Evaluating the Due Process and Crime Control Perspectives Using Rasch Measurement Analysis

Vanessa A. Edkins
Florida Institute of Technology

Kenneth D. Royal
University of Kentucky

Background: The biases jurors possess may influence everything from the interpretation of case evidence to impressions of the defendant to, ultimately, verdict and recognition of this has led to a number of juror attitude scales attempting to tap into important biases. A common ideology discussed in legal research is that individuals attitudes toward the law and the legal system differ along a continuum moving from due process (a concern for the preservation of individual rights) to crime control (a focus on swift and harsh punishment for those who break the law) although an agreed upon assessment of these perspectives has yet to be created.

Purpose: The current research addresses due process and crime control perspectives and uses the ideology as a source for a new measure of juror bias: The General Attitudes toward the Legal System (GALS) scale.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: The GALS scale was constructed based on existing theory and administered to nearly 700 undergraduate psychology students at a large Midwestern university. The psychometric properties of the instrument were then evaluated to determine instrument quality.

Data Collection and Analysis: The Rasch Rating Scale Model (RRSM) was used to evaluate the psychometric properties of the GALS. Evaluation focused on six characteristics of the instrument: dimensionality, reliability, rating scale quality, item quality, item hierarchy, and measure quality.

Findings: Results indicate the GALS is a psychometrically sound instrument for measuring juror bias.

Keywords: *juror decision making; juror bias; scale development; instrument validation; psychometric evaluation*

Two individuals sit side by side, listen to the same information, and apply the same instructions to their task. Afterwards, they meet to discuss and realize that not only have they reached different conclusions, they have very different interpretations of the information. An American jury is composed of up to twelve of these individuals. The same reason that the jury system works in the United States is also why it may fail, and why it generates so

Vanessa A. Edkins and Kenneth D. Royal

much research. Asking jurors to hear a case is interjecting human nature into the legal process.

The possible effects of bias in the courtroom often drive suggestions to reform or limit the use of juries (Hans & Appel, 1999). Research has shown that personality characteristics and attitudes can affect how a juror perceives trial information and decides a verdict (Casper, Benedict & Perry, 1989; Fitzgerald & Ellsworth, 1984; Hastie, Penrod & Pennington, 1983). For example, jurors who are pro-prosecution, and unable to set aside this attitude, may have a decision of guilt that begins at a level closer to conviction than to acquittal. Presumably, less evidence would be required to secure a guilty verdict in such persons than if jurors are pro-defense or if they apply no weight to their initial biases (Ostrom, Saks & Werner, 1978).

In cases where attitudes may exert an influence, the best predictor of juror bias is attitudes toward specific, key elements of the case (Lieberman & Sales, 2007). For example, when a lawyer has been accused of a drug-related offense good predictors of guilty verdicts are attitudes toward drugs and attitudes toward lawyers (Moran, Cutler, & Loftus, 1990). In the same respect, attitudes toward rape is a good predictor of the verdict in a rape case (Field, 1978) and endorsement of myths about battered women is a good predictor of the verdict when such women are accused of murdering their batterers (Vidmar & Schuller, 1989). This is in line with Ajzen and Fishbein's (1977) review of research on attitudes and behavior showing that if researchers want to use attitudes to predict behavior, their best bet is to focus on specific attitudes toward that behavior.

Information integration theories suggest that one of the initial components affecting juror decision-making is an a priori bias relevant to beliefs about the legal system (Kaplan & Kemmerick, 1974; Ostrom et al., 1978). In testing this theory, Kaplan and Miller (1978) pre-identified respondents as being supportive of either harsh or lenient treatment toward law-breakers before exposing them to trial information. This "general bias" of jurors played an important role in decisions of guilt, such that those pre-identified as harsh were more likely to convict and were more punitive than were those identified as lenient. While on the surface this may seem to violate Ajzen and Fishbein's (1977) contention that general attitudes are poor predictors of specific behaviors, Kaplan and Miller (1978) actually showed that it was the attitude's effect on a larger range of behaviors that, cumulatively, were predicting verdicts.

This "general bias" concept is important because many of the differences in attitudes toward the legal system have underlying similarities; specifically, general attitudes about the state and purpose of the criminal legal system. The goal of the present paper is to show that those different viewpoints regarding the "is" and "ought" (Packer, 1964) of the criminal legal system can be defined and measured. We do so by capitalizing on a theoretical distinction between two conceptions of the goal of the legal system, first advanced over 40 years ago by law professor Herbert Packer, and termed the due process model and the crime control model (Packer, 1964).

Due Process and Crime Control Attitudes

Packer's (1964; 1968) distinction between a due process and a crime control perspective suggests a template that

society uses to assess the justice system. He proposed that the state of legal affairs in the United States, and the ideological differences between its citizens, may be summarized as falling somewhere in between two extremes. At one end of the spectrum is crime control, and with it the idea of the swift hand of justice, and at the other end is due process and a vigilant attempt to preserve individual rights.

For crime control, what is important is the finale and accordingly sanctioning those who commit crimes. The crime control model has a low tolerance for the adjudicative process. If a case has been thoroughly investigated then it is safe to assume the suspect is likely guilty. This "presumption of guilt" (Packer, 1964) is what will allow the system to move in a more expeditious manner.

