Asking the Right Questions: Examining the Efficacy of Question Trails as a Method of Improving Lay Comprehension and Application of Legal Concepts

Benjamin Spivak^a, James R. P. Ogloff^a and Jonathan Clough^b

^a*Centre for Forensic Behavioural Science, Swinburne University of Technology and Forensicare, Melbourne, Australia;* ^b*Faculty of Law, Monash University, Melbourne, Australia*

The present study examines the ‘fact based’ approach to jury instructions, which embeds legal concepts in a series of logically ordered written factual questions that the jury must answer to reach a verdict. The study utilised a sample of 1007 adults called for jury service in Victoria, Australia. Four instructional types (standard, plain language, checklist, fact based) were compared on paraphrase and application measures across three time points. Results indicated that paraphrase performance was significantly lower for standard instructions compared to all other instructional types at the pre-deliberation stage. Findings around application of law were mixed. At the pre-deliberation stage, participants receiving fact based instructions had significantly higher scores on true/false application questions compared with participants in other conditions, whereas there were no significant differences between conditions for multiple-choice application. However, testing following deliberation revealed that participants in the fact-based condition had significantly higher scores on multiple-choice application items.

Keywords: checklists; comprehension; decision-making; instruction; jury; question trails.

Over recent decades, an increasing number of judges, law reform bodies, legal scholars and researchers have raised concerns regarding the ability of jurors to understand and apply jury instructions in criminal trials (Dann, 1993; Lieberman & Sales, 1997; Neave, 2012; New South Wales Law Reform Commission, 2011; Thomas, 2010; Victorian Law Reform Commission, 2009; Weinberg, 2012). This has led a number of researchers to devise and test methods for improving comprehension and application of jury instructions, which have met with some significant, albeit limited, success (Ogloff & Rose, 2005). This article examines a new

method of instructing the jury that has been developed primarily by judges – the so-called fact-based approach to instruction.

Past research on comprehension and application of jury instructions

Plain language instructions

Most studies on improving comprehension of jury instructions have focused on so-called ‘plain language’ revisions to instructions (Ogloff & Rose, 2005). Plain language revision involves revising instructions at the sentence level, removing syntactic and semantic elements thought to impede comprehension.

Correspondence: Benjamin Spivak, Centre for Forensic Behavioural Science, Swinburne University of Technology and Forensicare, 505 Hoddle Street, Clifton Hill, Melbourne, Victoria 3068, Australia. Email: bspivak@swin.edu.au

Researchers have mostly found that plain language revision has a modest improvement on the ability of lay participants to paraphrase instructions and answer multiple-choice items relating to the text in simulated trial scenarios (Charrow & Charrow, 1979; Diamond & Levi, 1996; Severance, Greene, & Loftus, 1984; Steele & Thornburg, 1988; Tiersma & Curtis, 2008).

The limited improvements recorded with plain language revised instructions are not surprising considering what a juror must do with those instructions. Where written instructions are not provided, they are expected to be able to recall these concepts and relate them to one another coherently, identify the factual circumstances that they refer to, and finally to execute a decision-making strategy that adheres to the various criteria set out by the judge. While simplifying sentence structure and using familiar terminology are necessary to assist jurors to successfully encode sentential information in memory, such revisions alone are insufficient to guide jurors to utilise the legal concepts communicated by the judge.

The use of checklists

Consistent with the limitations identified above, Wiener et al. (2004) argued that standard instructions, being a one-trial learning task, provided limited opportunity to develop cognitive skills required to apply the law correctly. They proposed using a document that parsed the language of instructions into concepts and steps necessary to reach a decision. Checklists (also known as ‘step-down directions’, ‘jury guides’, or ‘flowcharts’) consist of a written list of ordered questions corresponding to the substantive elements in the case. The use of a checklist ought to obviate the need for a juror to undergo the mentally taxing process of constructing an appropriate sequence of decision-making steps to verdict.

While checklists may offer benefit by assisting jurors to apply the law, results of research evaluating the efficacy of checklists has been mixed. Wiener et al. (2004)

examined the effect of checklists (which they termed ‘flowcharts’) on application and comprehension of instructions in a simulated trial. While improvements were found in comprehension and application of state law relative to the pattern instructions provided in the case, there was no significant difference between the checklist condition and other conditions on comprehension and application of constitutional law. When Semmler and Brewer (2002) tested the effect of checklists on comprehension of self-defence, they found that participants who received checklists scored significantly higher when asked to paraphrase legal concepts than participants who received instructions only. However, the authors found no differences between groups in the ability to apply the law correctly to a factual scenario.

Although checklist-aided instructions have shown some improvement over standard instructions, they have not been a panacea. A limitation of checklist-aided instructions is the linguistic complexity of the instructions themselves, which were not measured or specified in either study. Research on the interaction between linguistic revision and the use of written documentation suggests that the effectiveness of written materials in improving comprehension is influenced by the complexity of the language used in instructions (Prager, Deckelbaum, & Cutler, 1989).

Although the use of plain language revision in concert with checklists may improve legal comprehension beyond either method alone, theoretically there remains considerable room for improvement – a point that has prompted several scholars to call for a new approach to instructing the jury (Cork, 2002; Ogloff & Rose, 2005).

