually have a valid reason when they arrest people	0.850	-0.30	0.06	0.89	0.91
6. Police officers always treat people with respect	0.848	2.05	0.07	1.15	1.05
7. Some police are trustworthy	0.817	-2.10	0.07	1.38	1.32
8. Police officers are excellent at communicating with people	0.877	1.27	0.06	.78	0.78
9. Some police officers uphold values that are important to me	0.841	-1.42	0.07	1.31	1.20
10. The explanations that police officers give for a stop are always fair	0.804	1.61	0.06	1.25	1.34
11. If I call the police, they will try to help me	0.895	-1.17	0.06	0.78	0.77

for participants was a bachelor's degree (~36%), followed by some college (~22%) and a high school diploma (16%). The most common yearly income level was less than \$10,000 (~21%) followed by \$20,001 to \$30,000 (~12%).

## Materials

### Demographics

Participants were given a brief demographics form including questions about their age, sex, income, education level, and race/ethnicity.

**APLS-SF** Participants completed the APLS-SF which consisted of 11 items answered on a Likert-type response from 1 = *strongly disagree* to 5 = *strongly agree*. The order of the items was randomized (Cronbach's  $\alpha = 0.95$ ).

### Social Dominance Orientation

The Social Dominance Orientation scale (SDO; Pratto et al. 1994) assesses the degree to which one believes the current social hierarchy is legitimate and just. The SDO scale consists of 16 items, ranging from 1 = *very negative* to 7 = *very positive* (some items are reverse scored). Items were averaged, with higher scores indicating higher beliefs in social hierarchies (Cronbach's  $\alpha = 0.95$ ). The order of the items was randomized.

### Right-Wing Authoritarianism

The Right-Wing Authoritarianism short scale version (RWA) assesses participants' willingness to obey authority they believe to be legitimate (Altemeyer 2004; Zakrisson 2005). The scale contains 15 items, ranging from 1 = *very negative* to 7 = *very positive* (some items are reverse scored).

Items were averaged, with higher scores indicating a greater willingness to obey authority (Cronbach's  $\alpha = 0.87$ ). The order of the items was randomized.

### Hypothetical Donation Task

In this task, participants were asked to imagine they had \$100 to donate to five fictional charities: The Law Enforcement Support Fund, End Hunger Now, The International Association for the Treatment of Communicable Diseases, Americans Against Child Abuse, and South American Outreach Ministry (Reynolds et al. 2018). A description of each charity was given with each being similar in length (58 to 59 words). The order of the charity descriptions and the order of the answers in the donation question were randomized.

### Officer-Involved Shooting Justification Task

In this task, a video from a real fatal police shooting (occurring in 2014) was shown. In the video, an officer enters the home of a hit-and-run suspect, who is holding a knife. The officer attempts to get the suspect to drop the knife, but disobeying the officers' orders, walks towards the officer, who then discharges his weapon four times. This video has been used previously to examine justification of an officer-involved shooting (Culhane et al. 2016; Reynolds et al. 2018). Participants were asked whether the officer was justified in shooting the suspect (0 = no, 1 = yes).

### Attention Check

There were three attention check items. The first item was embedded in the APLS-SF, "This is an attention check question; for this question select response two." The second was for the hypothetical donation task, "Below are several causes



that people would be interested in donating to. Which of the ones listed below were actual options you were given to donate to in the donation task (you may choose more than one)?” There were seven options, and the correct answer on this item meant choosing both “police support” and “child abuse.” The third was for the officer-involved shooting task, “What was Eric Johnson (NOT the police officer) holding?” Four options were given, and the correct answer was “a knife.”

## Procedure

Participants signed up for the study and were redirected to an external survey delivered through Qualtrics. Participants completed the informed consent, answered the demographic questions, completed the APLS-SF, SDO, and RWA ( $N=600$ ). Participants were then randomly assigned to either the hypothetical donation task ( $N=281$ ) or the officer-involved shooting task ( $N=319$ ).

