



**Manchester
Metropolitan
University**

Willmott, Dominic ORCID logoORCID: <https://orcid.org/0000-0002-7449-6462>, Boduszek, Daniel, Debowska, Agata and Woodfield, Russell (2018) Introduction and validation of the Juror Decision Scale (JDS): An empirical investigation of the Story Model. *Journal of Criminal Justice*, 57. pp. 26-34. ISSN 0047-2352

Downloaded from: <https://e-space.mmu.ac.uk/625313/>

Version: Accepted Version

Publisher: Elsevier BV

DOI: <https://doi.org/10.1016/j.jcrimjus.2018.03.004>

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Please cite the published version

<https://e-space.mmu.ac.uk>

Manuscript Details

Manuscript number	JCJ_2018_57_R1
Title	Introduction and Validation of the Juror Decision Scale (JDS): An Empirical Investigation of the Story Model
Article type	Full Length Article

Abstract

Purpose: To develop and validate a self-report measure of individual juror decision making within criminal trials, based on theoretical features set out in the Story Model of juror decision making. Methods: The Juror Decision Scale (JDS) and Acceptance of Modern Myths about Sexual Aggression (AMMSA) measure were completed by 324 jury-eligible participants split across 27 jury panels, after observing a rape trial re-enactment high in ecological validity. Dimensionality and construct validity of the JDS was investigated using traditional confirmatory factor analysis (CFA) techniques alongside confirmatory bifactor analysis at two time points (individual juror verdict decisions pre- and post-deliberation). Three competing models of the JDS were specified and tested using Mplus with maximum likelihood robust estimation. Results: Bifactor model with three meaningful factors (complainant believability, defendant believability, decision confidence) was the best fit for the data at both decision points. Good composite reliability and differential predictive validity were observed for the three JDS subscales. Conclusion: Alongside demonstrating its multidimensional conceptualisation, the JDS development permits future empirical testing of the Story Model theoretical assertions surrounding juror decision making. Present findings also provide early evidence of a certainty principle assessment process governing individual verdict decision formation. Theoretical and practical applications are discussed.

Keywords	Juror Decision Scale (JDS); Story model; Confirmatory factor analysis; Differential predictive validity; Jury decision making; certainty principles
Taxonomy	Forensic Psychology, Psychology Research Method, Psychometrics
Corresponding Author	Daniel Boduszek
Corresponding Author's Institution	University of Huddersfield
Order of Authors	Dominic Willmott, Daniel Boduszek, Agata Debowska, Russell Woodfield

Submission Files Included in this PDF

File Name [File Type]

Revision Response letter.docx [Response to Reviewers (without Author Details)]

Highlights.docx [Highlights]

title.docx [Title Page (with Author Details)]

Complete Final Revised Manuscript_JDS Validation_AD.docx [Manuscript (without Author Details)]

Figures.docx [Figure]

Tables_revised.docx [Table]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Manuscript ID: JCJ_2018_57

Title: Introduction and Validation of the Juror Decision Scale (JDS): An Empirical Investigation of the Story Model

Dear Professor DeLisi,

Please find attached our revised manuscript entitled “*Introduction and Validation of the Juror Decision Scale (JDS): An Empirical Investigation of the Story Model*” which you kindly invited a revised manuscript be submitted. We have taken into account and addressed all reviewer comments to the original manuscript and have made revisions accordingly. We hope that the responses below and the revisions made, will allow the reviewers decision alongside your final editorial decision, to be made with ease.

Sincerely

Prof Daniel Boduszek

Reviewer 1

I have but two minor comments/suggestions for the authors to consider: The bigger picture, however, is the potentially important contribution the paper makes through the development of the juror decision scale.

Comment 1: The authors should note the limitations to this study posed by the use of a non-representative sample.

Response 1: Thank you for your comments. The limitations surrounding the study sample have now been added within the discussion section (please see page 23).

Comment 2: The authors should either (a) identify and defend in the front-end of the manuscript their decision to include the self-esteem scale or (b) simply delete all of this material from the paper. As it is, the inclusion of self-esteem does little to nothing to help the authors address their research questions; its inclusion appears pointless.

Response 2: Taking these comments into account, all reference to self-esteem has now been removed from the manuscript. Reflecting upon both reviewers making this point, the authors agree that using self-esteem as an external variable to display differential predictive validity with JDS sub-scales added little but confusion to the manuscript.

Reviewer 2

In a single experiment with mock jurors, the author(s) introduce and validate a new scale (Juror Decision Scale or JDS) of juror decision making. Scale construction is based on the theoretical assumptions of the Story Model explanation of decision making. Using this scale, they found that a model with complainant believability, defendant believability, and decision confidence provided the best fit for the data. There were some spectacular aspects to this study including the development of the experimental trial and the nature and size of the participant sample. In its present form, however, there are number of issues with the paper that limit the impact of this very important data set:

Comment 1: In my view, the author(s) has failed to capitalise on the strengths of their data—leaving us with a somewhat clinical (and not very interesting) series of statistical analyses.

- (a) For example, the reader gets no sense of what decision the jurors made or whether their decisions changed after deliberation.
- (b) Did any of the jurors' characteristics (gender, age, education) influence their initial decision? Their subsequent decision? I understand that the focus of the paper is on an experimental evaluation of the Story Model, but in limiting their data analysis to this point they have robbed the reader of much more interesting story.

Response 1: Thank you for your comments and the value you see in the publication of the manuscript after revisions. Please see how we have addressed/respond to your specific points below.

- (a) Whilst we understand the reviewer's concerns surrounding the statistics provided, these are the standard series of analyses necessary within any validation of a new scale. Having said this, an additional table (Table 1) has now been included which displays individual and collective verdict decision frequencies, with a description also provided in the results section (page 16) detailing the outcome of the McNemar Chi-square test for association between pre- and post-deliberation verdict decisions. Additional analysis surrounding the association between the three JDS sub-scales and verdict decisions at both time points has also now been included in the manuscript (Table 6) with a description in the results section (page 18) and interpretation of these associations provided within the discussion. We hope the inclusion of this additional information and interpretation surrounding verdict outcomes, leads the reviewer to conclude a more interesting analysis is now provided.
- (b) The inclusion of analysis surrounding the role of juror psychosocial characteristics upon verdict decision making is a substantial area in itself and forms the basis of a separate manuscript. In line with scientific convention surrounding what should be included within the validation of a new measure, the inclusion of analysis pertaining to such psychosocial characteristics is beyond the scope and purpose of the present paper.

Comment 2: Many of the potentially interesting (and sometimes puzzling) correlations between variables only emerge in the Discussion without a description of the full correlation matrix in the Results.

- (a) Although it makes sense that there would be a negative relation between complainant believability and rape myth acceptance (although we have no idea the size or strength of this relation here).

(b) The negative relation between self-esteem and decision confidence makes less sense. Given that the full correlation matrix is not presented, the reader has no way to evaluate the size of strength of this relation. What other interesting correlations were not discussed?

Response 2: A correlation matrix is not included in the manuscript as the associations with external variables such as rape myth acceptance serve only to display differential predictive validity with the three JDS sub-scales e.g. that they measure distinct latent constructs.

(a) The size and strength of the association between all JDS sub-scales and external variables (rape myth acceptance and verdict decisions) are already presented in Table 6. We respectfully draw the reviewer's attention to fact that the standardized beta value (β) in Table 6 is the value that is universally used to display the size and strength of the relationship between continuous variables in linear regression analyses and the odds ratio value (OR) universally reported for effect size within binary logistic regression. Nonetheless, for clarity of expression and to further assist the reader, these effect size indicators have now also been included within the discussion section where the associations between such variables are discussed.

(b) As highlighted above self-esteem has now been entirely removed from the manuscript. Please see the above points outlining what values display the size/strength of associations.

Comment 3: There are places in the paper where a potentially important idea is glossed over or not fully explained. For example, on page 22 it is not clear to me why the defendant would uniformly be considered part of the in-group while the complainant would be part of the out-group.

Response 3: Taking these comments into account, all reference to self-esteem has now been removed from the manuscript. Reflecting upon both reviewers making this point, the authors agree that using self-esteem as an external variable to display differential predictive validity with JDS sub-scales added little to the manuscript but confusion. Greater depth of explanation of the JDS sub-scales association with AMMSA scores and verdict decisions is now provided, with the findings further situated in line with past research (please see page 21 & 22).

Comment 4: No attempt is made to tell the reader what the possible practical implications of the data might be. What conclusions can be drawn that would be useful to a lawyer or a judge?