Fact-based directions

The fact-based approach has its origins in work conducted by judicial officers (Auld, 2001; Chambers, 2011) to improve the comprehensibility and applicability of jury instructions. Under the ‘fact-based’ approach, judges

Figure 1. Structure of standard instructions.

determine the elements of the crime that must be proved, as well as those of any relevant defences. These elements are converted into the form of questions relating directly to the facts of the case, and are then placed in their most logical order with evidentiary directions inserted where necessary. Consistent with other direction practices, the judge typically begins with an explanation of the standard and burden of proof; however, unlike other forms of directions, the judge does not proceed to define the charge and explain the elements or the defences. Rather the judge prepares a series of questions that embeds the law and relevant facts.¹ The jury is directed to approach each question in turn, and they are told that they should reach a decision on the first question before they turn to the next, and so on. If they follow the direction, they will have addressed all the elements necessary to reach their verdict (Judicial College, 2016; Leveson, 2015). An example of the difference in structure is presented in Figures 1 and 2.

In addition to the changes in structure, the fact-based approach also involves the provision of written assistance to the jury, in the form of a question trail. The question trail is a written copy of the factual questions for the jury placed in the same logically sequential order as the judge's instructions. Unlike checklists, which typically utilise a sequence of abstract questions of law, question trails utilise a sequence of concrete questions that make

explicit reference to the facts of the case. Furthermore, under the fact-based approach, the oral delivery of directions mirrors the question trail, meaning that the judge instructs the jury using the same sequence of questions as those provided in the question trail.

It is important to distinguish the fact-based approach from other innovations designed to enhance jury comprehension. For example, a number of researchers have previously implemented structural changes alongside plain language revision. Elwork, Sales, and Alfini (1982) supported the use of hierarchical-algorithmic-associational (HAA) structure – a method that involved presenting information in such a way that higher level concepts are reduced to fundamental concepts (hierarchical), that concepts are ordered in such a way that comprehension of each is necessary to understand the next (algorithmic) and that associated concepts are presented together (associational). In contrast, the structure implemented in the fact-based approach is determined by logically ordering the series of decisions that need to be made by the jury to reach the verdict. Ultimately, the HAA structure can complement the fact-based approach, with sentence-level organisation determined by HAA and higher level text units organised using fact-based sequencing.

Experimental research on question trails and factual framing

No known empirical studies have investigated the utility of fact-based instructions. However, in recent years, experimental psychologists have tested various elements of the fact-based approach (Ede & Goodman-Delahunty, 2013; McKay, Nolan, & Smithson, 2013; Smith & Haney, 2011). For example, Smith and Haney (2011) tested the effect of 'pin-point' instructions that integrated evidence with key legal concepts. The authors compared participants receiving pin-point instructions with those receiving standard instructions on the law and plain language instructions in a simulated trial

Figure 2. Structure of fact-based instructions.

scenario. The results indicated that participants receiving pin-point instructions scored significantly higher than those in both other conditions on tests of application and comprehension.

While the research on factual framing of instructions has been limited, a number of experiments have sought to evaluate the effect of written question trails on comprehension and application. Ede and Goodman-Delahunty (2013) considered comprehension of instructions by comparing jury-eligible participants in three conditions: those receiving no instructions, those receiving a written copy of instructions, and those receiving the same written instructions with a question trail. Participants were scored on their ability to recall legal concepts by indicating whether a statement that closely worded the jury instructions were true or false, and their ability to apply legal concepts to facts. The authors found that participants who received a question trail performed better on their ability to apply substantive law than participants in other conditions. However, they also found that participants who received a written copy performed better on their ability to comprehend substantive law than participant receiving question trails. Similar findings were reported by Essex and Goodman-Delahunty (2014), who compared comprehension of the law for participants with a question trail and without. The authors found no statistically significant difference between the conditions.

The results of the above studies suggest that question trails may assist jurors to apply

the law but are less impressive as methods of enhancing legal comprehension. Given the study design, this is not altogether surprising. Ede and Goodman-Delahunty (2013) utilised the same instructions in both conditions, varying only the aides given to jurors. The benefits of question trails, like checklists, should primarily consist in developing procedural knowledge for a task, which should in turn enhance application. It is unclear how the question trail would assist in learning legal concepts given that explanations of law are typically not included in question trails.

On the whole, the evidence adduced in favour of factual framing and question trails appears to suggest that these methods are a useful means of enhancing juror comprehension and application of law. On the other hand, the studies reviewed have examined factual framing and question trails separately. None of the studies to date have examined the effect of utilising both factual framing and question trails simultaneously, as is done under the fact-based approach. Given that both of these methods appear to improve comprehension and application of the law, it may be that the combination of both would improve performance further still.