According to this perspective, a perfect legal system would allow criminal investigations as much freedom as possible, demand efficiency and thoroughness in these investigations, advocate swift punishment if the investigation has deemed there is enough evidence to assume the individual in question is probably guilty (or expeditious exoneration if it is determined the defendant could not possibly be guilty), and employ adequate deterrents of future crime (generally in the form of harsh punishments).

On the other hand, the old adage that, "It is better for ten guilty men to go free than for one innocent man to suffer an injustice," is the foundation of the due process model. The tolerability for error is low. If there are any qualms about the guilt of an individual then that person should go free. In the crime control model, allowing investigators to use all necessary power to determine factual guilt is the goal of the legal system. The due process model makes a distinction between those who are factually guilty and those who are legally guilty (Packer, 1964). Whether or not the individual actually committed the crime is secondary. What is important is whether, through legal means and by preserving individual rights, a valid adjudicative process finds that individual guilty.

Due process does not condone crime nor does endorsement of this perspective equal a view that crime is not a problem in society (Packer, 1964). Instead, recognition of potential problems in criminal investigations (e.g., faulty eyewitness identifications and coerced confessions) and in legal proceedings (e.g., biased witnesses and inadequate counsel) suggests that we should never assume a case closed. Still, Packer suggests that the due process orientation, as he proposes it, is not void of affect, strictly concerned about preserving individual rights. He describes how a due-process leaning may promote anti-authoritarian views and a distrust of the players in the legal system.

Assessing Due Process and Crime Control

Even with knowledge of the two perspectives (and the face validity of the assumption that people differ along this dimension), attempts to evaluate whether the distinction actually exists have been scarce.

The first step is to define exactly what one means by a due process and crime control distinction. Those addressing the topic have used Packer's (1964) outline of the models as a starting point (Thompson, Cowan, Ellsworth, & Harrington, 1984; Fitzgerald & Ellsworth, 1984; Liu & Shure, 1993).

Vanessa A. Edkins and Kenneth D. Royal

A simplistic view in assessing these perspectives is an assumption that we can define due process or crime control by their correlations with other attitudes (see Fitzgerald and Ellsworth, 1984, for their interpretations of previous surveys assessing public attitudes); in other words, proposing the definitions of due process and crime control a posteriori from observing the relationships between attitudes concerning the death penalty, pro-prosecution or pro-defense standings, authoritarianism and conservatism (see Fitzgerald & Ellsworth, 1984).

The research on death-qualified jurors indicates that attitudes toward the death penalty relate to attitudes toward other aspects of the legal system (Fitzgerald & Ellsworth, 1984). Death-qualified jurors are pro-prosecution, suspicious of defendants, distrusting of defense attorneys, and unlikely to favor due process initiatives in the criminal justice system. Fitzgerald and Ellsworth suggest that this shows attitudes toward the death penalty predicted whether a juror was due-process oriented (focused on preserving individual rights in the justice system) or crime-control oriented (focused on swift, efficient justice).

Liu and Shure (1993) believed that the ideas of due process based on Packer's (1964) model did not clearly distinguish between due process as an ideological concept and due process as a basis for legal decision making. The authors propose that what researchers have been calling due process or crime control actually represents social and political ideas of justice when in fact these concepts are not necessarily related; specifically, due process attitudes do not always coincide with political liberalism. The example they provide is the American Civil Liberties Union's—strong supporters of all that represents due process and preservation of individual rights—decision to support the right of a Nazi party to march in a pre-dominantly Jewish suburb of Skokie, Illinois. Liberals would not likely set aside their distaste for such a group in order to preserve individual rights.

What Liu and Shure (1993) are suggesting is correct, but a strict legal definition of "due process" is not that defined by Packer's (1964) model, nor that perceived by the general public (Thompson et al., 1984; Fitzgerald & Ellsworth, 1984). Packer (1964) states that a part of the due process perspective that is important, though often not acknowledged, is a "mood of skepticism about the morality and the utility of the criminal sanction, taken either as a whole or in some of its applications" (p. 20). Removing the emotional and ideological content from the due process and crime control orientation is ignoring important parts of the constructs, such as the anger a crime-control individual may feel when an offender is given a light sentence or the cynicism a due-process individual may feel when hearing about an illegal search.

In their research, Liu and Shure (1993) found that strictly legal due process was not related to "due process/crime control" (as represented by 7 items from the Fitzgerald and Ellsworth (1984) scale and one new, similar item composed by the authors). Interestingly, the due process/crime control ideology based attitude assessment accounted for more of the overall variance in a scale combining this measure and the legal due process measure. The authors did not report which scale was a better predictor of overall verdicts for their simulated trial.

While all of the previous studies focused on what effect crime control/due process attitudes may have on an individual juror, and the definitions of the

concepts (as ideologies) across the studies are near identical (Casper et al., 1989; Liu & Shure, 1993; Fitzgerald & Ellsworth, 1984; Thompson et al., 1984), none of them have developed a valid measure built from the theoretical constructs to examine exactly how due process and crime control perspectives are structured and how they may relate to one another. The current paper addresses this topic.