## Results

### Attention Check

One participant failed the first attention check item embedded in the APLS-SF. Subsequent analyses removed this participant. Fifty-six participants failed the hypothetical donation task attention check and were removed from those analyses ( $N=224$ ). Twelve participants failed the officer-involved shooting task attention check and were removed from those analyses ( $N=307$ ).

### Rasch Analysis

The Rasch model was fit in WINSTEPS 5.2.4.0 (Linacre 2022b). For additional information on all analyses, see Supplementary Information.

### Reliability

The person separation index (588 non-extreme) was 3.97 and person reliability 0.94; and the item separation index was 18.61 and item reliability 1.00, all indicating excellent separation (Linacre 2022c).

### Dimensionality

The raw variance explained by measures was high (Eigenvalue = 31.28; 74%) and the unexplained variance in the first contrast (Eigenvalue = 2.09; 4.9%) and subsequent contrasts was low. While the Eigenvalue for the first contrast

was slightly higher than 2, the percentage variance was less than 5, and an Eigenvalue below even 3 is still acceptable for shorter instruments (< 20 items; Linacre 2022c). All other contrasts had Eigenvalues less than 1.50. These results indicate the APLS-SF is unidimensional.

### Category Function

For each response category, the observed average was similar to the sample expectations. Infit MNSQ values were between 0.92 and 1.1 and Outfit MNSQ values were between 0.87 and 1.10 (see Table 4). In the Andrich thresholds, the differences in adjacent categories were all larger than 1.20 but smaller than 5 logits. In the plot of the response category probability curves, the categories appear to function well (see Fig. 3).

### Item-Person Map

Examining the Wright map (see Fig. 4), similar to study 1, the majority of person abilities fell within  $\pm 3$  logits. Item 6, “Police officers always treat people with respect,” was the most difficult item to endorse at 2.05 logits and item 7, “Some police are trustworthy,” was the easiest to endorse at  $-2.10$  logits. For the item statistics, see Table 1. The APLS-SF captured a reasonable range of person abilities on the police legitimacy latent trait. While there do appear to be small gaps in assessing the person scores, there was no longer any redundancy among the items.

### Item Fit

Infit and Outfit statistics are available in Table 1. All items had Infit and Outfit MNSQ values between 0.77 and 1.38, indicating acceptable item fit.