Response 4: Practical implications/applications in light of the legal system are now discussed on page 23.

Minor issues:

1. Throughout the paper, the author(s) uses human verbs with non-human nouns. For example, research does not seek, examine, conclude—researchers do these things.
 - This has now been addressed throughout the manuscript.
2. McCabe, Krauss, & Lieberman (2010) is not in the reference section.
 - This has now been included (see page 28)

3. To give more depth to the procedure, it would be helpful to provide some sample items from each of the measures (i.e., JDS, AMMSA, RSES).
 - This has now been included for the AMMSA scale (see page 14) and reference made to the table where the JDS items/scale can be found (see page 13 & 14). As outlined above the RSES is no longer included in the manuscript.

- We introduce a new Juror Decision Scale (JDS) validated within an experimental paradigm
- JDS consists of three subscales (complainant & defendant believability, and decision confidence)
- Good composite reliability and differential predictive validity of JDS is reported
- This empirical contribution adds to the Story Models conceptualisation of the certainty principles underlying juror decisions

Introduction and Validation of the Juror Decision Scale (JDS):
An Empirical Investigation of the Story Model

Dominic Willmott ^a, Daniel Boduszek ^{a,b}, Agata Debowska ^c & Russell Woodfield ^d

Author Note:

^a University of Huddersfield, Huddersfield, UK

^b SWPS University of Social Sciences and Humanities, Katowice, Poland

^c The University of Sheffield, Sheffield, UK

^d Leeds Trinity University, Leeds, UK

Correspondence concerning this article should be addresses to Daniel Boduszek, University of Huddersfield, Department of Psychology, Edith Key Building, Queensgate Campus, Huddersfield HD1 3DH, United Kingdom, contact email: d.boduszek@hud.ac.uk

Acknowledgements

The authors wish to thank Nigel Booth, criminal barrister at the St. John's Buildings Chambers, Manchester (UK) who ensured all features of the mock trials accorded with English law and whose legal expertise gave the experimental design a level of credibility largely unrivalled in past research. Thanks are also given to specialist sexual offence investigators within West Yorkshire and Cheshire Constabularies (UK), whose insights into rape investigations also proved invaluable, alongside the many other legal practitioners who volunteered their time to the research.

Conflict of Interest:

Authors declare that they have no conflict of interest.

Introduction and Validation of the Juror Decision Scale (JDS):
An Empirical Investigation of the Story Model

Abstract

Purpose: To develop and validate a self-report measure of individual juror decision making within criminal trials, based on theoretical features set out in the Story Model of juror decision making.

Methods: The Juror Decision Scale (JDS) and Acceptance of Modern Myths about Sexual Aggression (AMMSA) measure were completed by 324 jury-eligible participants split across 27 jury panels, after observing a rape trial re-enactment high in ecological validity. Dimensionality and construct validity of the JDS was investigated using traditional confirmatory factor analysis (CFA) techniques alongside confirmatory bifactor analysis at two time points (individual juror verdict decisions pre- and post-deliberation). Three competing models of the JDS were specified and tested using *Mplus* with maximum likelihood robust estimation.

Results: Bifactor model with three meaningful factors (complainant believability, defendant believability, decision confidence) was the best fit for the data at both decision points. Good composite reliability and differential predictive validity were observed for the three JDS subscales.

Conclusion: Alongside demonstrating its multidimensional conceptualisation, the JDS development permits future empirical testing of the Story Model theoretical assertions surrounding juror decision making. **Present findings also provide early evidence of a certainty principle assessment process governing individual verdict decision formation. Theoretical and practical applications are discussed.**

Keywords: Juror Decision Scale (JDS); Story model; Confirmatory factor analysis; Differential predictive validity; Jury decision making, **certainty principles**

Introduction and Validation of the Juror Decision Scale (JDS):

An Empirical Investigation of the Story Model

Numerous theoretical models have been advanced in an attempt to explain how jurors arrive at verdict decisions within criminal trials. Competing explanations differ in their attempt to account for individual decision formation or collective group decision-making, which constitute two distinct processing tasks jurors must undertake throughout the duration of a trial. Yet despite distinctions between juror-versus-jury level decision models, most theorising to date has centred upon individual juror processing. Dual process models such as Epstein's (1994) Cognitive-Experiential Self-Theory, alongside Tversky and Kahneman's (1974) heuristic processing shortcuts, have gained plentiful support, with empirical explorations reporting features of both models to account for many processing stages jurors undertake (Brekke & Borgida, 1988; Bornstein & Greene, 2011; Hawkins & Scherr, 2017; Kovera McAuliff, & Hebert, 1999; Krauss, Lieberman, & Olson, 2004; Lieberman, 2002; Mears & Bacon, 2009). Bayesian models have also been drawn upon, proposing a process by which individual jurors judge discrete pieces of information upon a theorised continuum of guilt. Juror weightings are posited to shift as every new piece of evidence is independently assessed, allowing an overall probability of guilt to be constructed by the end of trial (Ostrom, Werner, & Saks, 1978). Nonetheless, despite aforementioned explanations accounting for many processes thought to underlie juror decisions, no theory has been so widely adopted or comprehensive in its account of juror decision formation as Pennington and Hastie's (1992) Story Model.

Attempting to provide a complete account of the decision-making process undertaken, the Story Model posits jurors to be actively engaged in a narrative construction of information surrounding a case. A combination of evidence presented during trial, existing world knowledge, and preconceived attitudes are said to be used by jurors to construct one or more possible interpretations of the event, termed stories (Pennington & Hastie, 1992). Personal inferences and pre-existing bias are considered most likely to be incorporated within the narrative interpretations jurors construct when key elements of the stories are not presented as evidence (Pennington & Hastie, 1988). Thus, trials lacking compelling evidence, including CCTV or eyewitness testimony, appear most at risk from juror bias. In essence, the theory suggests that when hearing competing accounts of the same incident during trial, typically including one version put forward by a defendant and an alternative account put forward by a complainant, individual jurors construct differing narrative interpretations of what they believe actually occurred. At the end of trial and prior to deliberation, jurors then select one such narrative as the dominant, accepted version of events, they believe to be true (Pennington & Hastie, 1992). Yet before this occurs, competing stories or narratives are thought to undergo three differing phases of processing termed; story construction, verdict representation, and story classification.

Whilst the *verdict representation* phase relates to juror's ability to identify and understand differing verdict options available and the *story classification* phase surrounds juror's determination of which verdict option best matches the story accepted (according to the perceived goodness of fit between the two), the *story construction* phase is considered most important for individual decision formation (Pennington & Hastie, 1992). Here, jurors are thought to draw primarily upon evidence presented during trial, as well as prior knowledge held around what typically occurs in similar events, in making sense of the case (Pennington & Hastie, 1993). Prior knowledge is conceptualised as factual information, alongside

assumptions and attitudes jurors bring to trial that are relevant to the issues under scrutiny. From the combination of such information, competing stories are thought to be concurrently constructed as variants of what may have happened in the case, though only one of multiple stories constructed will ultimately be selected (Pennington & Hastie, 1992).

Competing defendant and complainant stories are subsequently assessed by individual jurors according to what Pennington and Hastie (1992) term, *certainty principles*. Thus, a story constructed will only be accepted by an individual juror when considered to have adequate: *coverage* of crucial pieces of evidence integrated within an account (i.e., good fit between evidence presented and a given version of events), *coherence* regarding how *consistent* (i.e., lacks internal contradictions), *complete* (i.e., no aspects of the story are missing from the evidence available), and *plausible* (i.e., the story is credible and could possibly have happened) a story appears to be, and finally the *uniqueness* of the story, surrounding whether alternative equally credible and comprehensive explanations could emerge from the evidence available. Pennington and Hastie (1992, 1993) posit only upon satisfying each of these certainty principle elements within the story construction stage, will any story be accepted by an individual juror, over other competing possibilities. Once one story is accepted and matched to a verdict option available, a verdict decision will be made. Taken together, the Story Model considers individual juror decision formation is best conceptualised as representing two core factors surrounding, belief in a defendant's story and belief in a complainant's story, distinct factors thought to be independently ascertained through certainty principle assessments. Consideration of theoretical discussion surrounding the role of confidence in jurors' story assessments and verdict classifications (Pennington & Hastie, 1993), as well as the importance attributed to confidence in decision pathways more broadly within jury literature (Hawkins & Scherr, 2017; Matthews, Hancock, & Briggs, 2004; Willmott & Sherretts, 2016), a third theorised factor of decision confidence is also conceptualised.