The Current Research

One can think of the due process and crime control orientations as individual characteristics. A due process orientation represents a multitude of beliefs related to the law and to those involved in the legal system with the unifying characteristic of concern for individual rights. The same can be said of crime control beliefs and a concern for keeping order in the legal system. The content of these perspectives provides specific predictions about how an individual endorsing either perspective would react toward different aspects of the legal system; that is, due process individuals would frown upon illegal searches and seizures while crime control individuals would be disturbed if evidence indicative of guilt collected from the searches was deemed inadmissible. Assessing the due process and crime control attitudes of prospective jurors can help predict how they will react to different parts, and participants, of a criminal trial.

Various measures of juror bias have been validated over the years, including the Juror Bias Scale (JBS)(Kassin & Wrightsman, 1983; Myers & Lecci, 1998), the Legal Attitudes Questionnaire (LAQ) and Revised Legal Attitudes Questionnaire (RLAQ) (Kravitz, Cutler, & Brock, 1993), and the more recent Pretrial

Juror Attitude Questionnaire (PJAQ)(Lecci & Myers, 2008). Although both the JBS and the PJAQ have shown predictive validity for individual juror verdicts—the JBS predicts around 3% of the variance in verdict decisions (Myers & Lecci, 1998) and the PJAQ predicts upwards of 7% of variance in verdict decisions (Lecci & Myers, 2008)—the development of the scales was not based in theory, and their content is very broad. Similar to the LAQ, which measures authoritarianism ideologies as they relate to the legal system and in a meta-analysis was shown to account for about 4% of the variance in a juror's verdict (Narby, Cutler, & Moran, 1993), a measure looking at due process and crime control orientations would be theory-based and able to proffer a number of predictions for the future behavior of individuals endorsing the items.

The existence of these current measures of juror bias does not preclude the utility of the new measure constructed for this study, the General Attitudes toward the Legal System scale (GALS). This new measure will tap into due process and crime control endorsement which, although often discussed in research and the legal realm (Casper et al., 1989; Fitzgerald & Ellsworth, 1984; Liu & Shure, 1993; Thompson et al., 1984) and with generally agreed upon construct content, has not been empirically tested by its own, independent measure. The content of the crime control and due process perspectives offer predictions for individual differences at every major part of the legal system: attitudes toward police, investigation tactics, lawyers, the court, sentencing, constitutional rights, etc. With the theoretical constructs as a reference, one can tap into specific legal views that together create an overall conception of the legal system.

Vanessa A. Edkins and Kenneth D. Royal

Method

Participants

A total of 744 Introductory to Psychology students at a large Midwestern university in the United States filled out an online version of the GALS scale for partial course credit. The sample had been previously randomly divided in order to perform independent exploratory and confirmatory factor analysis for another purpose. For the 372 participants designated to the exploratory factor analysis (EFA), 12 participants were removed from the sample for answering "Neither agree nor disagree" to all items. Eight participants were removed from the 372 participant confirmatory factor analysis (CFA) sample for the same reason.

An ANOVA test was performed on the EFA sample to assess whether there was a difference in overall scale score for participants taking less than two minutes ($N = 25$, $M = 109.3$, $SD = 12.8$), between two and three minutes ($N = 79$, $M = 102.9$, $SD = 13.2$), and three minutes or more ($N = 256$, $M = 101$, $SD = 11.5$). Results showed there was a significant difference with regard to how long a participant took to complete the survey ($p < .01$, $\eta^2 = .03$). Post hoc analyses confirmed that those participants finishing in less than two minutes had overall scale scores significantly different from those spending three minutes or more ($p < .01$). For the EFA these individuals could be filtered out, but because the CFA was going to be performed with a statistics package that could not filter out these individuals, the 27 individuals in this sample that took less than 2 minutes to complete the questionnaire had to be removed. When

the data were recombined for the purpose of the current study—to utilize a Rasch modeling approach—a total of 697 participants remained.

Measures

Items generated spanned multiple aspects of the legal system, including attitudes toward police and policing, lawyers, judges, jurors and general legal system concepts such as sentencing and parole. DeVellis' (1991) guidelines for scale construction were followed and items were generated to directly assess the due process and crime control perspectives as they are commonly discussed and as they are defined by Packer (1964; 1968). As recommended for scale construction (Kerlinger, 1984), previous scales were consulted for item ideas including the Juror Bias Scale (Kassin & Wrightsman, 1983), the Revised Legal Attitudes Questionnaire (Kravitz et al., 1993), Attitudes toward the Criminal Legal System (Martin & Cohn, 2004), the Scale of Attitudes toward the Legal System (Wrightsman & Schiffhauer, 1995), and Attitudes toward Entrapment (Butler & Wrightsman, 2002). Direct item adaptation from these scales was as follows:

From Wrightsman and Schiffhauer's (1995) Scale of Attitudes toward the Legal System the item, "Defendants who are guilty often "get off" because of technicalities" was changed to "Too often a defendant gets off because of technicalities." From Kassin and Wrightsman's (1983) Juror Bias Scale, the item "Out of every 100 people brought to trial, at least 75 are actually guilty of the crime with which they are charged" was adapted and changed to "Out of every 100 persons brought to trial, at least 90 are actually guilty" and the item "Defense

Vanessa A. Edkins and Kenneth D. Royal

lawyers really don't care about guilt or innocence, they are just in business to make money" was adapted and changed to "Defense attorneys are only in it for the money" each trying to tap into a Crime Control attitude/affect or a Due Process attitude/affect. From Butler and Wrightsman's (2002) Attitudes toward Entrapment scale the items "Use of informants (who pose as friends but actually work for the police) is allowable if that is what it takes to catch lawbreakers," "The crime rate is so high that we should give the police the power to catch criminals, whatever it takes," and "Police should be allowed to do whatever is necessary to catch those who are breaking the law" were adapted as is.