The present study

Evidence from controlled mock jury experiments has shown that approaches such as plain language revision and checklists are helpful in enhancing comprehension and application of

the law. While these approaches have proven useful in enhancing comprehension and application, there are clear limitations on their capacity to maximise juror performance. The proposals made by Auld (2001) and Chambers (2011) for question trails and fact-based approaches may enhance jury comprehension and application of the law beyond the use of plain language, checklists and written materials. The integration of directions with the question trail provides something lacking in previous approaches – an overarching structure. The effect of structure has not been examined empirically in the literature on jury comprehension and application of directions to date. An examination of fact-based directions compared to other approaches to directing the jury offers the opportunity to extend research in this direction and introduce advances in discourse processing that have not been previously applied in the jury context. The present study attempts to address this gap in the literature by evaluating fact-based instructions against both standard approaches to instruction and plain language revised instructions using a simulated sexual assault trial.

The study was based on the law concerning sexual assault in the state of Victoria, Australia at the time of the study. The complexity of the relevant law makes it a useful test case to determine whether fact-based approaches have utility in assisting the jury to understand and apply legal concepts in their decision-making. In Victoria, judges are provided with a ‘Bench Book’, which provides model instructions on the law (Judicial College of Victoria, 2015). The model instructions on sexual assault are conceptually complex, requiring the person hearing them to isolate and execute various subtasks in order to apply the instructions to facts. The element relating to the state of mind of the accused is an example of this sort of conceptual complexity. In the case of rape, the model instructions, at the time, read:

The **fourth element** relates to the accused’s state of mind about the complainant’s

consent. The prosecution must prove beyond reasonable doubt that at the time of sexual penetration the accused: was aware that the complainant was not or might not be consenting; **or** was not giving any thought to whether the complainant was not or might not be consenting (Judicial College of Victoria, 2015).

A juror hearing these instructions not only must overcome the hurdle of linguistic complexity, but must then extract the three tasks contained in the element, order them appropriately and understand the consequences involved in deciding each of the tasks (e.g. If I decide that the complainant was not giving any thought to whether the complainant was not consenting, there is no need to engage with the other questions). Plain language revision is likely to be of only limited help in navigating this element successfully. The checklist and fact-based approaches are designed to assist with this sort of issue by providing jurors with an obvious structure to guide decision-making – taking them through each element, question by question, with the relationship to the ultimate verdict laid out clearly.

While both the checklist and fact-based approaches are both designed to deal with the same problem – namely, application of legal concepts in decision-making – one might expect to see better application performance in groups that receive fact-based directions. The abstract nature of sexual assault instructions presents an additional mental task for jurors. They must be able to substantiate abstraction with the concrete details of the case. For example, when told that they must determine whether the defendant was aware that the complainant might not be consenting, interpretation requires the juror to tie the correct referents to the terms complainant and defendant and also to refer to the correct factual situation under dispute. A checklist, even accompanied by plain language directions, does not assist with this task.

To examine the differences in comprehension and application between fact-based instructions relative to standard instructions,

instructions revised according to plain language principles and plain language instructions accompanied by a checklist, a simulated sexual assault trial was conducted to examine three hypotheses: (a) Participants in the standard condition would have significantly lower proportion of correct paraphrase, true/false and multiple-choice application scores than those in the plain language condition; (b) participants in the plain language condition would have significantly lower proportion of correct true/false and multiple-choice application scores than those in both the checklist and fact-based conditions; (c) participants in the checklist condition would have significantly lower proportion of correct true/false and multiple-choice application scores than those in the fact-based condition.

Method

Participants

Participants were recruited from the regular pool of people summonsed to serve on juries in cases in the County and Supreme Courts of Victoria over a one-year period. After arriving at the court house, people were invited to volunteer to participate in this study in lieu of regular jury service. All participants were jury eligible.

A total of 1007 participants took part in the study. Participants were almost evenly divided by gender ($n = 525$, 52.1% males; $n = 478$, 47.5% females; and $n = 4$, 0.40% unspecified), and they ranged in age from 18–87 years ($M = 44$, $SD = 14.1$). In total 89 groups participated in the study. Groups varied from 10–12 members, with a median of 11 members for the sample.

Materials

Simulated trial video

A simulated trial was filmed based on an actual criminal trial transcript from the County Court of Victoria. Actors volunteered to play members of the public, the jury and the

witnesses in the case. The judge was played by a sitting County Court judge, and the lawyers were played by two practising lawyers. The video ran for approximately one hour.

In addition to the trial video, videos were prepared containing the judicial instructions that the judge gave to the jury. Although there were four instructional conditions, only three instructional videos were required. One of the instructional videos (plain language) was used for two conditions (plain language; plain language + checklist).

Instructions

Standard instructions

Standard instructions were drafted using the Judicial College of Victoria's model texts as a template (Judicial College of Victoria, 2015). To ensure accuracy, several drafts were reviewed and amended by senior Victorian judges and members of an advisory committee overseeing this research. The standard instructions followed the typical structure utilised in Australian jury trials in which the judge instructs the jury on law, summarises the evidence presented and summarises the addresses from the lawyers (Ogloff et al., 2006). Running time of the video was approximately 1 hour 15 min.