Empirical Support for the Story Model

Early attempts to examine jurors' mental representation of evidence offered initial support of a story construction process underpinning juror decision making. In one study, Pennington and Hastie (1986) exposed participants to a videotaped re-enactment of a murder trial and asked mock jurors to provide individual verdict decisions, before probing the decision-making process undertaken. Jurors reported constructing evidence into a story structure format in order to make sense of the evidence and described a process by which they drew more heavily on evidence that supported their accepted version of events than other evidence presented. In fact, the authors found evidence presented during mock trials that did not directly fit with the story constructed, was much less likely to be discussed by the jurors, regardless of its individual merit. Where important elements of a juror's story were not presented as evidence, the researchers found mock jurors simply made inferences based upon personal experiences and assumptions, ensuring the accepted story was deemed coherent and complete. Adopting an alternative approach, Pennington and Hastie (1988) presented mock jurors with a written summary of a case which they were required to render a verdict upon before undertaking a memory recognition test of trial evidence. Results displayed memory of trial information was best when information being recalled was consistent with a story matching the verdict decision participants had made and poorest for story inconsistent evidence. Further, in studies that varied the presentation of evidence from the traditional narrative format (where witnesses were asked questions about the event sequentially), to an item-by-item format (where witnesses were asked about discrete aspects of the case non-sequentially), results displayed presentation order not only differentially affected a juror's memory of evidence but led to different verdicts being returned in respect of the same case. The traditional narrative format was found to allow easier credibility assessments of witness testimony to be undertaken than item-by-item evidence presentation (Pennington & Hastie, 1992). The authors report that when

asked to make global judgements of the evidence (rather than item-by-item evaluations), jurors seemingly adopted a system of certainty principle processing of competing witness stories before deciding upon a chosen verdict (Pennington & Hastie, 1992). The authors made this assertion based upon qualitative responses mock jurors gave when asked to describe their decision-making process and thus more objective, quantitative analysis of the data gathered was not possible. To date, all studies report jurors' mental representations of trial evidence were underpinned by causally connected sequences of events, in which selected testimony appeared to be constructed into story formats. Whilst Pennington and Hastie's (1988) study displayed the same evidence would be considered stronger when presented in a story format, the greatest influence upon final decisions was found to be the strength of one story when compared to another (Pennington & Hastie, 1988, 1993). Contemporary studies appear to support the Story Model assertions surrounding juror's narrative construction of evidence underpinning individual decision-making (Blume, Johnson, & Paavola, 2007; Huntley & Costanzo, 2003). Ellison and Munro's (2015) qualitative examination of the role of written judicial instructions upon juror comprehension of legal guidelines also explored the process by which mock jurors reached verdict decisions. Analysis of deliberations led the legal scholars to again conclude a narrative construction of trial evidence was apparent.

However, the Story Model is not beyond criticism. Pennington and Hastie (1992) offer little explanation surrounding the process by which individual juror decisions remain stable or change during group-deliberations and provide no account of the exact verdict decision-making process undertaken by individual juror's during or post-deliberation. With much juror-level research dismissed as unrepresentative of collective agreed jury-level decisions, ultimately required within criminal trials before a verdict can be given (Darbyshire, 2011; Kapardis, 2014), the need to examine how individual juror decisions made pre-deliberation may interact with the group deliberation process remains apparent. Despite being considered crucial to the

acceptance of one witness story over another, no researchers have directly empirically tested whether certainty principle assessments underpin the decisions individual jurors make during trial. Authors have sought to substantiate the premise that jurors construct competing stories during trial, however to date no researchers have directly sought to test whether the certainty principles set out within the Story Model do in fact govern the acceptance of one story over another. Individual constructs thought to be comprised within the story construction phase have been tested in isolation; including plausibility judgements between criminal narratives (Canter, Grieve, Nicol, & Benneworth, 2003; Jackson, 1996), as well as narrative coherence and completeness assessments of guilt (Voss & Van Dyke, 2001; Yale, 2013). Further, no empirical attempt to date has established whether a juror's greater belief in a complainant or defendant's story has any significant association with the verdict decision jurors ultimately make.

Current Study

Whilst the Story Model provides a detailed conceptualisation of the decision-making stages thought to underlie a juror's decision to vote guilty or not guilty and remains the dominant explanation within the field, a lack of empirical research exists which seeks to verify important features of the theory. In particular, a central component termed *certainty principles* have, to the authors' knowledge, never been empirically tested or verified. Therefore, the aim of the current study was to develop a valid measure of individual juror decision making relative to criminal trials, directly integrating theoretical features of Pennington and Hastie's (1992) Story Model into an empirically testable scale. Accordingly, the factorial structure and construct validity of the scale developed, termed the Juror Decision Scale (JDS) (complainant believability, defendant believability, decision confidence), was tested using confirmatory factor analysis and confirmatory bi-factorial techniques pre and post juror deliberation. Composite reliability and differential predictive validity of the JDS was also investigated.

Methods

Sample

A self-selected opportunity sample of 352 participants were recruited from a large urban town in the North of England. Based upon recent census data, the town has a population of approximately 140,000 people, making it the 11th largest town in Great Britain (Office for National Statistics, 2016). Of this population, electoral polls suggest around 96,000 live within the parliamentary constituency of the town, where approximately 75,400 are aged 18 and above, meaning that such individuals are eligible to vote in government elections and thus, in principle, eligible for jury service (Electoral Calculus, 2015). Targeting prospective participants through advertising posters distributed throughout the town centre and university campus, the present sample consisted of members of the general population, as well as undergraduate and postgraduate students. All volunteers were screened prior to participation in line with English juror eligibility criteria (i.e., age, residential status, criminal history, mental health), with most of those excluded from partaking declined on the basis of age or lack of permanent residency status. Due to the non-attendance of eight participants for their allocated mock trial, the data from three entire jury panels were removed prior to analysis. The remaining sample was therefore 324 participants, distributed across 27 separate mock trials, each with 12 jurors in total. Participants ranged in age from 18 to 70 years old ($M = 24.86$, $SD = 9.34$), and comprised of 210 females (64.8%) and 114 males (35.2%). In total, 213 participants reported their ethnicity as Caucasian (65.7%), 58 reported being of South Asian descent (17.9%), and 53 as Black Afro-Caribbean (16.4%). Most participants reported having a level of education below a bachelor's degree (76.0%), with a smaller proportion reporting being educated to at least university level (24.0%). The demographic profile of study participants was overall

representative of the general population of the local region surrounding ethnicity and educational attainment. All participants gave up their time voluntarily and received only a gift token of nominal value for taking part.

Study procedure

Attempting to improve upon methodological limitations present within much jury research to date, typically utilising brief written vignette trial scenarios and lacking any group deliberation element (for a review see McCabe, Krauss & Lieberman, 2010), the present study sought to undertake mock trials in a manner exhibiting greater ecological validity. Accordingly, study procedures were designed to reflect the same sequential stages undertaken within genuine criminal jury trials. Following consultation with an expert panel of criminal justice practitioners including, an experienced Crown Prosecution Service (CPS) lawyer, a practising criminal barrister, and three senior police investigators from two differing British police forces, a rape case collectively deemed to be typical of those often brought before the courts was selected as a case most likely to elucidate the greatest understanding surrounding how jurors reach decisions during trial. All panel members independently identified the same three features as present within many contested rape cases (voluntary intoxication, lack of independent witnesses, previous acquaintanceship with the alleged perpetrator) and accordingly it was decided that these components should form the basis of the case transcript selected for experimentation. A systematic trawl of legal case databases LexisNexis and the British and Irish Legal Information Institute (BILII) were conducted adopting the following search criteria; transcripts were required to include (1) voluntary intoxication, previous acquaintanceship, and a general lack of independent witnesses; (2) the sexual offence of rape was recorded; (3) sufficient detail surrounding the alleged offence and the events preceding/following the rape incident, as well as the competing accounts put forward by the complainant and defendant; (4) largely evidentially ambiguous, meaning that roughly equal information corroborated and

contradicted both parties' accounts of what happened. As these databases store genuine trial transcript information of cases that have previously been heard at trial within the UK, the legal threshold of any case selected had been met, with regards to the evidence available being deemed significant enough to warrant a criminal trial. Cases which met the stipulated criteria were qualitatively reviewed until a total fifteen cases were shortlisted. These transcripts were then further scrutinised on the basis of including enough detail of the legal arguments put forward by the prosecution and defence, such that mock trial re-enactment would be possible.