For each statement, individuals were asked to indicate their agreement on a 7-point Likert scale ranging from strongly disagree (-3) to strongly agree (+3).

Analysis

Rasch models are routinely used in medicine, various health sciences, psychology and other fields to validate the psychometric properties of survey instruments. Traditional validation studies often involve summing and averaging Likert-scale data and treating the data as interval measures. This is a statistical violation as Likert-scale data are actually ordinal in nature. Raw scores are not measures; therefore, a linear transformation process must take place in order to produce a continuum that has the qualities of a true metric. Rasch models are generally considered a superior approach to analyzing survey data as ordinal raw score data can be converted into interval measures as logarithmic values of odds (logits) and overcome these important, and often overlooked, assumptions (Bond & Fox, 2007).

Rasch models are also widely acclaimed because they possess the property of invariance. Measurement scholars have long noted that students who receive a high test score may do so because they received an easy test. Conversely, those who performed poorly may have done so because they received a difficult test. Being able to separate a person's ability from the instrument's difficulty is both desirable and necessary for objective measurement. Traditional methods to analyze survey data do not possess this property, thus they are inherently sample-dependent. Rasch models, on the other hand, are able to separate a person's ability (or other latent trait) and an item's difficulty and map both on a linear continuum so that more meaningful inferences can be made.

All Rasch models assume a more able person should always have a greater probability of success on any given item than a less able person. Similarly, an easier item should always have a greater probability of being answered correctly than a more difficult item. This notion seamlessly transcends into survey research where a latent trait such as attitudes, perceptions, beliefs, etc. is measured instead of ability, per se. With Rasch models, a person's response to an item is predicted based on the interaction between that person's ability (or "severity of opinion" in the case of a questionnaire) and the question's "difficulty to endorse". The probability that a respondent will agree with a particular item is a logistic function of the relative distance between the person and the item's hierarchical location on a linear continuum. Therefore, a participant who possesses a great amount of the latent trait would have a high probability of agreeing with an item that is relatively easy to endorse. Conversely, a participant who possesses a

small amount of the latent trait would have a lower probability of endorsing that same item.

Specific to this study, The Rasch Rating Scale Model (RRSM; Andrich, 1978) was utilized for data analysis. The RRSM is a Rasch model that is appropriate for analyzing survey data that was collected using static rating scale categories. The formulae for the model is:

$$\ln\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = B_n - D_i - F_j$$

where, P_{nij} is the probability that person n encountering item i is observed in category j, B_n is the "endorsability" measure of person n, D_i is the "difficulty" measure of item i, the point where the highest and lowest categories of the item are equally probable. F_j is the "calibration" measure of category j relative to category $(j-1)$, the point where categories $(j-1)$ and j are equally probable relative to the measure of the item. No constraints are placed on the possible values of F_j.

Winsteps (Linacre, 2010) measurement software estimated the parameters for the model using joint maximum likelihood estimation procedures (Wright & Masters, 1982). In the present study, person measures (B_n) refer to the survey respondents' tendency to endorse items based on their attitudes toward due process and crime control. Item measures (D_i) refer to the difficulty of endorsing a particular item. Threshold measures (F_j) refers to the difficulty of assigning a rating of j versus $j-1$ on the rating scale. Standard errors are estimated for every person, item and threshold measure as well. Estimating parameters can be problematic when respondents heavily select extreme categories on the rating scale, so Winsteps adjust the raw scores by adding or subtracting 0.3 raw score points from either the zero score or perfect score. This adjustment properly parameterizes the measures (Linacre, 1999; 2010a; Wright and Panchapakesan, 1969).

The evaluation of the psychometric properties of the GALS scale focused on six criteria: dimensionality, internal consistency, rating scale category effectiveness, item quality, item hierarchy and person measure quality. First, to investigate dimensionality we performed a principal components analysis of residuals. This procedure is useful in identifying the amount of variance explained by each extracted principal component. Second, we performed reliability analyses to evaluate the degree to which the measures are reproducible. Third, we evaluated rating scale category effectiveness to determine if the rating scale was functioning properly. Fourth, we investigated item quality by examining item fit statistics. Fifth, we evaluated the item hierarchy to determine if the empirical hierarchy was consistent with theory-based expectations. Sixth, we evaluated person measures by investigating person fit statistics.

Results

Dimensionality

A principal components analysis of residuals was performed by Winsteps measurement software to investigate unidimensionality. In total, 33.9% of the Rasch dimension was explained. The largest secondary dimension explained 9.5% of the variance. Overall, the variance explained by the items was 30.5%. This is about three times the variance explained from the first contrast, thus indicating a

second dimension is present. The eigenvalue of the first contrast is 3.6, indicating it has a strength of about 4 items (out of 25 total items). The Rasch dimension is strong enough to assert a significant primary dimension, however the strength of the second dimension is also somewhat noticeable. Considering this evidence, we consider the model to have reasonable unidimensional fit.