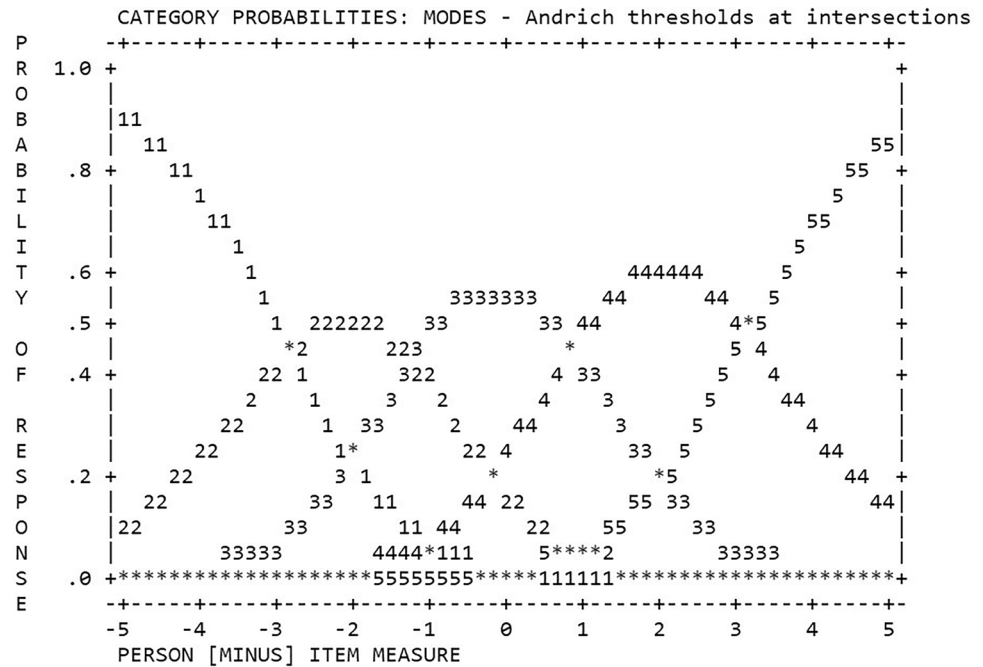
### Confirmatory Factor Analysis

A one-factor model with all 11 items was specified with a diagonally weighted least squares estimator. The model of the hypothesized one-factor structure was significant,  $\chi^2(44, N=599) 620.93, p < 0.001$ , and provided the

**Table 4** Category response statistics for study 2: APLS-SF

Category label	Observed count	Infit MNSQ	Outfit MNSQ	Andrich threshold
1	1036	1.10	1.10	None
2	1258	0.96	.97	-2.80
3	1802	0.94	0.94	-1.23
4	1584	0.92	0.87	0.86
5	909	1.10	1.07	3.17

**Fig. 3** Response category probability curves for study 2: APLS-SF



following fit indices: RMSEA = 0.148, 90% CI [0.138, 0.159], SRMR=0.053, CFI=0.978, TLI=0.973. As in study 1, all fit indices indicated very close fit except RMSEA. The standardized factor loadings which range from 0.804 to 0.895 are in Table 1.

**Construct Validity**

In the next set of analyses, the APLS-SF was used to predict relevant variables. Rasch person measures are naturally in logits and have been rescaled here so that the lowest reportable person measure score was 0 and the highest score 100. As with the raw scores, higher person scores indicate greater police legitimacy. Parallel analyses using the raw score averages on the APLS-SF are provided in the Supplementary Information (R-markdown file), with the results being the same.

**Police Legitimacy, Social Dominance, and Right-Wing Authoritarianism**

Higher APLS-SF person measure scores were related to higher scores on both RWA  $r(597)=0.52$ , 95% CI [0.46, 0.58],  $p < 0.001$ ; and SDO  $r(597)=0.35$ , 95% CI [0.28, 0.42],  $p < 0.001$ .

**Police Legitimacy and Donating to a Hypothetical Police Charity**

The APLS-SF person measure scores positively predicted the amount of money donated to the police charity  $\beta=0.65$ , 95% CI [0.52, 0.78],  $t(222)=9.8$ ,  $p < 0.001$ ,  $sr^2=0.302$ .

**Controlling for Age, Sex, Race, and Income**

The effect of police legitimacy on police donations could be biased because of several demographic factors, such as race. However, controlling for age, sex, race, and income, the APLS-SF person measure scores nonetheless still predicted greater police donations,  $\beta=0.60$ , 95% CI [0.46, 0.75],  $t(204)=7.99$ ,  $p < 0.001$ ,  $sr^2=0.20$  (see directed acyclic graph 1 in the R-markdown in Supplementary Information for the DAG and adjustment sets).

**Police Legitimacy and Justification of an Officer-Involved Shooting**

Using logistic regression, the odds of perceiving the police shooting as justified increased 1.10 times per increase in APLS-SF person measure scores ( $\beta=0.09$ ,  $SE=0.012$ ,