Having selected one case that matched the aforementioned criteria, the trial transcript was subsequently reduced in length to allow a shorter mock trial scenario to be devised. A clear narrative was constructed relative to the case, whereby a summary of the undisputed facts, the complainant's version of events, the defendant's version of events, a condensed version of both the prosecution and defence questioning of both parties, and the judge's instructions, were scripted.

In order that the judge's instructions were accurately summarised from the original trial transcript, as well as ensuring all evidence was in accordance with English law, lawyers from the expert CJS panel were again consulted. Next, this scripted scenario was developed into a videotaped mock trial simulation and thus a local filmmaker, as well as professional actors were recruited for the roles of the complainant (female), defendant (male), and court clerk. Finally, experienced criminal lawyers were enlisted to take on the role of the legal professionals including the role of the judge, to present mock-juror participants with the case. Special permission was granted to record the mock trial recreation within a genuine courtroom in the North of England and the duration of the final condensed version was 25 minutes long.

Adopting an experimental design, participants were recruited to take part in one of 27 replications of the same mock trial. In an attempt to simulate the randomisation of mock jurors

into respective trials, participants were assigned at random to different mock trials listed for experimentation over the coming weeks. On the day of experimentation, participants were first asked to complete a number of psychosocial assessments within the context of a mock courtroom. Immediately after, the twelve-person jury panels were shown the 25-minute videotaped rape trial re-enactment on a large screen within the mock courtroom. In an attempt to ensure that participants were actively attending to the video and approached the decision-making task in a similar way to that of a real jury, mock jurors were informed that whilst the video was a re-enactment, the content therein related to that of a genuine rape allegation that had previously gone to trial and that all of the testimony they would hear was drawn from evidence presented within the real case. Once the trial video had concluded and mock jurors had heard all testimony and evidence in the case, participants were asked to remain in their seats prior to deliberation. Each participant then completed the JDS and recorded their preferred verdict (“*How do you find the defendant on the charge of Rape? Guilty or Not Guilty*”) allowing pre-deliberation individual juror decisions to be assessed. Jurors were reminded not to discuss the case until they were in the deliberation room.

Next, in accordance with genuine trial deliberation procedures, participants were reconvened within a separate jury deliberation room where they were asked to collectively discuss the case in an attempt to reach a unanimous verdict. An experimenter was not present in the room during deliberations. Where participants agreed upon a unanimous verdict within the allotted one-hour time frame, they were reconvened within the mock courtroom to render their verdict. Where participants were not unanimous after one-hour deliberating, an extra thirty minutes was provided in an attempt to reach a majority verdict, after which point all 27 jury panels had successfully arrived at verdict, recorded by an experimenter. Finally, jurors were again asked to complete the JDS and record a final verdict preference (*Guilty/Not Guilty*), allowing post-deliberation juror decisions to be assessed. Before doing so, jurors were

instructed that the verdict decisions they were being asked to make related to them as an individual and may therefore not necessarily reflect the collective verdict that had just been returned. Once completed, participants were debriefed and thanked for partaking. In total, each mock trial experiment lasted between 120 and 180 minutes from arrival to debriefing.

Scale Development

In designing the JDS as a measure of individual juror verdict decision-making, we sought to incorporate theoretical features termed *certainty principles* described in Pennington and Hastie's (1992) Story Model. Specifically, the model suggests competing versions of events (i.e. the complainant versus defendant stories), are independently and implicitly assessed by individual jurors according to a number of prescribed certainty principles. Item generation for the JDS relied directly upon the Story Model's theoretical conceptualisation of these certainty principles. As such, seven items pertaining to the extent to which a juror/respondent felt a complainant's story had coverage, coherence, consistency, completeness, plausibility, uniqueness, and overall believability, were devised. Seven identical items pertaining instead to the defendant's story were also included in the scale. In accordance with the story model assertions, these complainant versus defendant certainty principle items were hypothesised to constitute two separate dimensions within the scale, which, in line with the Pennington and Hastie's (1992) theory, should be highest for the individual whose story is matched to a verdict decision. Consideration of theoretical discussion surrounding the role of confidence in jurors' story assessments/verdict classifications (Pennington & Hastie, 1993), two global items pertaining to decision confidence were also included, hypothesised to comprise a separate dimension within the scale. In total, the scale developed comprised of 16-items distributed across three hypothesised dimensions (Complainant Believability, Defendant Believability, Decision Confidence) (see Table 4 for all scale items). Where the Story Model is accurate in its theoretical account of the role of certainty principles underpinning individual

juror decision formation, higher scores would be expected to be found on the complainant believability sub-scale for jurors who returned a guilty verdict, and higher on the defendant believability sub-scale where jurors returned a not guilty verdict. Therefore, all JDS scale items measure respondents' self-reported assessments of how believable they determine a complainant and defendant to be, having heard all evidence in a particular jury trial (or mock trial for research purposes), as well as their self-reported confidence relating to the individual verdict decision made in a given case.

Measures

The Juror Decision Scale (JDS) is a 16-item self-report measure designed to assess individual juror decision making (see Table 4 for all scale items). The measure consists of three subscales; Complainant Believability (seven items), Defendant Believability (seven items), and Decision Confidence (two items), with all items scored on a 5-point Likert scale (1 = "not at all" to 5 = "extremely"). Higher scores on the Complainant Believability sub-scale indicates greater juror/respondent belief in the complainant's story, with lower scores indicating reduced belief in a complaints account. Higher scores on the Defendant Believability sub-scale indicates greater juror/respondent belief in the defendant's story, with lower scores indicating a reduced belief. Higher scores on the Decision Confidence subscale indicates greater juror/respondent confidence in the accuracy of the verdict decision given. All scale items are measured according to the individual juror/respondents' decisions relative to the evidence heard within a particular case or trial. The JDS should be administered pre and post group deliberation.

Acceptance of Modern Myths about Sexual Aggression (AMMSA; Gerger, Kley, Bohner, & Siebler, 2007) is a 30-item unidimensional self-report inventory developed to measure attitudes held towards rape and sexual aggression in diverse populations (e.g. item 9 "*If a woman invites a man to her home for a cup of coffee after a night out this means that she*

wants to have sex” and item 27 “Many women tend to misinterpret a well-meant gesture as a sexual assault”. Responses are measured on a seven-point Likert scale (1 = “completely disagree” to 7 = “completely agree”). Total scores range from 30 to 210, with higher scores indicating greater acceptance of myths surrounding sexual aggression (Cronbach’s alpha = 0.92).

Analytical Procedure

To investigate the dimensionality and construct validity of the JDS, traditional confirmatory factor analysis (CFA) techniques and confirmatory bifactor analysis procedures (Reise, Moore, & Haviland, 2010) were undertaken at both verdict decision time points (VD1 – pre-deliberation and VD2 – post-deliberation). Three alternative models of the JDS were specified and assessed using *Mplus 7.4* (Muthen & Muthen, 2015) with maximum likelihood robust (MLR) estimation. The CFA was used to determine factor loadings and identify the best factorial structure.

At both verdict decision time points, Model 1 is a one-factor solution, where all 16 JDS items load onto a single latent factor. Model 2 is a correlated three-factor solution, where items load on the complainant believability factor (items 2, 3, 4, 5, 6, 7 and 8), defendant believability factor (items 9, 10, 11, 12, 13, 14 and 15), and decision confidence factor (items 1 and 16). Model 3 (see Figure 1) is a bifactor conceptualisation with one general factor (all items) of juror decision making, alongside three subordinate factors as described in Model 2.

Insert Figure 1 about here

The overall fit of each model and the relative fit between the three differing models were assessed using a range of goodness-of-fit indices. The Chi-square statistic (χ^2), Comparative Fit Index (CFI; Cronbach, 1990), Tucker Lewis Index (TLI; Tucker & Lewis,

1973), Root-Mean-Square Error of Approximation (RMSEA; Steiger, 1990) with the associated 90% confidence interval (90% CI), Standardised Root Mean Square Residual (SRMR), and Bayesian Information Criterion (BIC; Schwarz, 1978) were reported for all models. For CFI and TLI, values above or approaching 0.95 are indicative of good model fit and above 0.90 acceptable model fit (Bentler, 1990; Hu & Bentler, 1999). Likewise, for RMSEA and SRMR, values less than 0.05 suggest good model fit and below .08 acceptable model fit (Bentler, 1990; Hu & Bentler, 1999). For BIC values comparing alternate models, the lowest value is indicative of the best fitting model (Nylund, Asparouhov, & Muthen, 2007).