Reliability

Reliability and separation measures indicate the extent to which scores are reproducible (see Table 1). Here, "Real" and "Model" reliability and separation are reported. Real reliability refers to the "worst case estimates", and model refers to "best case estimates"; true reliability falls somewhere in between. Person reliability for this sample ranges from .66 to .72, indicating moderate internal consistency. Item reliability estimates are stable at 1.00 indicating high item reliability. Separation measures indicate the number of statistically distinguishable levels in the data. Separation estimates ranging from 1.38 to 1.60 for persons indicates sufficient spread. Item separation measures of 14.33 to 15.10 also indicate adequate spread of items.

Table 1
Reliability and Separation Measures

Category	Real Reliability	Model Reliability	Real Separation	Model Separation
Persons	.74	.78	1.71	1.90
Items	1.00	1.00	14.33	15.10

Rating Scale Effectiveness

An investigation of the rating scale effectiveness involved evaluating the sample's use of the rating scale and the scale's inferential value (see Table 2). Linacre (2002) suggests a number of quality control checks to ensure adequate rating scale functioning. Counts and percents were provided to determine the extent to which the various rating scale categories were utilized by survey respondents. Infit and outfit mean square fit statistics indicate the extent to which each rating scale category is "noisy", or producing calibrations that are not desirable for productive measurement. Structure calibration refers to the calibrated measure of transition between categories. Also called "step calibration", this measure indicates how difficult it is to observe each category. Results indicate the sample aptly utilized the full rating scale. Fit statistics are well within the recommended ceiling range of 2.0, indicating relatively noise-free calibrations. Step calibrations advance appropriately from smallest to largest in accordance with the direction of the scale.

Vanessa A. Edkins and Kenneth D. Royal

Table 2
Rating Scale Effectiveness

Category	Count	Percent	INFIT MNSQ	OUTFIT MNSQ	Structure Calibration	Category Measure
Strongly Disagree	792	5	.96	.98	NONE	-2.79
Moderately Disagree	1,796	11	1.05	1.06	-1.37	-1.43
Slightly Disagree	3,073	19	.97	.97	-.87	-.66
Neutral	3,622	23	.92	.92	-.29	-.06
Slightly Agree	3,858	24	1.01	1.02	.03	.59
Moderately Agree	2,066	13	1.03	1.04	.96	1.48
Strongly Agree	843	5	1.03	1.02	1.53	2.93

Item Measure Quality

Item measure quality is determined by investigating the extent to which the items vary in difficulty to endorse, the size of the standard errors and the degree to which the items fit the model's expectations (see Table 3). Item difficulty calibrations ranged from -1.14 to .92 logits, indicating adequate discrimination for data analyzed via the RRSM. Standard error estimates for each item were quite small, ranging between .03 and .04. As mentioned previously, fit statistics are useful for identifying noisy measures. Wright and Linacre (1994) indicate for rating scales, values of .6 to 1.4 are ideal, although values between .5 and 2.0 are not degrading for measurement. With regard to the present data, only one item stands out as being potentially problematic. Q1, *In our legal system, everyone receives a fair trial*, appears to slightly misfit the model's expectations. Further qualitative investigation into this item is necessary before considering its removal from the survey.

Table 3
Item Fit Statistics

Item	Measure	*SE*	INFIT MNSQ	OUTFIT MNSQ
Q1 - In our legal system, everyone receives a fair trial	.04	.03	1.65	1.65
Q2 - Judges only take the facts of a case into consideration during sentencing	.11	.03	1.18	1.18
Q3 - Plea-bargaining allows too many criminals back on the street	-.36	.03	.91	.91
Q4 - People with more money are more likely to win their trials	-1.14	.04	1.36	1.33
Q5 - Lawyers will often lie to win their case	-.72	.03	1.19	1.17
Q6 - Prosecutors are only concerned with their	-.13	.03	.80	.79

Vanessa A. Edkins and Kenneth D. Royal

Item	Measure	*SE*	INFIT MNSQ	OUTFIT MNSQ
conviction rates				
Q7 - Too many criminals are given lighter sentences than they deserve	-.41	.03	.97	.96
Q8 - Sentences given to racial minorities are often too harsh	-.03	.03	1.04	1.05
Q9 - Too often a defendant gets off because of technicalities	-.37	.03	.67	.67
Q10 - It is too easy for criminals to appeal their case and get off	-.11	.03	.75	.75
Q11 - Police often overstep their boundaries when pursuing a suspect	-.22	.03	1.13	1.14
Q12 - Defense attorneys are generally very honest	.36	.03	.85	.88
Q13 - Police will charge the first suspect that comes along	.24	.03	.95	.95
Q14 - Prosecutors are generally very honest	.24	.03	.83	.85
Q15 - Defense attorneys are only in it for the money	.16	.03	.80	.80
Q16 - It is too easy for a violent offender to be paroled	-.34	.03	.73	.73
Q17 - Out of every 100 persons brought to trial, at least 90 are actually guilty	.07	.03	.83	.84
Q18 - If the police arrest someone, then that person is almost certainly guilty	.92	.03	1.10	1.11
Q19 - Too often a defendant is convicted on evidence that was collected illegally	.04	.03	.71	.71
Q20 - Use of informants (who pose as friends but actually work for the police) is allowable if that is what it takes to catch lawbreakers	-.20	.03	1.35	1.34
Q21 - Too often people who commit crimes are given too harsh of sentences	.45	.03	1.00	1.00
Q22 - The crime rate is so high that we should give the police the power to catch criminals, whatever it takes	.62	.03	1.16	1.17
Q23 - Police should be allowed to do whatever is necessary to catch those who are breaking the law	.69	.03	1.27	1.27
Q24 - Police often bend the rules in order to strengthen a weak case	-.28	.03	.87	.88
Q25 - All evidence collected against a person, no matter how it was collected, should be admissible in court	.38	.03	1.18	1.20