Plain language instructions

These instructions were a revised version of the standard instructions using empirically supported plain language principles (Charrow & Charrow, 1979; Elwork, Sales, & Alfini, 1982). While the structure of topics remained similar to that in the standard condition, language was revised to remove multiple negation in sentences, nominalisations and repetition. The syntax was also revised where possible to conform to a simple subject–verb–object structure. In addition, participants were provided with a written copy of the instructions that they could utilise during the instructional video and during deliberation.

Running time of the video was approximately one hour.

Plain language instructions plus checklist

These instructions were identical to the plain language instructions except for the inclusion of a written checklist taken from the Judicial College of Victoria Charge Book (JCV, 2015). The checklist contains all of the elements for each offence and asks jurors to decide whether they are convinced beyond reasonable doubt that each element has been proved. The checklist was presented at the same time as the videotaped instructions, and participants were permitted to utilise it during deliberation.

Fact-based directions

The fact-based directions were drafted with the assistance of a New Zealand Supreme Court justice. The drafting of the fact-based instructions was different from that in the previous three conditions in two ways. First, the legal directions were structured around a written question trail that the participants were required to follow. Secondly, where possible, directions were explained with reference to concrete examples taken from the case and to the lawyers' respective positions on the issues. This differed substantially from the previous three conditions, where the summary of evidence and the lawyers' addresses were summarised independently of the legal instructions. The question trail and a transcript of the directions were presented to participants at the same time as the videotaped instructions, and participants were permitted to utilise them during deliberation.

Readability

To quantify the extent of linguistic revision in both the plain language and fact-based conditions, the Flesch–Kincaid Grade Level (FKGL) was calculated for both sets of instructions and compared to the score for the standard condition. The standard condition text had the highest score (FKGL = 9.9). Both

the plain language text (FKGL = 8.3) and the fact-based text (FKGL = 7.7) had relatively lower scores.

Assessment materials

Assessments of pre-trial knowledge

Pre-trial knowledge of the law was assessed through the use of a multiple-choice recognition task and a recall task. The recognition questionnaire consisted of 12 questions. Each question contained one correct answer and three distractors. The recall task was a five-item short answer recall questionnaire that assessed the ability to reinstate legal concepts upon request. Items for both assessments were scored as either correct or incorrect.

Short answer paraphrase questionnaire

The short answer paraphrase questionnaire consisted of seven short answer questions designed to assess participant understanding of legal concepts contained in the trial instructions. Paraphrase tasks have been utilised in prior studies of juror comprehension (Charrow & Charrow, 1979; Rose & Ogloff, 2001; Steele & Thornburg, 1988). Charrow and Charrow (1979) argue that 'the validity of the paraphrase task as a measure of comprehensibility rests on the premise that a subject will not be able to paraphrase accurately material that he or she has not understood' (p. 1310). Responses were coded as either correct or incorrect based on evaluation against the definitions provided in the instructions.

Application questionnaires

Two application questionnaires were developed for this study. Previous jury research has utilised both multiple-choice form questions and true/false choice (e.g. yes/no, true/false) type questions in relation to legally unambiguous statements or scenarios that relate to the case (Ede & Goodman-Delahunty, 2013; Rose & Ogloff, 2001; Wiener et al., 1998). Both formats were utilised in this study. A multiple-choice questionnaire was developed to test understanding of

legal concepts contained in the trial instructions. Participants were instructed to select one answer for each question. In total, the questionnaire consisted of 12 questions. Each question contained one correct answer and three distractors.

A true/false choice questionnaire was developed to test application of legal concepts contained in the trial instructions. The questionnaire consisted of 10 questions testing application of concepts such as the standard of proof, the consent element and the *mens rea* element. Each question contained a hypothetical reason for a verdict. Participants were instructed to indicate whether the reasoning was legally correct or incorrect.

Procedure

Following the selection of a group, the group was allocated to one of the four experimental conditions on an iterative basis. Once allocated, the individual members of the group were seated in a jury room in the courthouse, introduced to the researcher and provided with explanatory statements and consent forms. After signing the consent forms, participants were asked to complete the pre-test assessment materials. The process of explanation, obtaining informed consent and testing participants took approximately one hour.

Each group selected a foreperson and viewed the videotaped trial. Screening time varied from 105 min to 130 min depending on condition. A break was provided before the judge's directions were given. Following the completion of the directions, the participants were instructed to complete the paraphrase task, multiple-choice application and true/false-choice application, in that order. Following completion of the post-trial assessment, the group was asked to begin their deliberations. They were instructed to try and reach a unanimous verdict. Mean deliberation time was approximately 55 min.

On days where the group started the trial exercise early enough, they were asked to complete a final exercise following deliberation. They were asked to complete an identical multiple-choice questionnaire to the one

presented at the pre-deliberation phase as a group. This group exercise was recorded and completed at the participants' own pace. The post-deliberation exercise took approximately 15 min on average to complete. Following completion of the group exercise, the participants were thanked by the researcher and reminded that a copy of the results would be made available upon request.

Ethics approval

The Department of Justice Human Research Ethics Committee provided ethical approval for the study on 2 July 2012, and Monash University Research Ethics Committee provided ethics approval on 1 August 2012.