Finally, due to criticisms surrounding Cronbach's alpha coefficient indicators of internal consistency (Raykov, 1997; 1998), composite reliability was used within the present analysis to assess internal reliability of the JDS factors, with values above 0.60 typically considered acceptable (Diamantopoulos & Siguaw, 2000).

Results

Frequencies of individual juror verdict decisions pre- and post-deliberation and collective group decisions are presented in Table 1. Whilst the number of participants who recorded a guilty verdict decreased post-deliberation with the number of individual not guilty verdict preferences increasing, results of a McNemar's Chi-square test for association displayed no significant change occurred between pre- and post-deliberation, $\chi^2(1, N = 324) = 2.16, p = .142$. Overall, most trials resulted in a collective not guilty verdict returned ($N = 22$) with just a minority of jury panels returning a guilty verdict ($N = 5$).

Descriptive statistics for the three JDS factors (Complainant Believability, Defendant Believability, and Decision Confidence) at both verdict decision time points (VD1 = pre-deliberation; VD2 = post-deliberation), are presented in Table 2.

Insert table 1 about here

Insert table 2 about here

Fit indices for three alternative models of the JDS at both verdict decision time points (VD1, VD2) are presented in Table 3. At both time points, the one-factor model and correlated three-factor model were rejected, based upon exhibiting CFI and TLI values considerably below the 0.95 approximate level of acceptance (Bentler, 1990; Hu & Bentler, 1999) and RMSEA and SRMR values considerably above the 0.05 level of acceptance. Taken together the combination of fit statistics indicate the bifactor model of the JDS provides the best fit to the data at both verdict decision time points: VD1 (CFI = 0.94, TLI = 0.92, SRMR = 0.04, RMSEA = 0.07 [90%CI = 0.05/0.08], BIC = 11119.99), VD2 CFI = 0.96, TLI = 0.94, SRMR = 0.04, RMSEA = 0.07 [90%CI = 0.05/0.08], BIC = 10700.38). Notably, the BIC statistic for the bifactor model, at both verdict decision time points, was lower than that displayed for all alternative models.

Insert Table 3 about here

The appropriateness of the bifactor model of the JDS can also be determined through examination of factor loadings for statistical significance. Inspection of the factor loadings for the three JDS factors (Table 4 and 5) provides clear evidence of the appropriateness of including these latent factors in the scoring of the JDS. Overall, standardized factor loadings are higher for three grouping factors than for the general factor. Therefore, all three JDS subscales (complainant believability, defendant believability, and decision confidence) should be considered in research and practical application.

Insert Table 4 about here

Insert Table 5 about here

The correlations between the three JDS factors, relative to Verdict Decision 1, were low (complainant believability and decision confidence $r = 0.10$, $p > 0.05$; complainant believability and defendant believability $r = -0.30$, $p < 0.001$; defendant believability and decision confidence $r = 0.14$, $p < 0.05$) indicating little overlap between the variables. Correlations between the JDS latent factors relative to Verdict Decision 2 were also low (complainant believability and decision confidence $r = 0.05$, $p > 0.05$; complainant believability and defendant believability $r = -0.36$, $p < 0.001$; defendant believability and decision confidence $r = 0.25$, $p < 0.001$) further indicating little overlap between the variables. Nonetheless, whilst there appears to be no significant overlap between JDS variables at either decision points, assessing differential predictive validity of a multidimensional scale is recommended (Boduszek & Debowska, 2016; Carmines & Zeller, 1979). In the present analysis, this involved ensuring that the three dimensions of the JDS were associated differentially with external variables.

Table 6 displays the results of the regression analyses at both decision time points. In relation to Verdict Decision 1, complainant believability forms a significant negative relationship with rape attitudes (AMMSA), whereas a significant positive relationship is observed between defendant believability and AMMSA scores. While negatively correlated, decision confidence was non-significantly related to AMMSA scores. For Verdict Decision 2, AMMSA was again significantly negatively correlated with the complainant believability dimension and significantly positively correlated with defendant believability. Though positively correlated, decision confidence was non-significantly related to AMMSA scores. Both pre- and post-deliberation logistic regression results display a significant positive relationship between complainant believability and guilty verdict selections, whereas a significant negative relationship is observed between defendant believability and guilty verdict

selections. Decision confidence was positively associated with guilty verdicts at both decision points, though this relationship was statistically non-significant. These differential correlations between JDS subscales and external variables confirm the correctness of multidimensional solution of JDS.

Insert Table 6 about here

Internal reliability of the JDS factors was calculated using composite reliability in place of traditional Cronbach's alpha (as suggested by Boduszek & Debowska, 2016; Raykov, 1997). Using the formula displayed below where; CR = composite reliability of the factor score, λ_i = standardised factor loading, and $\text{Var}(\epsilon_j)$ = standard error variance, results demonstrate good internal reliability for the JDS factors pre- and post-deliberation.

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \text{Var}(\epsilon_j)}$$

At Verdict Decision 1, results displayed that confidence in decision = 0.82, complainant believability = 0.70, and defendant believability = 0.79, alongside the general factor = 0.74, exhibited good internal reliability. Likewise, at Verdict Decision 2, confidence in decision = 0.83, complainant believability = 0.72, and defendant believability = 0.85, as well as the general factor = 0.79, exhibited good internal reliability.

Discussion

Pennington and Hastie's (1992) Story Model provides a detailed conceptualisation of the information processing stages thought to underlie juror verdict decisions. Yet whilst

credited for its comprehensiveness and widely regarded as the dominant explanation of juror-level decision making, researchers are yet to empirically test and verify important theoretical features underlying the model. One central feature suggests whilst hearing competing defendant and complainant accounts during trial, jurors assess the extent to which they believe such stories according to a subscribed set of certainty principles. Based upon the varying extent to which each story is considered to be consistent, complete, and plausible (amongst other certainty principles), the theory posits competing stories are rated in terms of overall believability, with the account deemed to be most believable, used to construct the individual juror's chosen verdict decision. Yet with no prior inventory in existence, the need to develop a self-report scale directly integrating the theoretical certainty principles into an empirically testable measure remained apparent. The main objective of the current exploration was therefore to develop a valid and reliable scale permitting the Story Model's conceptualisation of individual juror decision formation (certainty principles) to be examined. Another objective was to evaluate the dimensionality and construct validity of the proposed Juror Decision Scale (JDS) using confirmatory techniques. Specifically, confirmatory factor analysis (CFA) and confirmatory bi-factor analysis was undertaken upon a large community sample of mock-jurors following their exposure to a simulated rape trial and completion of the JDS pre and post-deliberation.

It has previously been suggested that in order to fully explore the factorial structure of a proposed measure, a number of alternate conceptually sound solutions should be tested (Boduszek & Debowska, 2016; Reise, Moore, & Haviland, 2010). In the current study three alternative models of the JDS were identified and tested at both pre- and post-deliberation verdict decision time points (a one-factor model, a three-factor model, and a bifactor model with three grouping factors), using confirmatory techniques. Results displayed that the only acceptable solution for the 16-item JDS at both verdict decision points (as indicated by all

model fit statistics) was the bifactor model with three grouping factors (Complainant Believability, Defendant Believability, and Decision Confidence), while controlling for a general factor. Since the majority of covariation between the observed indicators were explained by the three grouping factors, at both decision points, these factors formed the basis for creating the instrument's subscales (see Reise, Moore, & Haviland, 2010). According to Boduszek and Debowska (2016), when compared with traditional CFA procedures, bifactor modelling allows the validity of a single factor to be assessed alongside incorporating elements of construct multi-dimensionality. Adopting this approach subsequently elucidated the JDS as a multi-dimensional concept.