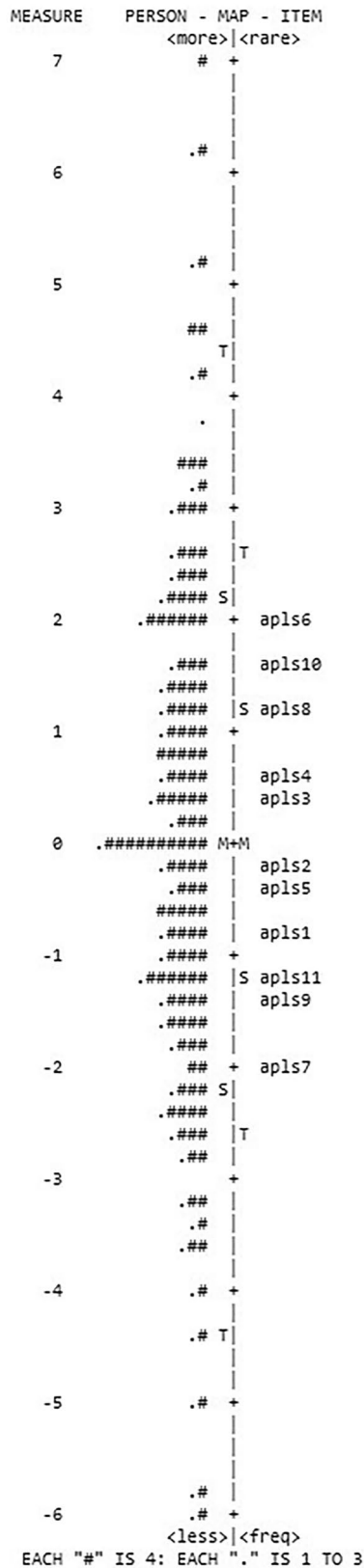


Fig. 4 Wright map for study 2: APLS-SF

$p < 0.001$ , 95% CI for  $OR$  [1.07, 1.12]). The odds of viewing it as justified when APLS person scores were 0 was 0.005 95% CI [0.001, 0.017].

### Controlling for Age, Sex, Race, and Income

As in the police donation analyses, the effect of police legitimacy on justification of a shooting could be biased because of several demographic factors. Controlling for age, sex, race, and income, the APLS-SF person measure scores predicted greater odds of viewing the shooting as justified ( $\beta = 0.102$ ,  $SE = 0.014$ ,  $p < 0.001$ , 95% CI for  $OR = 1.11$  [1.08, 1.14]; see directed acyclic graph 2 in the R-markdown in Supplementary Information for the DAG and adjustment sets).

### Discussion

The results of the Rasch modeling in study 2 indicated that the APLS-SF had excellent person and item separation, the 1–5 response categories functioned well, there was evidence of unidimensionality, the items fit the Rasch model, and the items covered a range of person abilities. In the CFA results, all items had high factor loadings and all fit indices, except RMSEA, indicated close fit. To examine construct validity of the APLS-SF and make comparisons to the 34-item APLS, several other hypothesized relationships were examined. The APLS-SF was correlated with social dominance and right-wing authoritarianism. The APLS-SF predicted donating more money to the hypothetical police charity and predicted greater odds of viewing the officer-involved shooting as justified. After controlling for age, sex, race, and income, the effects of APLS-SF on police donations and justification of a shooting remained similar.

### General Discussion

The purpose of this research was to use Rasch analysis to develop a shorter version of the 34-item APLS (Reynolds et al. 2018). An initial Rasch analysis of previous data indicated that there were many redundant items, some misfit items, and a potential issue with the 1–7 response categories, from the Rasch perspective. Based on this, the scale was reduced to 13 items and 300 participants were collected in study 1. The results of study 1 were promising but indicated several redundant items. In study 2, 600 participants were collected and the 11-item APLS-SF functioned

very well. The items fit the Rasch model, and the reliability indices indicated excellent reliability. Therefore, the final items of the APLS-SF are available in Table 3 (the paper version of the APLS-SF is available in Supplementary Information: <https://osf.io/jf2qn/>). Additionally, all the relationships including right-wing authoritarianism, social dominance, police support, and justification of an officer-involved shooting were comparable to results based on the 34-item APLS in Reynolds et al. (2018).