Results

Pre-trial understanding of the law

Total pre-trial comprehension and application of legal concepts

Scores from both questionnaires were summed to provide an overall score for pre-trial understanding (possible range = 0–27). The means and standard deviations for each of the conditions are presented in Table 1. Results of a one-way analysis of variance (ANOVA) revealed that the pre-trial understanding level was slightly higher for participants in the fact-based condition $F(3, 889) = 3.13$, $p = .03$, $\eta_p^2 = .01$.

Comprehension and application at the pre-deliberation phase

Participant scores on the paraphrase, true–false application and multiple-choice application items were summed and divided by the maximum possible score to produce percentage scores for each participant. To test for co-linearity, each of the three scales were correlated with one another. Pearson correlation coefficients ranged from $r = .31$ to $r = .41$; all correlations were significant at an alpha level of .05, with p values all below .001. Means and

Table 1. Means, standard deviations and confidence intervals for total pre-trial understanding between conditions.

Condition	Mean % correct	95% CI	SD
Standard	39.70	[37.90, 41.50]	13.83
Plain language	40.55	[38.56, 42.53]	13.80
Checklist	41.69	[39.89, 43.48]	13.80
Fact-based	43.75	[42.00, 45.49]	13.42
Overall	41.46	[40.55, 42.38]	13.77

Note: CI = confidence interval.

Table 2. Mean percentage correct and standard deviations for total paraphrase, total true/false application and total multiple-choice application.

Condition	Paraphrase <i>M (SD)</i>	T/F application <i>M (SD)</i>	MC application <i>M (SD)</i>
Standard	37.45 (26.14)	75.33 (15.30)	49.87 (21.48)
Plain language	61.56 (24.91)	79.33 (15.97)	56.83 (22.44)
Checklist	58.44 (25.60)	78.65 (15.08)	53.81 (22.46)
Fact-based	68.77 (24.26)	82.81 (12.83)	56.43 (21.54)
Overall	56.62 (27.83)	79.03 (15.02)	54.13 (22.10)

Note: T/F = true/false; MC = multiple-choice.

standard deviations for each condition are presented in Table 2.

A multilevel model was created for each of the tasks to account for non-independence of observations. Intra-class correlations for each of the three percentage scores ranged from $\rho = .03$ to $\rho = .15$, indicating that between 3% and 15% of the variation in scores was accounted for by variation within individual groups. As a result, the models each included a random intercept fitted for the group factor. Total pre-trial knowledge scores and education scores were included as fixed-effect Level 1 covariates in the model. The covariate estimates, intercept values and model fit indices for each of the models are presented in Table 3.

In each model, a set of mean dependent variable scores estimated at the mean level of the covariates were produced for each condition. The estimated means and standard deviations for each condition across the three models are presented in Table 4. As with the

descriptive means presented in Table 1, the standard condition appears to have uniformly the lowest average scores.

Hypothesis 1: participants in the standard condition would have a significantly lower proportion of correct paraphrase, true/false and multiple-choice application scores than those in the plain language condition

Hypothesis 1 was addressed by way of three two-tailed contrasts with a Bonferroni correction applied to the alpha level ($\alpha = .017$) conducted between the standard and plain language condition for each of the paraphrase, $t(79.07) = -8.35$, $p < .001$, $r_{\text{contrast}} = .68$, true/false, $t(80.81) = -1.35$, $p = .18$, $r_{\text{contrast}} = .15$, and multiple-choice, $t(79.66) = -1.77$, $p = .08$, $r_{\text{contrast}} = .19$, application scores. The test results indicate that the only condition on which there was a statistically significant difference between means was for paraphrase scores. For both true/false and multiple-choice application, there was not enough evidence to

Table 3. Covariate estimates, model fit indices and intercept values for total paraphrase, total true/false application and total multiple-choice application models.

Model	Pre-trial understanding of law	University qualification	Vocational qualification	High school	Fixed intercept	Random intercept	Model summary (-2LL)	Number of parameters
Paraphrase	0.70	~	-10.65	-14.12	43.36	14.70	7418.40	10
T/F application	0.48	~	-2.06	-5.58	63.81	4.86	6619.25	10
MC application	0.76	~	-9.97	-10.92	28.01	4.92	7338.89	10

Note: T/F = true/false; MC = multiple-choice; LL = Log Likelihood.

conclude that the variation was beyond what would be expected by chance.

Hypothesis 2: participants in the plain language condition would have a significantly lower proportion of correct true/false and multiple-choice application scores than those in the checklist and fact-based conditions

Two contrasts were conducted to test Hypothesis 2. Each of the contrasts assessed the difference between scores in the plain language condition and scores in the checklist and fact-based conditions. The results of two-tailed tests with a Bonferroni correction applied to the alpha level ($\alpha = .025$) on the true/false, $t(76.05) = -0.98$, $p = .33$, $r_{\text{contrast}} = .11$, and multiple-choice scores, $t(78.36) = 1.27$, $p = .21$, $r_{\text{contrast}} = .14$, indicated that there was not enough evidence to conclude that the mean differences between conditions was beyond what would be expected by chance.