Vanessa A. Edkins and Kenneth D. Royal

Item Hierarchy

The item hierarchy investigates the extent to which the items rank-order themselves in a manner that is consistent with theory. The item map presents an illustration of the item hierarchy (see Figure 1). Item Q18, *If the police arrest someone, then that person is almost certainly guilty*, is the most difficult item for respondents to endorse, followed by Q23, *Police should be allowed to do whatever is necessary to catch those who are breaking the law*, and Q22, *The crime rate is so high that we should give the police the power to catch criminals, whatever it takes*. Item Q4, *People with more money are more likely to win their trials*, is the easiest

item for respondents to endorse, with item Q5, *Lawyers will often lie to win their case*, being the next easiest item. Notice, the mean for both person and item measures falls almost exactly at 0.0 logits. This indicates the average participant in this sample would have a 50/50 probability of endorsing items that constitute the average difficulty, namely items Q1 and Q8. The probability of endorsing an item increases the farther an item is located down the map. Likewise, the probably of endorsing an item decreases the farther an item is located up the map. The precision as to how much these probabilities increase or decrease is contingent upon the location of the person measure used for the comparison.

```
                Person - MAP - Attitude
                  <more>|<rare>
      2              .  +                      Most Difficult to Endorse
                        |
                        |
                        |
                     .  |
                     .  |
                     .  |
      1              .  +T Q18
                     .  |
                     .  | Q23
                     .  | Q22
                   .## T|
                    .#  |S Q21
                  .#### |  Q12    Q25
                 .##### S|  Q13    Q14
                .########|  Q15
               .#########|  Q2     Q17    Q19
      0   .###########  M+M Q1     Q8
           .############|  Q10
            #########  |  Q6     Q11    Q20
             .#### S|  Q24
              .####  |  Q3     Q9     Q16
               ##   |S Q7
                .  T|
                .   |
                .   |
                .   |  Q5
     -1              +T
                .   |
                .   |
                    |  Q4
                    |
                    |
                    |
                    |
                    |
     -2              +
                  <less>|<frequent>             Least Difficult to Endorse
         EACH "#" IS 8. EACH "." IS 1 TO 7
```

Figure 1. Item Map

Vanessa A. Edkins and Kenneth D. Royal

Person Measure Quality

Person measure quality is determined by investigating the stability of the measures, the size of their associated standard errors, and the extent to which the measures are noisy (see Table 4). For this sample, the measures are relatively stable. The average standard error associated with the person measure is about .15, indicating acceptable stability. Mean square infit and outfit statistics are close to the ideal value of 1.0, indicating relatively noise-free calibrations. Using the .6-1.4 criteria as suggested by Wright and Linacre (1994) as reasonable fit statistics ranges, person misfit was evaluated. Results indicate about 19% of the sample misfit the model, which on the surface seems fairly large. However, as

Karabatsos (2000) points out, the sampling distribution for the outfit mean square is unknown, so this information could potentially be misleading. Wright and Linacre also point out that only fit statistics outside the range of .5-2.0 are unproductive or degrading for measurement. When considering this criteria, only about 8% of the persons misfit the RRSM. Rasch practitioners regularly remove persons and items who misfit the model (Linacre, 2010b). For the purposes of this validation study, no misfitters were removed. The authors chose to retain all the data, including those that might be noisy, or otherwise problematic. However, in practice, it would be perfectly reasonable to remove all misfitters from the sample given the proportion of misfitters is so small.

Table 4
Overall Data to Model Fit Statistics

	Measure	Model Error	INFIT MNSQ	OUTFIT MNSQ
Persons				
M	.01	.15	1.01	1.01
SD	.29	.00	.56	.57
Items				
M	.00	.03	1.01	1.01
SD	.44	.00	.24	.23

Discussion

In 1989, Messick discussed the notion of construct validity as a uniform concept. Messick suggests construct validity "is based on an integration of any evidence that bears on the interpretation or meaning of the test scores..." (p. 7). Specifically, Messick identified six distinguishable aspects that comprise construct validity: content, structural,

substantive, generalizability, external, and consequential. When determining the construct validity of the GALS scale, it is helpful to evaluate the results from this perspective.

Item quality measures appeared sound with the possible exception of one item: Q1, *In our legal system, everyone receives a fair trial.* This speaks to the content validity aspect. With regards to Q1, perhaps the reason that this item misfit

the model was that it did not adequately distinguish between the crime control and due process orientation. The item may have been interpreted differently by different people and even the idea of a "fair trial" may have had different connotations. If this item stays in the GALS scale, future research may want to investigate exactly what the roughly 50% of individuals have in mind when they disagree with the item and denote that people are not receiving a fair trial.