Further, whilst the three JDS factors displayed little overlap with one another, the need to establish differential predictive validity between sub-scales on a multidimensional scale is considered advantageous (Carmines & Zeller, 1979). Ensuring sub-scales measure separate theoretical, as opposed to statistical, factors by establishing differential predictive validity thereby allows conceptual distinctiveness to be reliably ascertained (Boduszek & Debowska, 2016). Indeed, the present results displayed that across both verdict decision time points, the three JDS factors correlated differently with external measures. For example, complainant believability was significantly negatively associated with rape myth acceptance (as measured using the AMMSA; Gerger et al., 2007) both pre-deliberation ($\beta = -0.16$) and post-deliberation ($\beta = -0.24$). Conversely, defendant believability significantly positively associated with AMMSA scores, both upon pre-deliberation ($\beta = 0.23$) and post-deliberation ($\beta = 0.16$) verdict decisions. Such relationships display the important role that pre-trial bias appears to have upon juror decision making, with rape attitudes shown to be directly associated with juror beliefs in a *defendant's* account of an alleged rape, though unsurprisingly, not with that of the *complainant*. Such findings directly support those reported in prior research in that, greater acceptance of sexually aggressive myths appears to reduce a juror's propensity to believe a

rape complainant's testimony (Dinos, Burrowes, Hammond, & Cunliffe, 2015; Ellison & Munro, 2010; Finch & Munro, 2005; Pollard, 1992; Raitt & Zeedyk, 1997; Temkin & Krahe, 2008; Whatley, 1996). Relative to mock jurors' individual verdict decisions both prior to (OR = 1.62) and following group deliberation (OR = 1.45), complainant believability was also shown to be significantly positively associated with guilty verdict selections, whereas defendant believability was instead, significantly negatively associated with guilty verdict decisions pre- (OR = 0.68) and post-deliberation (OR = 0.78). Particularly interestingly, this indicates that the greater a juror's belief in a rape complainant's testimony, the more likely it is that a guilty verdict will be returned. Yet contrastingly, where a juror exhibits greater belief in the defendant's story, a not guilty verdict is more probable. The totality of such findings thereby provides early support for Pennington and Hastie's (1992) certainty principle assertions in that, heightened scores on the complainant believability subscale not only appear to co-exist with reduced scores in defendant believability, but when measured in association with verdict preferences, appears to be directly predictive of the verdict decision that jurors ultimately make. As such, results support Pennington and Hastie's (1992) assertions that prior to selecting a verdict, jurors appear to assess competing witness accounts in terms of a subscribed set of certainty principles, in order to determine which story they deem most believable, before voting accordingly. Shown to be significantly associated in opposing directions with rape myth acceptance and verdict decisions, assessments made according to certainty principle items included in the JDS (surrounding a stories completeness, plausibility, coherence amongst others), may not only be drawn upon to decide which story will ultimately be selected (Pennington & Hastie, 1992) but appear to be influenced in themselves by prior attitudes jurors held. Alternatively put, whilst the certainty principle items comprised with the JDS are clearly important determinants upon individual decision formation as first theorised, these assessments appear to be influenced in themselves by preconceived attitudes jurors hold.

The present study ought to be considered in light of some limitations. Most pertinent is the use of self-report measurement that the JDS relies upon in its assessment of juror decisions and as such, is associated with possible response bias (see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Additionally, the current study procedure relied upon validating the JDS within the context of an English mock rape trial utilising a sample consisting of both community and student mock jurors. With debate surrounding the generalisability of student samples within jury research and studies reporting differences in both attitudes and cognitive processing styles in student as opposed to community samples (for competing reviews see Keller & Weiner, 2011 and Bornstein et al., 2017), future explorations should see researchers seek to replicate the validation of the JDS within a more representative sample. Thus, to ensure the scale's utility as an accurate assessment of certainty principle processing applicable to genuine juror decision making, future samples should be drawn from electoral voting and driver registration registers adopting the same process in which trial jurors are drawn.

A number of practical implications emerge from the present research and development of the JDS. Firstly, results display greater belief in a complainant's version of events to be directly associated with the juror's propensity to return a guilty verdict and contrastingly, greater belief in the defendant's account, associated with juror's reluctance to return a guilty verdict. Whilst it is perhaps routinely taken for granted by justice systems around the world that jurors return verdicts which match true and accurate interpretations of the evidence, questions around juror comprehension of legal instructions and malicious or biased decision making has long brought this assumption into question (Dhimi, Lundrigan & Mueller-Johnson, 2015; Ellison & Munro, 2010; Ellsworth & Reifman, 2000; Semmler & Brewer, 2002). Though specific instruction comprehension was not directly tested here, the present findings offer justice systems and trial judges at least some minor reassurance that individual juror decisions were related and matched to juror interpretations of the facts. Whilst further

systematic research clearly remains necessary to examine the role of prejudice upon verdict decision making as well as instruction comprehension, it appears that jurors do not routinely make verdict selections that do not match the evidence as they interpret it. The development of the JDS also provides an opportunity for criminal justice practitioners, in particular trial lawyers, to utilise the measure as a tool for examining likely juror interpretations of particular evidence pre-trial. Within a North American trial consultancy context where evidence can be shown to mock jurors prior to the genuine case being presented in court, the JDS provides a more reliable empirically grounded measure of individual juror decision making than traditionally less scientific approaches (see Oostinga & Willmott, 2017; Seltzer, 2006). Within future research, additional mock trial scenarios varying by crime type and the judicial procedures of that country should also be tested to examine the validity of the JDS more broadly. Where legislation permits, future explorations should also see researchers seek to revalidate the JDS beyond experimental conditions with genuine juror respondents pre and post juror-deliberation, particularly within a North American context where jurors can discuss cases post-trial.

Overall, current findings provide an empirical contribution to an almost exclusive theoretical literature surrounding the Story Model's conceptualisation of the certainty principles. Whilst several studies have sought to substantiate claims that jurors construct competing narratives during trial and others have examined the importance of particular story features in isolation including plausibility, coherence, and completeness upon an assessors determination of credibility and guilt (Campbell, Menaker, & King, 2015; Canter, Grieve, Nicol, & Benneworth, 2003; Hine & Murphy, 2017; Jackson, 1996; Voss & Van Dyke, 2001; Yale, 2013), no researchers to date have developed and validated a complete scale which permits comprehensive testing of such an assertion. With existing juror bias scales adequately testing the importance of legal attitudes upon verdict decisions (Kassin & Wrightsman, 1983;

Lecci & Myers, 2008; Lundrigan, Dhimi, & Mueller-Johnson, 2016), the need for an empirical test of the decision making process itself remained apparent. As such, the development of the Juror Decision Scale alongside demonstrating its validity and multidimensional conceptualisation, permits future empirical testing of the Story Model theoretical assertions surrounding juror decision making and provides early evidence of a certainty principle assessment process governing individual juror verdict decision formation.

References

- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *217*, 238–246.
- Blume, J. H., Johnson, S. L., & Paavola, E. C. (2007). Every Juror Wants a Story: Narrative Relevance, Third Party Guilt and the Right to Present a Defense. *American Criminal Law Review*, *44*, 1069-1113.
- Boduszek, D., & Debowska, A. (2016). Critical evaluation of psychopathy measurement (PCL-R and SRP-III/SF) and recommendations for future research. *Journal of Criminal Justice*, *44*, 1–12.
- Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and human behavior*, *41*(1), 13.
- Bornstein, B. H., & Greene, E. (2011). Jury decision making: Implications for and from psychology. *Current Directions in Psychological Science*, *20*(1), 63-67.
- Brekke, N., & Borgida, E. (1988). Expert psychological testimony in rape trials: A social-cognitive analysis. *Journal of Personality and Social Psychology*, *21*, 372-384.
- Campbell, B. A., Menaker, T. A., & King, W. R. (2015). The determination of victim credibility by adult and juvenile sexual assault investigators. *Journal of Criminal Justice*, *43*(1), 29-39.

Canter, D. V., Grieve, N., Nicol, C., & Benneworth, K. (2003). Narrative plausibility: The impact of sequence and anchoring. *Behavioral Sciences & the Law*, 21(2), 251-267.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage Publications.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York, NY: Harper & Row.

Darbyshire, P. (2011). *Darbyshire on the English Legal System* (10th Ed.) London: Sweet and Maxwell.

Dhami, M. K., Lundrigan, S., & Mueller-Johnson, K. (2015). Instructions on reasonable doubt: Defining the standard of proof and the juror's task. *Psychology, Public Policy, and Law*, 21(2), 169.

Diamantopoulos, A., & Sigauw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. London: Sage.

Dinos, S., Burrowes, N., Hammond, K., & Cunliffe, C. (2015). A systematic review of juries' assessment of rape victims: Do rape myths impact on juror decision-making? *International Journal of Law, Crime and Justice*, 43(1), 36-49.

Electoral Calculus. (2015). *Huddersfield: Seat, Ward and Prediction Details*. Retrieved from <http://www.electoralcalculus.co.uk/cgi-bin/seatdetails.py?seat=Huddersfield>

Ellison, L., & Munro, V. E. (2015). 'Telling tales': exploring narratives of life and law within the (mock) jury room. *Legal studies*, 35(2), 201-225. doi: 10.1111/lest.12051

Ellison L, & Munro, V. (2010). Getting to (Not) Guilty: Examining Jurors Deliberative Processes in and Beyond the Context of a Mock Rape Trial. *Legal Studies*, 30(1) 74-97.