Hypothesis 3: participants in the checklist condition would have a significantly lower proportion of correct true/false and multiple-choice application scores than those in the fact-based condition

Two contrasts were conducted to test Hypothesis 3. The results of two-tailed tests ($\alpha = .025$) on true/false, $t(76.51) = 2.46$, $p = .02$, $r_{\text{contrast}} = .27$, and multiple-choice application, $t(80.34) = 0.361$, $p = .72$, $r_{\text{contrast}} = .04$, scores indicated a significant difference between conditions on true/false items and no significant difference between conditions on multiple-choice scores.

Post-deliberation analysis

Where time permitted, groups of participants were asked to complete the multiple-choice application questionnaire as a group following deliberation. In total, 23 groups (26% of total sample) completed the questionnaire. There were five groups in the fact-based condition; all other conditions contained six groups.

Table 4. Estimated mean percentage correct and standard deviations for each condition in paraphrase, true/false application and multiple-choice application models.

Condition	Paraphrase <i>M (SD)</i>	T/F application <i>M (SD)</i>	MC application <i>M (SD)</i>
Standard	38.43 (25.94)	76.41 (14.35)	51.04 (18.76)
Plain Language	59.47 (24.85)	79.33 (14.30)	54.45 (19.12)
Checklist	56.33 (25.35)	78.65 (16.09)	51.96 (14.47)
Fact-based	64.49 (24.15)	82.81 (15.01)	52.62 (15.11)
Overall	57.15 (27.63)	79.03 (15.02)	54.13 (22.10)

Note: T/F = true/false; MC = multiple-choice.

Table 5. Mean percentage correct and standard deviations for pre- and post-deliberation multiple-choice application.

	Pre-delib. <i>M</i>	Post-delib. <i>M</i>	Pre-delib. <i>SD (range)</i>	Post-delib. <i>SD (range)</i>
Standard	52.08	79.83	23.08 (0–91.66)	17.08 (33.33–100)
Plain language	56.92	83.33	18.33 (16.67–91.66)	13.33 (58.33–100)
Checklist	51.17	79.58	24.50 (8.33–91.66)	18.58 (58.33–100)
Fact-based	59.42	92.83	24.67 (8.33–100)	6.91 (83.33–100)
Overall	54.50	83.50	23.00 (0–100)	15.83 (33.33–100)

Note: delib. = deliberation.

Descriptive comparisons between the participants’ pre-trial results and post-trial results are presented in Table 5.

The pre- and post-trial pattern of results suggests a significant improvement for each condition. However, the pattern of differences between conditions changed somewhat from pre-deliberation to post-deliberation. To assess the difference in improvements between each condition, the total number of multiple-choice questions correct at the pre-deliberation testing phase was subtracted from the total number of multiple-choice questions correct at the post-deliberation stage. The total score on the pre-deliberation multiple-choice application task was also included as a predictor to evaluate the extent to which ability on the previous task was predictive of performance on the subsequent task. The results are summarised in Table 6.

Participants in the fact-based condition had significantly higher scores than participants in all other conditions, even after pre-trial

understanding of the law and previous scores on the multiple-choice application task were accounted for by the model. Post hoc tests revealed no significant differences between other conditions at a Bonferroni corrected $p = .0125$ level. Of the other factors considered in the model, only the scores on the pre-deliberation multiple-choice application task were significantly predictive of post-deliberation performance.

Discussion

The results of the study provide mixed evidence to support the notion that reorientating jury instructions to support the fact-finding process is likely to improve comprehension and application of the law relative to traditional approaches to instructing the jury. In support of the notion, fact-based directions were associated with gains in comprehension relative to standard directions and gains in

Table 6. Coefficients and model summary for post-deliberation multiple-choice application scores.

	Predictors	Regression coefficient	95% CI	<i>p</i>
<i>Experimental conditions</i>	Intercept	76.59	[−79.88, 233.06]	<.001
	Fact-based condition	~	~	~
	Standard condition	−12.64	[−18.01, −7.20]	<.001
	Plain language condition	−9.49	[−14.91, −4.01]	.001
<i>Personal factors</i>	Checklist condition	−11.91	[−17.26, −6.56]	<.001
	Pre-trial understanding of law	0.14	[−0.02, 0.30]	.00
	Pre-deliberation multiple-choice scores	−0.83	[−0.93, −0.73]	<.001
<i>Random coefficients</i>				
<i>Model summary</i>	Intercept	99.13	—	—
	−2LL	1793.50	—	—
	Number of parameters	8		

Note: ~ indicates a baseline condition. CI = confidence interval; LL = Log Likelihood.

true/false application relative to all other conditions tested. Furthermore, this difference remained when pre-trial knowledge of the law was statistically controlled for as a covariate.

However, the level of improvement associated with fact-based instructions indicates that the method is not a panacea for difficulties facing members of the public in trying to understand and apply the law provided to them in jury trials. The improvements associated with application items at the pre-deliberation phase of the trial were minor, and the differences between multiple-choice application items were not significant at the pre-deliberation stage. Conversely, when groups of participants were asked to deliberate on various questions of law, participants receiving fact-based directions not only had higher average proportions correct relative to participants in other conditions, but the dispersion around the mean was

also multiples smaller than that in other conditions. The finding is indicative of an important corrective influence that deliberation may have on incorrect understandings of the law when guided by appropriately developed instructions.