Each of the rating scale effectiveness quality control checks was acceptable and evidence was provided that the rating scale was functioning appropriately. This speaks to the structural aspect of validity. Also with regard to the rating scale, there appears to be some evidence for communicative validity (Lopez, 1996), or the extent to which the rating scale categories are sufficient and appropriately interpreted by respondents. Reliability estimates were moderate for persons and high for items. These estimates lend support for relatively reproducible scores, thus speaking to the generalizability aspect of validity.

With respect to the substantive aspects of validity, looking at the item hierarchy in Figure 1, it seems that Packer's (1964) vision has garnered support. For example, items Q3, *Plea-bargaining allows too many criminals back on the street*; Q9, *Too often a defendant gets off because of technicalities*; Q16, *It is too easy for a violent offender to be paroled*; and Q7, *Too many criminals are given lighter sentences than they deserve*, are all very close together on the item map and all four reflect crime control views that the current system is too lenient.

Recall that the "presumption of guilt" aspect of the crime control ideology discussed by Packer (1964) encompassed the idea that if investigations are conducted properly and efficiently, we should assume the individual charged is likely guilty. Looking at the item hierarchy, one can see this faith in the police and endorsement of increasing police power construct as represented by the close proximity of three specific items: Q18, *If the police arrest someone then that person is almost certainly guilty*, Q22, *The crime rate is so high that we should give the police the power to catch criminals, whatever it takes*, and Q23, *Police should be allowed to do whatever is necessary to catch those who are breaking the law*. Interestingly, these items were also the most difficult to endorse (see Figure 1). The fact that these low endorseability items are so similar in nature, construct-wise, could tell someone a lot about the individual that endorses them. That individual would likely be more forgiving of police misconduct, if that is a component in a trial, and more conviction-prone in general. Future research will need to address whether these questions have predictive validity in an actual trial scenario.

The low endorseability of these staunch crime control items may also be telling us something about the current state of society. Packer (1964) felt that throughout history, society moves back and forth along the ideological continuum between favoring more due process initiatives to favoring more crime control initiatives. It may be tempting to think that the low endorseability of items related to a key component of crime control—a trust in investigations and police to uncover the truth; a belief in the factual over legal guilt—mean that we are currently a 'due-process society' but in reality, it may be showing us that the relationship is much more complex than Packer originally thought. Recall that Packer's ideologies are complex and

Vanessa A. Edkins and Kenneth D. Royal

multi-faceted—another important aspect of crime control is distrust in, and a low tolerance for, adjudication as well as a strong support of punishment as a deterrent. The three easiest to endorse items show that these beliefs are garnering plenty of support: Q4, *People with more money are more likely to win their trials*, Q5, *Lawyers will often lie to win their case*, and Q7, *Too many criminals are given lighter sentences than they deserve.* Perhaps society can be due-process oriented toward some parts or players of the legal system while at the same time, crime-control oriented when it comes to others.

The substantive validity of the GALS scale is further buttressed by the apparent acknowledgement of the cynicism inherent in a due process ideology. Packer's conception of a due process perspective included a certain amount of cynicism or distrust of the legal system and its players; an anti-authoritarian view that the system was biased. Three items very close together on the item hierarchy reflect this perspective, Q6, *Prosecutors are only concerned with their conviction rates*, Q11, *Police often overstep their boundaries when pursuing a suspect*, and Q24, *Police often bend the rules in order to strengthen a weak case.* These items were about mid-range for endorseability.

We provide no evidence for the external or consequential aspects of validity, although these facets are certainly important. Future research with the GALS scale will assess Differential Item Functioning (DIF) across various demographic characteristics to test for item bias, thus addressing the systematic validity of the scores. And while our validity assessments presented here offer a solid argument for the generalizeability of our results, future research should assess the structure of the GALS scale using a more diverse sample.

Conclusion

The due process and crime control model offers an excellent theoretical basis for differentiating between jurors by their beliefs about the state and purpose of the criminal justice system. Packer's (1964) proposal of the model, and subsequent interpretations (Fitzgerald & Ellsworth, 1984; Thompson et al., 1984; Liu & Shure, 1993), offers predictions for individual differences at every level of the criminal justice system. The current paper presents one of the first empirical attempts to assess the true content and structure of the perspectives. We have low endorseability items that seem to reflect a crime control orientation expressing faith in the police and their actions. Individuals endorsing these items may differ from other potential jurors in their interpretation—and subsequent acceptance—of questionable police behaviors in an investigation. We also have due process orientation items that seem to express cynicism toward the justice system. Individuals endorsing these items are explicitly stating that police should not be trusted and may interpret some trial evidence in a manner opposite the first group.

Unlike most other IRT models, Rasch models do not require normally distributed data. However, when data happen to be relatively normally distributed researchers can make inferences at the summative level. As illustrated in Figure 1, data for the present study are rather normally distributed. Therefore, we can make inferences about how the average person feels about the criminal justice system. This last point is

why expanding the study beyond our current sample could help attorneys or trial consultants in planning trial strategy. If we know that the average person is highly unlikely to endorse an item that states, "Police should be allowed to do whatever is necessary to catch those who are breaking the law," then the defense would want to highlight any police missteps in an investigation and the prosecution should work to explain away or downplay any questionable police actions. On the other hand, knowing that the average person is very likely to endorse the statement, "People with more money are more likely to win their trials," suggests some trial strategy possibilities as well. Perhaps the defense attorneys of a well-off client would want to keep their client looking 'ordinary' – avoid outwardly flashy dress and have only one or two lawyers sitting with the client (opposed to a 'dream team'), for example. Prosecutors, on the other hand, would likely want to accentuate the obvious resources of the client whenever possible. Expanding the current study to include a more diverse population in terms of age, ethnicity, and geographical location, can further highlight how the average U.S. jury-eligible citizen feels about the criminal justice system and its players.