Ellsworth, P. C., & Reifman, A. (2000). Juror comprehension and public policy: Perceived problems and proposed solutions. *Psychology, Public Policy, and Law*, 6(3), 788.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, 49, 709-724.

Finch, E., & Munro, V. E. (2005). Juror Stereotypes and Blame Attribution in Rape Cases Involving Intoxicants the Findings of a Pilot Study. *British Journal of Criminology*, 45(1), 25-38.

- Gerger, H., Kley, H., Bohner, G., & Siebler, F. (2007). The acceptance of modern myths about sexual aggression scale: Development and validation in German and English. *Aggressive Behavior, 33*(5), 422-440.
- Hawkins, I., & Scherr, K. (2017). Engaging the CSI effect: The influences of experience-taking, type of evidence, and viewing frequency on juror decision-making. *Journal of Criminal Justice, 49*, 45-52.
- Hine, B., & Murphy, A. (2017). The impact of victim-perpetrator relationship, reputation and initial point of resistance on officers' responsibility and authenticity ratings towards hypothetical rape cases. *Journal of Criminal Justice, 49*, 1-13.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1-55.
- Huntley, J. E., & Costanzo, M. (2003). Sexual harassment stories: testing a story-mediated model of juror decision-making in civil litigation. *Law and human behavior, 27*(1), 29-51.
- Jackson, B. S. (1996). 'Anchored narratives' and the interface of law, psychology and semiotics. *Legal and Criminological Psychology, 1*(1), 17-45.
- Kapardis, A. (2014). *Psychology and Law: A Critical Introduction* (4th Ed.). Cambridge: Cambridge University Press.
- Kassin, S. M., & Wrightsman, L. S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality, 17*(4), 423-442.
- Keller, S. R., & Wiener, R. L. (2011). What are we studying? Student jurors, community jurors, and construct validity. *Behavioral sciences & the law, 29*(3), 376-394.
- Kovera, M. B., McAuliff, B. D., & Hebert, K. S. (1999). Reasoning about scientific evidence: Effects of juror gender and evidence quality on juror decisions in a hostile work environment case. *Journal of Applied Psychology, 84*(3), 362-348.
- Krauss, D. A., Lieberman, J. D., & Olson, J. (2004). The effects of rational and experiential information processing of expert testimony in death penalty cases. *Behavioral sciences & the law, 22*(6), 801-822.

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*(8), 2010-2038.

Lieberman, J. D. (2002). Head Over the Heart or Heart Over the Head? Cognitive Experiential Self-Theory and Extra-legal Heuristics in Juror Decision Making. *Journal of Applied Social Psychology, 32*(12), 2526-2553.

Lundrigan, S., Dhimi, M. K., & Mueller-Johnson, K. (2016). Predicting verdicts using pre-trial attitudes and standard of proof. *Legal and Criminological Psychology, 21*(1), 95-110.

Matthews, R., Hancock, L., & Briggs, D. (2004). *Jurors' perceptions, understanding, confidence and satisfaction in the jury system: A study in six courts*. Research Development and Statistics Directorate, Home Office.

McCabe, J. G., Krauss, D. A., & Lieberman, J. D. (2010). Reality check: A comparison of college students and a community sample of mock jurors in a simulated sexual violent predator civil commitment. *Behavioral Sciences & the Law, 28*(6), 730-750.

Mears, D. P., & Bacon, S. (2009). Improving criminal justice through better decision making: Lessons from the medical system. *Journal of Criminal Justice, 37*(2), 142-154.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modelling, 14*(4), 535-569.

Office of National Statistics. (2016). *Criminal Justice Statistics 2015 England and Wales: Statistics Bulletin*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524429/criminal-justice-statistics-quarterly-update-_annual_-2015.pdf

Oostinga, M. & Willmott, D. (2017). Scientific Jury Selection. In B. Baker, R. Minhas, & L. Wilson (Eds.), *Psychology and Law: Factbook* (2nd ed.). European Association of Psychology and Law. (ISBN 9781326989651)

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

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51(2), 242-263.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 521-538.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, 62(2), 189-224.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, 49(1), 123-163.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-893.

Pollard, P. (1992). Judgements about victims and attackers in depicted rapes: A review. *British Journal of Social Psychology*, 31(4), 307-326.

Raitt, F. E., & Zeedyk, S. M. (1997). Rape Trauma Syndrome: Its corroborative and educational roles. *Journal of Law and Society*, 24(4), 552-568.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, 22, 375-385.

doi:10.1177/014662169802200407

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544-559.

Seltzer, R. (2006). Scientific jury selection: Does it work?. *Journal of Applied Social Psychology*, 36(10), 2417-2435.

Semmler, C., & Brewer, N. (2002). Using a flow-chart to improve comprehension of jury instructions. *Psychiatry, psychology and law*, 9(2), 262-270.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.

Temkin, J., & Krahe, B. (2008). *Sexual assault and the justice gap: A question of attitude*. Oxford: Hart Publishing.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Voss, J. F., & Van Dyke, J. A. (2001). Narrative structure, information certainty, emotional content, and gender as factors in a pseudo jury decision-making task. *Discourse processes*, 32(2), 215-243.

Whatley, M. A. (1996). Victim characteristics influencing attributions of responsibility to rape victims: A meta-analysis. *Aggression and Violent Behavior*, 1(2), 81-95.

Willmott, D., & Sherretts, N. (2016). Individual differences in eyewitness identification accuracy between sequential and simultaneous line-ups: consequences for police practice and jury decisions. *Current Issues in Personality Psychology*, 4(4), 228-239.

Yale, R. N. (2013). Measuring Narrative Believability: Development and Validation of the Narrative Believability Scale (NBS-12). *Journal of Communication*, 63(3), 578-599.

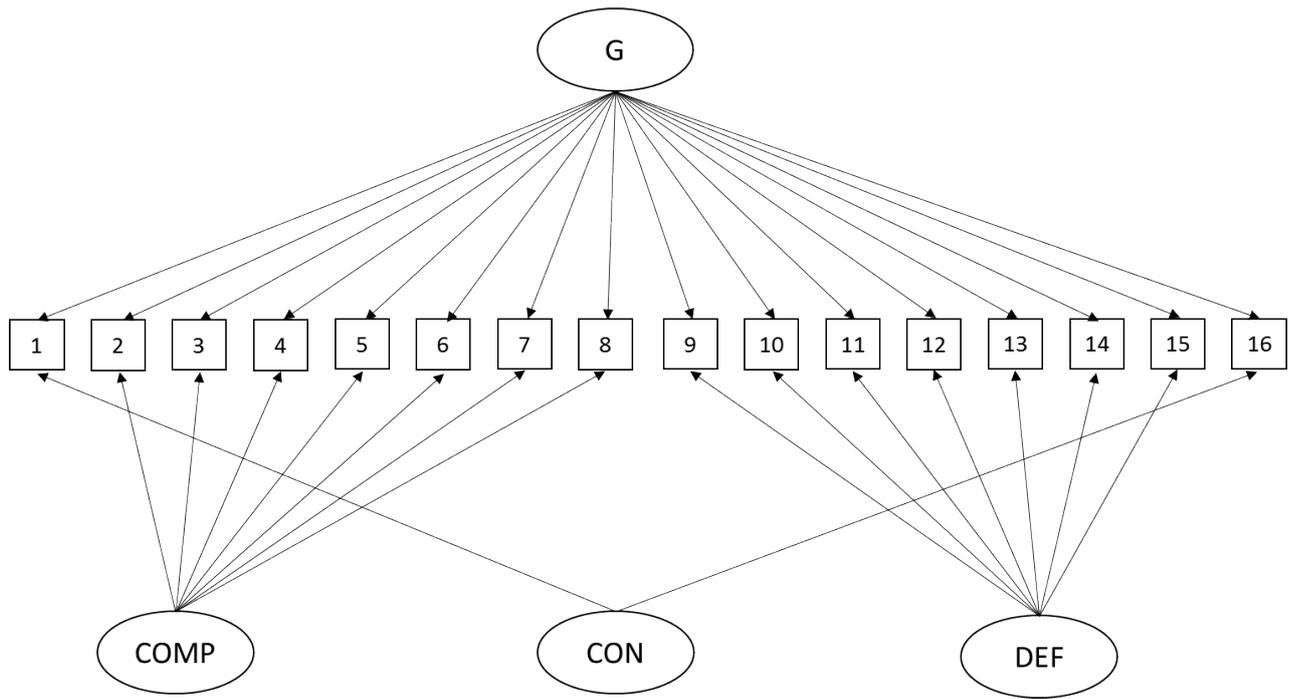


Figure 1. Bifactor model of the JDS; G = general factor of JDM (items 1-16); COMP = Complainant Believability (items 2-8); DEF = Defendant Believability (items 9-15); CON = Decision Confidence (items 1 & 16)

Table 1*Individual and Collective Verdict Decision outcomes (n =324)*

Decision	Guilty N (%)	Not Guilty N (%)
Group Verdict	5 (18.5%)	22 (71.5%)
Individual VD1	145 (44.8%)	179 (55.2%)
Individual VD2	133 (41.0%)	191 (59.0%)

Note: Group Verdict = collective jury panel decision; VD1 = Individual Verdict decision 1 (pre-deliberation); VD2 = Individual Verdict decision 2 (post-deliberation).