The findings also support previous research in relation to the differences between standard, plain language and checklist approaches to instructing the jury (Charrow & Charrow, 1979; Diamond & Levi, 1996; Tiersma & Curtis, 2008; Wiener et al., 2004). Both plain language instructions and plain language instructions with a checklist resulted in improved levels of comprehension relative to standard instructions, a finding consistent with those of previous studies (Charrow & Charrow, 1979; Diamond & Levi, 1996; Prager et al., 1989; Severance, Greene, & Loftus, 1984; Severance & Loftus, 1982; Tiersma & Curtis, 2008; Wiener et al., 2004).

However, participants who received plain language instructions alone outperformed participants who received the same revisions alongside the checklist. This suggests that the use of these revisions does not have a straightforward additive effect on comprehension and application of directions. Indeed, it may be that the use of a checklist has little benefit when the instructions are already made clear. Alternatively, it may be that the checklist had little utility when the instructions are not integrated with the checklist as with fact-based directions.

Support for the idea that the lack of difference between the two conditions was caused by a failure to integrate the two different modalities (e.g. making sure that the order of instructions was identical to the order used in the checklist, making reference to the checklist in the instructions) can be found in a number of studies in instructional psychology that have considered the effect of utilising multiple modalities to enhance task performance. Sweller and colleagues have found that task performance decreases where students are expected to mentally integrate disparate sources of information presented in different modalities – a finding they have termed the ‘split-attention effect’ (Chandler & Sweller, 1992).

If such an effect is assumed to have been at play in the current study, it may also explain why fact-based directions, which integrate a question trail into the instructions, performed relatively better than participants receiving plain language and a checklist. The participants receiving fact-based directions did not have to go through the process of mentally linking each question in the checklist with the appropriate part of the instructions, reducing the amount of cognitive load required to successfully apply the instructions.

Limitations

The current study manipulated multiple elements across conditions simultaneously, rather than isolating each element tested. In the

standard condition participants were provided with a set of standard verbal directions that were relatively linguistically complex. The instructions were developed based on the manual judges rely upon and were reviewed and approved by several judges. In the plain language condition, participants were provided with both a written copy of the directions and multiple linguistic variations. The extent to which changes in comprehension and application were caused by any one specific linguistic variation or written assistance cannot be ascertained from the present study. The only reasonable inference that one can make is that there seems to be an improvement in comprehension when both are utilised. Similarly, there were multiple linguistic variations between the fact-based condition and other conditions, meaning that similar uncertainty exists about the precise element or combination of elements that is responsible for any observed effects.

In any applied psychological research, a tension exists between the extent to which experimental results are internally valid and the extent to which they are seen as practical and useful for the community they are designed to inform and assist. Elwork, Sales, and Alfini (1982) pointed to this tension when they performed their work on simplifying pattern jury instructions, noting that the number of controlled experiments one would have to run to ascertain the exact combination of linguistic variables that might result in consistently improved comprehension would run into the thousands given the combinatorial possibilities. The present study sought to follow the path of other applied research in this area that has foregone strong internal validity for the sake of applicability (Elwork, Sales, & Alfini, 1982; Severance & Loftus, 1982; Wiener et al., 2004). The standard directions were written with the assistance of judges, and the absence of a written copy reflected the practice of a large number of judges at the time the study was conducted (Ogloff et al., 2006). Following Wiener et al. (2004), both written

copy and linguistic variations were utilised in an attempt to boost comprehension and application to the greatest extent possible. An additional set of conditions testing linguistic variations without the use of written copy would have been ideal; however, due to the practical limitations associated with recruiting such a large sample, this option was rejected.

Conclusion

The findings of the present study support two claims that have been made repeatedly by legal professionals and academics in recent years. The first is that lay people struggle to comprehend and apply legal directions (Lieberman & Sales, 1997; Ogloff & Rose, 2005). Specifically, in the Victorian criminal justice system, the results support claims made by several judges and law reform bodies concerning the difficulty involved in comprehending substantive directions in rape trials involving consent and awareness (Victorian Law Reform Commission, 2009; Weinberg, 2012) and evidentiary directions involving consciousness of guilt and delayed complaint (Eames, 2007; Neave, 2012; Victorian Law Reform Commission, 2009).

Secondly, the findings also support claims that comprehension and application can be improved through the use of revisions to directions such as plain language revision, written copies of directions and decision-making aides like checklists and question trails (Charrow & Charrow, 1979; Diamond & Levi, 1996; Ede & Goodman-Delahunty, 2013; Rose & Ogloff, 2001; Smith & Haney, 2011; Tiersma & Curtis, 2008; Wiener et al., 2004). They demonstrated that while lay people struggle to comprehend and apply these directions in standard form, comprehension and application can be improved. In particular, the findings provide support for proponents of integrated directions and question trails as a means by which comprehension and application can be improved. However, they also suggest that more modest revisions to directions such as

plain language with written copy improve comprehension beyond standard directions. This is also supported by studies that suggest that jurors themselves would like to receive written directions to assist their deliberations (Thomas, 2013).