### Limitations and Future Directions

The APLS-SF was only able to capture individuals  $\pm 2$  logits on the latent trait. While this is useful for most populations, there is clearly much higher and much lower abilities on this latent trait. Therefore, when assessing those at the extreme ends, caution should be used. An area for future scale improvement could be adding several items that capture those on the extremes of the police legitimacy latent trait. Relatedly, as can be seen in the Wright map of Fig. 4, there were several small gaps in the items. In future research, several items could be added fill those gaps. However, on the whole, the scale performed well, with no redundancy among the items and capturing a range of person abilities.

In examining people's perception of police legitimacy, what defines their legitimacy may differ based on the context. While the APLS-SF is arguably an improvement over the APLS, the APLS-SF cannot necessarily be used in all contexts. In some countries, for example, items that concern trust may not be of key importance to those people's perception of legitimacy. Future research in international samples would be useful in this regard.

While this research was primarily concerned with Rasch analysis, CFAs were also conducted. One area of potential concern is the high RMSEA values, indicating poor fit. However, the factor loadings were consistently high, the other fit indices suggested close fit, and conditions were present (around 10 items of high quality) to produce a misleadingly high RMSEA value. Inconsistency in fit indices is not particularly diagnostic of model issues and generally the literature warns of overinterpreting fit indices (Lai and Green 2016). The Rasch unidimensionality results and previous research on the APLS have consistently shown a one-factor solution; however, this issue deserves more attention in future research.

Another important area for future research is examining if legitimacy has quantitative structure (i.e., it has order and additivity). Demonstrating that a variable has quantitative structure is an extensive process. For example, it took centuries to understand the properties of, and develop measures for, temperature (Chang 2004; Michell 1990; Sherry 2011). It appears that legitimacy has order, but not necessarily additivity. This makes it more difficult to meaningfully compare scores. Examining the APLS-SF and other

legitimacy scales through additive conjoint measurement would be a useful future direction in this regard (Luce et al. 1990; Luce and Tukey 1964; Michell 1990). Examination of measurement invariance is also necessary and should be pursued in future research.

Attention checks were used throughout this research. Failing the attention checks was rare, except in the case of the hypothetical donation task. The donation task attention check question was substantially harder, and thus not unexpected that this item would have high failure. Given the number that failed, this was of some concern. Results of the hypothetical donation task were reanalyzed, this time including all who completed the task, and the results were the same. Therefore, although a higher number of participants failed the attention check, this did not appear to bias those results.

### Conclusion

Defining and measuring police legitimacy is complicated, contested, and elusive (Trinkner 2019). The APLS was developed to assess people's perception of police legitimacy (i.e., empirical legitimacy). The evidence indicates that the APLS does tap into legitimacy rather than just general attitudes towards the police (see Reynolds et al. 2018); however, it does not necessarily capture all aspects of legitimacy and further theory development and tests are important here. This view of legitimacy is in line with Cao et al. (1996) and Cao and Wu (2019), where the focus is on the perception of police activities and having confidence that the police can be trusted and are fair, for example. Theory is essential to measurement and understanding legitimacy requires much more precise theory and understanding of the causal relationships. The goal in this research was a pragmatic one, to reduce the lengthy 34-item APLS to the much shorter 11-item APLS-SF, while simultaneously improving the psychometric qualities of the scale. Given the brevity of the scale and its other strengths, it may be of particular use in examining police legitimacy over time or examining differences between groups, as discussed in Cao and Wu (2019). Future research will undoubtedly improve the scale and our understanding of police legitimacy.

**Funding** This work was supported by internal funding.

**Data Availability** The data that support the findings of this study are openly available on: <https://osf.io/jf2qn/>.

### Declarations

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and/or national research committee and with

the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Conflict of Interest** The author declares no competing interests.