While trial-specific material will always be the best predictor of verdicts, and rightfully so, juror bias does play a role, especially in cases with a certain degree of ambiguity. Since the vast majority of cases in the United States never make it to trial, one could infer that those for which a jury is formed are those with some question of whether or not the defendant is actually guilty. Measuring individual differences regarding the "is" and "ought" (Packer, 1964) of the purpose of legal system may help us understand why those two jurors, sitting side-by-side, can come to different conclusions.

References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracin, B. Johnson & M. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Lawrence Erlbaum Associates.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model. Fundamental measurement in the human sciences, 2nd edition.* Lawrence Erlbaum Associate.

Butler, K., & Wrightsman, L. (2002, March). *Attitudes toward law enforcement and mock jurors' reactions to an entrapment defense.* Paper presented at the American Psychology and Law Society, Austin, TX.

Casper, J., Benedict, K., & Perry, J. (1989). Juror decision making, attitudes, and the hindsight bias. *Law and Human Behavior, 13*, 291-310.

DeVellis, R. F. (1991). *Scale development: Theory and applications* (Vol. 26). Newbury Park, CA: Sage Publications.

Field, H. (1978). Juror background characteristics and attitudes toward rape: Correlates of jurors' decisions in rape trials. *Law and Human Behavior, 2*, 73-93.

Fitzgerald, R., & Ellsworth, P. (1984). Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior, 8*(1/2), 31-51.

Hans, V., & Appel, A. (1999). The jury on trial. In W. Abbott & J. Batt (Eds.), *A handbook of jury research.*

Philadelphia, PA: The American Law Institute.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury.* Cambridge, MS: Harvard University Press.

Kaplan, M., & Kemmerick, G. (1974). Juror judgment as information integration: Combining evidential and nonevidential information. *Journal of Personality and Social Psychology, 30*, 493-499.

Kaplan, J., & Miller, L. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology, 36*(12), 1433-1455.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*, 152-176.

Kassin, S., & Wrightsman, L. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality, 17*, 423-442.

Kerlinger, F. (1984). *Liberalism and conservatism: The nature and structure of social attitudes.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kravitz, D., Cutler, B., & Brock, P. (1993). Reliability and validity of the original and revised Legal Attitudes Questionnaire. *Law and Human Behavior, 17*, 661-677.

Lecci, L., & Myers, B. (2008). Individual differences in attitudes relevant to juror decision making: Development and validation of the Pretrial Juror Attitude Questionnaire (PJAQ). *Journal of Applied Social Psychology, 38*, 2010-2038.

Lieberman, J., & Sales, B. (2007). *Scientific jury selection.* Washington, DC: American Psychological Association.

Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of Outcome Measurement, 3*(4), 382-405.

Linacre, J. M. (2002). Understanding Rasch measurement: Optimal rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Linacre, J. M. (2010a). WINSTEPS® (Version 3.69.1). Computer Software. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2010b). When to stop removing items and persons in Rasch analysis? *Rasch Measurement Transactions, 23*(4), 1241.

Liu, J., & Shure, G. (1993). Due process orientation does not always mean political liberalism. *Law and Human Behavior, 17*(3), 343-360.

Lopez, W. "Communication validity and rating scales." *Rasch Measurement Transactions, 10*(1) (1996): 482-483.

Martin, T., & Cohn, E. (2004). Attitudes toward the criminal legal system: Scale development and predictors. *Psychology, Crime & Law, 10*(4), 367-391.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Moran, G., Cutler, B., & Loftus, E. (1990). Jury selection in major controlled substance trials: The need for extended voir dire. *Forensic Reports, 3*, 331-348.

Myers, B., & Lecci, L. (1998). Revising the factor structure of the Juror Bias Scale: A method for the empirical validation of theoretical constructs. *Law and Human Behavior, 22*(2), 239-256.

Narby, D., Cutler, B., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability. *Journal of Applied Social Psychology, 78*, 34-42.

Ostrom, T., Saks, M., & Werner, C. (1978). An integration theory of analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology, 36*(4), 436-450.

Packer, H. L. (1964). Two models of the criminal process. *University of Pennsylvania Law Review, 113*(1), 1-68.

Packer, H. L. (1968). *The limits of the criminal sanction.* Stanford, CA: Stanford University Press.

Thompson, W., Cowan, C., Ellsworth, P., & Harrington, J. (1984). Death penalty attitudes and conviction proneness: The translation of attitudes into verdicts. *Law and Human Behavior, 8*(1/2), 95-114.

Vidmar, N., & Schuller, R. (1989). Juries and expert evidence: Social framework testimony. *Law and Contemporary Problems, 52*, 133-176.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch Measurement.* Chicago: MESA Press.

Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23-48.

Wrightsman, L., & Schiffhauer, K. (1995, November). The measurement of attitudes toward the legal system: Unpublished paper. Department of Psychology, University of Kansas, Lawrence, KS.