Table 2

Descriptive Statistics for JDS factors pre-deliberation (verdict decision 1), post-deliberation (verdict decision 2), and AMMSA.

Variables	<i>M</i>	<i>SD</i>	<i>Mdn</i>	Observed Min.	Observed Max.
JDS VD1					
Decision Confidence	7.09	1.44	7.00	2	10
Complainant Believability	22.32	4.72	22.00	10	33
Defendant Believability	22.71	4.29	23.00	8	35
AMMSA	93.70	25.74	91.00	37	161
JDS VD2					
Decision Confidence	7.57	1.60	8.00	2	10
Compliant Believability	21.27	5.12	21.00	7	35
Defendant Believability	23.03	4.83	23.00	10	35

Note: JDS = Juror Decision Scale; VD1 = Individual Verdict decision 1 (that participants made pre-deliberation); VD2 = Individual Verdict decision 2 (that participants made post-deliberation); AMMSA = Acceptance of Modern Myths about Sexual Aggression total score.

Table 3

Fit Indices for Three Alternative Models of the JDS, during stage VD1 (pre-deliberation) and stage VD2 (post-deliberation).

Stage	Models	χ^2	<i>df</i>	CFI	TLI	RMSEA	90% CI	SRMR	BIC
VD1	1. One-factor	1149.72*	104	.49	.41	.16	.17/.19	.16	12034.38
	2. Correlated 3 factors	813.79*	101	.90	.89	.10	.08/.12	.07	11201.76
	3. Bifactor	204.42*	85	.94	.92	.07	.05/.08	.04	11119.99
VD2	1. One-factor	1606.68*	104	.52	.45	.21	.20/.22	.17	11980.06
	2. Correlated 3 factors	353.86*	101	.90	.89	.09	.08/.10	.07	10785.36
	3. Bifactor	199.60*	85	.96	.94	.07	.05/.08	.04	10700.38

Note. JDS = Juror Decision Scale; CFI = Comparative Fit Index; CI = Confidence Interval; *df* = degrees of freedom; RMSEA = Root-Mean-Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker Lewis Index; BIC = Bayesian Information Criteria; χ^2 = chi square goodness of fit statistic. * Indicates χ^2 are statistically significant ($p < .05$).

Table 4

Standardized Factor Loadings for the Three JDS Factors and General Factor (G) pre-deliberation (VD1).

MCSI-R items	G	CONF	COMP	DEF
1. Thinking about your individual verdict decision of ‘guilty’ or ‘not guilty’, how confident are you that you have made the correct decision?	.14*	.82***		
2. How well did the evidence match and cover what the complainant said happened?	.34***		.42***	
3. How complete was the complainant’s story in the sense that no aspects were missing or left unsupported by the evidence?	.35***		.54***	
4. How plausible was the complainant’s version of events, in that you think what they said happened, is both possible and likely?	.64***		.43***	
5. How coherent was the complainant’s story, meaning that the different stages described as happening were logically connected?	.49***		.53***	
6. How unique was the complainant’s account, in that you feel it was the only possible explanation of the evidence heard?	.34***		.61***	
7. How consistent was the complainant’s version of events with the evidence presented overall?	.36***		.51***	
8. Overall, how much do you believe the complainant’s version of events?	.76***		.43***	
9. How well did the evidence match and cover what the defendant said happened?	.08			.61***
10. How complete was the defendant’s story in the sense that no aspects were missing or left unsupported by the evidence?	.03			.73***
11. How plausible was the defendant’s version of events, in that you think what they said happened, is both possible and likely?	.57***			.55***
12. How coherent was the defendant’s story, meaning that the different stages described as happening were logically connected?	.25*			.75***
13. How unique was the defendant’s account, in that you feel it was the only possible explanation of the evidence heard?	.54***			.39***
14. How consistent was the defendant’s version of events with the evidence presented overall?	.38**			.63***
15. Overall, how much do you believe the defendant’s version of events?	.74***			.47***
16. Finally, how confident are you overall that you have reached the correct verdict decision in this case?	.01	.84***		

Note. Factor loadings are statistically significant at * $p < .05$; ** $p < .01$; *** $p < .001$. CONF = Decision Confidence; COMP = Complainant Believability; DEF = Defendant Believability.

Table 5

Standardized Factor Loadings for the Three JDS Factors and General Factor (G) post-deliberation (VD2).

MCSI-R items	G	CONF	COMP	DEF
1. Thinking about your individual verdict decision of ‘guilty’ or ‘not guilty’, how confident are you that you have made the correct decision?	.11	.76***		
2. How well did the evidence match and cover what the complainant said happened?	.39***		.62***	
3. How complete was the complainant’s story in the sense that no aspects were missing or left unsupported by the evidence?	.29**		.79***	
4. How plausible was the complainant’s version of events, in that you think what they said happened, is both possible and likely?	.76***		.37***	
5. How coherent was the complainant’s story, meaning that the different stages described as happening were logically connected?	.41***		.51***	
6. How unique was the complainant’s account, in that you feel it was the only possible explanation of the evidence heard?	.47***		.57***	
7. How consistent was the complainant’s version of events with the evidence presented overall?	.63***		.49***	
8. Overall, how much do you believe the complainant’s version of events?	.83***		.35***	
9. How well did the evidence match and cover what the defendant said happened?	.30***			.67***
10. How complete was the defendant’s story in the sense that no aspects were missing or left unsupported by the evidence?	.24***			.72***
11. How plausible was the defendant’s version of events, in that you think what they said happened, is both possible and likely?	.48***			.66***
12. How coherent was the defendant’s story, meaning that the different stages described as happening were logically connected?	.17*			.76***
13. How unique was the defendant’s account, in that you feel it was the only possible explanation of the evidence heard?	.41***			.53***
14. How consistent was the defendant’s version of events with the evidence presented overall?	.32***			.72***
15. Overall, how much do you believe the defendant’s version of events?	.63***			.62***
16. Finally, how confident are you overall that you have reached the correct verdict decision in this case?	.12	.92***		

Note. Factor loadings are statistically significant at * $p < .05$; ** $p < .01$; *** $p < .001$. CONF = Decision Confidence; COMP = Complainant Believability; DEF = Defendant Believability.

Table 6*Associations between the three JDS Factors and External Variables.*

Variable	Verdict Decision 1		Verdict Decision 2	
	Guilty Verdict ($\chi^2 = 236.50, p < .05$) OR (95% CI)	AMMSA (F [3, 319] = 11.61, $p < .001$) β (95% CI)	Guilty Verdict ($\chi^2 = 190.10, p < .001$) OR (95% CI)	AMMSA (F [3, 319] = 12.76, $p < .001$) β (95% CI)
Decision Confidence	1.13 (.88/1.45)	-0.03 (-.13/.08)	1.03 (.84/1.27)	0.01 (-.11/.11)
Complainant Believability	1.62*** (1.45/1.81)	-0.16** (-.27/-.05)	1.45*** (1.32/1.59)	-0.24*** (-.35/-.12)
Defendant Believability	0.68*** (.60/.76)	0.23*** (.11/.34)	0.78*** (.71/.85)	0.16** (.04/.27)

Note: Verdict Decision 1 = Individual Verdict decision 1 (made pre-deliberation); Verdict Decision 2 = Individual Verdict decision 2 (made post-deliberation); Guilty Verdict = Individual juror guilty verdict selections; AMMSA = Acceptance of Modern Myths about Sexual Aggression total score; ** $p < .01$, *** $p < .001$