## References

- Altemeyer B (2004) The other “authoritarian personality.” In: Jost JT, Sidanius J (eds) *Political psychology: key readings*. Psychology Press, pp 85–107
- Andrich D (1978) A rating formulation for ordered response categories. *Psychometrika* 43(4):561–573. <https://doi.org/10.1007/BF02293814>
- Anselmi P, Colledani D, Robusto E (2019) A comparison of classical and modern measures of internal consistency. *Front Psych* 10:2714. <https://doi.org/10.3389/fpsyg.2019.02714>
- April K, Cole LM, Goldstein NES (2022) Let’s “talk” about the police: the role of race and police legitimacy attitudes in the legal socialization of youth. *Curr Psych: A Journal for Diverse Perspectives on Diverse Psychological Issues*. Advance online publication. <https://doi.org/10.1007/s12144-022-02709-8>
- Bailes LP, Nandakumar R (2020) Get the most from your survey: an application of Rasch analysis for education leaders. *Int J Leadership* 16(2). <https://doi.org/10.22230/ijelpl.2020v16n2a857>
- Bentler PM (1990) Comparative fit indexes in structural models. *Psychol Bull* 107(2):238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bond TG, Yan Z, Heene M (2021) *Applying the Rasch model. Fundamental measurement in the human sciences* (4<sup>th</sup> ed.). Routledge
- Boone WJ, Staver JR, Yale MS (2014) *Rasch analysis in the human sciences*. Springer
- Bottoms A, Tankebe J (2012) Beyond procedural justice: a dialogic approach to legitimacy in criminal justice. *J Crim Law Criminol* 102:119–170. <http://www.jstor.org/stable/23145787>
- Cao L, Frank J, Cullen FT (1996) Race, community context, and confidence in the police. *Am J Police* 15(1):3–22. <https://doi.org/10.1108/07358549610116536>
- Cao L, Graham A (2019) The measurement of legitimacy: a rush to judgment? *Asian J Criminol* 14:291–299. <https://doi.org/10.1007/s11417-019-09297-w>
- Cao L, Wu Y (2019) Confidence in the police by race: taking stock and charting new directions. *Police Pract Res* 20(1):3–11. <https://doi.org/10.1080/15614263.2017.1396460>
- Chang H (2004) *Inventing temperature*. Oxford University
- Culhane SE, Bohman JH, Schweitzer K (2016) Public perceptions of the justifiability of police shootings: the role of body cameras in a pre- and post- Ferguson experiment. *Police Q* 19(3):251–274. <https://doi.org/10.1177/10986111166651403>
- Ewanation L, Bennell C, Blaskovits B, Baldwin S (2019) Validating the police legitimacy scale with a Canadian sample. *Can J Criminol Crim Justice* 61(4):1–23. <https://doi.org/10.3138/cjccj.2018-0036>
- Green DA, Evans AM (2021) Undue police violence toward African Americans: an analysis of professional counselors’ training and perceptions. *J Couns Dev* 99(4):363–371. <https://doi.org/10.1002/jcad.12389>
- Hart CO, Mueller CE, Royal KD, Jones MH (2013) Achievement goal validation among African American high school students: CFA and Rasch results. *J Psychoeduc Assess* 31(3):284–299. <https://doi.org/10.1177/0734282912466726>
- Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
- Jackson J, Bradford B (2019) Blurring the distinction between empirical and normative legitimacy? A methodological commentary on ‘Police legitimacy and citizen cooperation in China.’ *Asian J Criminol* 14(4):265–289. <https://doi.org/10.1007/s11417-019-09289-w>
- Jackson J, Milani J, Bradford B (2018) Empirical legitimacy and normative compliance with the law. In A. Farazmand (Ed.), *Global encyclopedia of public administration, public policy, and governance*. Springer International Publishing. [https://doi.org/10.1007/978-3-319-31816-5\\_1914-1](https://doi.org/10.1007/978-3-319-31816-5_1914-1)
- Kenny DA, Kaniskan B, McCoach DB (2015) The performance of RMSEA in models with small degrees of freedom. *Sociol Method Res* 44(3):486–507. <https://doi.org/10.1177/0049124114543236>
- Kline T (2005) *Psychological testing: A practical approach to design and evaluation*. Sage
- Korva N, Bennell C, Lalumière M, Karimullah M (2022) What do you see? Understanding perceptions of police use of force videos as a function of the camera perspective. *Behav Sci Law* 40(3). <https://doi.org/10.1002/bsl.2578>
- Lai K, Green SB (2016) The problem with having two watches: assessment of fit when RMSEA and CFI disagree. *Multivar Behav Res* 51(2/3):220–239. <https://doi.org/10.1080/00273171.2015.1134306>
- Linacre JM (1997) KR-20 or Rasch reliability: which tells the “truth”? *Rasch Measurement Transactions* 11(3):580–581
- Linacre JM (2022a) *Winsteps. Rasch measurement computer program (Version 5.2.3)*. Portland, Oregon: [Winsteps.com](http://Winsteps.com)
- Linacre JM (2022b) *Winsteps. Rasch measurement computer program (Version 5.2.4)*. Portland, Oregon: [Winsteps.com](http://Winsteps.com)
- Linacre JM (2022c) *Winsteps. Rasch measurement computer program User’s Guide. Version 5.2.3*. Portland, Oregon: [Winsteps.com](http://Winsteps.com)
- Linacre M, Wright B (1993) Constructing linear measures from counts of qualitative observations. Paper presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics, Berlin, Germany
- Luce RD, Krantz DH, Suppes P, Tversky A (1990) *Foundations of measurement, Vol 3: representation, axiomatization, and invariance*. Academic Press
- Luce RD, Tukey JW (1964) Simultaneous conjoint measurement: A new type fundamental measurement. *J Math Psychol* 1(1):1–27. [https://doi.org/10.1016/0022-2496\(64\)90015-X](https://doi.org/10.1016/0022-2496(64)90015-X)
- Marcus B, Bosnjak M, Lindner S, Pilischenko S, Schütz A (2007) Compensating for low topic interest and long surveys: a field experiment on nonresponse in web surveys. *Soc Sci Comput Rev* 25(3):372–383. <https://doi.org/10.1177/0894439307297606>
- Michell J (1990) *An introduction to the logic of psychological measurement*. Routledge
- Michell J (2014) The Rasch paradox, conjoint measurement, and psychometrics: response to Humphry and Sijtsma. *Theory Psychol* 24(1):111–123. <https://doi.org/10.1177/0959354313517524>
- Papachristos AV, Meares TL, Fagan J (2012) Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders. *J Crim Law Criminol* 102(2):397–440. Retrieved from [https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=5667&context=fss\\_papers](https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=5667&context=fss_papers)
- Patronoster R, Brame R, Bachman R, Sherman LW (1997) Do fair procedures matter? The effect of procedural justice on spouse assault. *Law Soc Rev* 31(1), 163–204. Retrieved from <https://www.jstor.org/stable/3054098>
- Peer E, Brandimarte L, Samat S, Acquisti A (2017) Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 70:153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>



- Pomerantz AL, Bell K, Green K, Foster S, Carvallo M, Schow P (2021) "Badge of honor": honor ideology, police legitimacy, and perceptions of police violence. *J Police Crim* 36(3):473–489 (2021). <https://doi.org/10.1007/s11896-021-09433-2>
- Pratto F, Sidanius J, Stallworth LM, Malle BF (1994) Social dominance orientation: a personality variable predicting social and political attitudes. *J Pers Soc Psychol* 67(4):741–763. <https://doi.org/10.1037/0022-3514.67.4.741>
- Provenza K (2021) Gatekeepers of accountability: Grand juror decision making in police use of force cases at the intersection of race and mental illness [Doctoral dissertation, University of Wyoming]. Pro Quest Dissertations and Theses Global
- R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reisig MD, Bratton J, Gertz MG (2007) The construct validity and refinement of process-based policing measures. *Crim Justice Behav* 34(8):1005–1028. <https://doi.org/10.1177/0093854807301275>
- Reynolds JJ, Estrada-Reynolds V, Nunez N (2018) Development and validation of the Attitudes Towards Police Legitimacy Scale. *Law Hum Behav* 42(2):119–134. <https://doi.org/10.1037/lhb0000281>
- Rosseel Y (2012) Lavaan: an R package for structural equation modeling. *J Stat Softw* 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sherry D (2011) Thermoscopes, thermometers, and the foundations of measurement. *Stud Hist Philos Sci A* 42(4):509–524. <https://doi.org/10.1016/j.shpsa.2011.07.001>
- Shi D, Lee T, Maydeu-Olivares A (2019) Understanding the model size effect on SEM fit indices. *Educ Psychol Measur* 79(2):310–334. <https://doi.org/10.1177/0013164418783530>
- Sturges HA, Flick C, Provenza K, Nunez N (2022) Civil juror compensation and judgments of police use of force at the intersection of race and mental illness. *Behav Sci Law* 40(4). <https://doi.org/10.1002/bsl.2561>
- Sun IY, Li L, Wu Y, Hu R (2018) Police legitimacy and citizen cooperation in China: Testing an alternative model. *Asian J Crim* 13:275–291. <https://doi.org/10.1007/s11417-018-9270-4>
- Sun IY, Wu Y, Hu R, Farmer AK (2017) Procedural justice, legitimacy, and public cooperation with police: does Western wisdom hold in China? *J Res Crime Delinq* 54(4):454–478. <https://doi.org/10.1177/002242781663870>
- Tankebe J (2013) Viewing things differently: The dimensions of public perceptions of police legitimacy. *Criminology* 51(1):103–135. <https://doi.org/10.1111/j.1745-9125.2012.00291.x>
- Tankebe J, Reisig MD, Wang X (2016) A multidimensional model of police legitimacy: a cross-cultural assessment. *Law Hum Behav* 40(1):11–22. <https://doi.org/10.1037/lhb0000153>
- Trinkner R (2019) Clarifying the contours of the police legitimacy measurement debate: a response to Cao and Graham. *Asian J Crim* 14(4):309–335. <https://doi.org/10.1007/s11417-019-09300-4>
- Tyler TR (1990) Why people obey the law. Yale University Press
- Tyler TR (2006) Psychological perspectives on legitimacy and legitimation. *Annu Rev Psychol* 57:375–400. <https://doi.org/10.1146/annurev.psych.57.102904.190038>
- Tyler TR (2009) Legitimacy and criminal justice: the benefits of self-regulation. *Ohio State J Crim Law* 7:307–359
- Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities. *Ohio State J Crim Law* 6:231–275. Retrieved from [https://kb.osu.edu/dspace/bitstream/handle/1811/73064/OSJCL\\_V6N1\\_231.pdf](https://kb.osu.edu/dspace/bitstream/handle/1811/73064/OSJCL_V6N1_231.pdf)
- Tyler TR, Jackson J (2014) Popular legitimacy and the exercise of legal authority: motivating compliance, cooperation, and engagement. *Psychol Public Policy Law* 20(1):78–95. <https://doi.org/10.1037/a0034514>
- West C, Baker A, Ehrich JF, Woodcock S, Bokosmaty S, Howard SJ et al (2018) Teacher disposition scale (TDS): construction and psychometric validation. *J Furth High Educ* 44(2):185–200. <https://doi.org/10.1080/0309877X.2018.1527022>
- Wright BD, Masters GN (1982) Rating scale analysis. MESA press
- Yan Z (2020) Developing a short form of the self-assessment practices scale: psychometric evidence. *Front Edu* 4:153. <https://doi.org/10.3389/educ.2019.00153>
- Zakrisson I (2005) Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality Individ Differ* 39(5):863–872. <https://doi.org/10.1016/j.paid.2005.02.026>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.