Of course, juror comprehension and application of the law is not the only factor that requires consideration before innovations are implemented. Trial judges face a significant workload that may reduce the amount of time available to draft directions and consequently restricts the ability of some judges to implement the revisions considered in the current study. The revision of model directions in line with plain language principles is one possible method that may improve comprehension without unduly increasing a trial judge's workload. However, the use of fact-based directions requires specific tailoring of directions to the evidence in a specific case. Thus, trial judges who wish to use the method that the present study found to be most efficacious in improving comprehension and application may be burdened with an increased workload. The cost effectiveness of fact-based directions is an important area to consider before its use is promoted among trial judges.

The study also raises a number of questions for future studies. As with all simulated research, the question of generalisability looms large. While the findings suggest that fact-based instructions are helpful to participants in the controlled conditions of a simulation, the extent to which these effects hold up across time, place and situation is at present unknown. A field study comparing jurors receiving fact-based directions to jurors receiving standard form directions would be a step toward addressing concerns around external validity.

A second area of interest raised by the present study relates to the potential for interaction between instruction type, deliberation and application of the law. The improvements in application for the fact-based condition at the post-deliberation stage suggest that

deliberation can interact with instruction type to enhance application. However, the question of how such improvements actually take place is at present unknown. An analysis of deliberation content utilising the same experimental design may assist in determining the factors at play.

Acknowledgments

The researchers gratefully acknowledge the assistance of the Victorian Juries Commissioner Paul Dore and the staff at the Victorian Juries Commissioner's Office in facilitating data collection.

Ethical standards

Declaration of conflicts of interest

Benjamin Spivak has declared no conflicts of interest

James R. P. Ogloff has declared no conflicts of interest

Jonathan Clough has declared no conflicts of interest

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

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

Funding

This research was funded by a Linkage Grant from the Australian Research Council in partnership with the Department of Justice and Regulation, Victoria.

Note

1. In England and Wales, a similar concept is known as a 'Route to Verdict' – that is, a sequence of questions posed in a logical and necessary sequence (Judicial College, 2016).

References

- Auld, R. (2001). *Review of the criminal courts of England and Wales*. London, UK: National Archives.
- Chambers, R. (2011). Is there another way? Paper presented at the Judicial College of Victoria, Melbourne.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233–246. doi:10.1111/j.2044-8279.1992.tb01017.x
- Charrow, R. P., & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, 79, 1306–1374.
- Cork, C. M. (2002). A better orientation for jury instructions. *Mercer Law Review*, 54, 1–30.
- Dann, B. M. (1993). 'Learning lessons' and 'speaking rights': Creating educated and democratic juries. *Indiana Law Journal*, 68, 1229–1279.
- Diamond, S. S., & Levi, J. N. (1996). Improving decisions on death by revising and testing jury instructions. *Judicature*, 79, 224–232.
- Eames, G. (2007). Tackling the complexity of criminal trial directions: What role for appellate courts? *Australian Bar Review*, 29, 161–194.
- Ede, T., & Goodman-Delahunty, J. (2013). Question trails in trials: Structured versus unstructured juror decision-making. *Criminal Law Journal*, 37, 114–136.
- Elwork, A., Sales, B. D., & Alfini, J. J. (1982). *Making jury instructions understandable*. Charlottesville, VA: The Michie Company.
- Essex, R., & Goodman-Delahunty, J. (2014). Judicial directions and the criminal standard of proof: Improving juror comprehension. *Journal of Judicial Administration*, 24, 75–94.
- Judicial College of Victoria. (2015). Victorian criminal charge book. Retrieved from <http://www.judicialcollege.vic.edu.au/publications/victorian-criminal-charge-book>
- Judicial College. (2016). The Crown Court compendium part I: Jury and trial management and summing up. Retrieved 27/11/2016 from <https://www.judiciary.gov.uk/wp-content/uploads/2016/05/crown-court-compendium-part-i-jury-and-trial-management-and-summing-up.pdf>
- Leveson, B. (2015). *Review of efficiency in criminal proceedings*. London, UK: The Stationery Office. Retrieved 23/11/16 from <https://www.judiciary.gov.uk/wpcontent/uploads/>

- 2015/01/review-of-efficiency-in-criminal-proceedings20151.pdf
- Lieberman, J. D., & Sales, B. D. (1997). What social science teaches us about the jury instruction process. *Psychology, Public Policy and Law*, 3, 589–644.
- McKay, C., Nolan, M., & Smithson, M. (2013). Effectiveness of question trails as jury decision aids: The jury's still out. *Psychiatry, Psychology and Law*, 21, 492–510 doi:10.1080/13218719.2013.839929
- Neave, M. (2012). Jury directions in criminal trials - Legal fiction or the power of magical thinking. Paper presented at the Supreme and Federal Court Judges' Conference, Melbourne.
- New South Wales Law Reform Commission. (2011). Jury directions. Sydney, NSW, Australia.
- Ogloff, J. R. P., Clough, J. A., Goodman-Delahunty, J., & Young, W. (2006). *The jury project: stage 1 - A survey of Australian and New Zealand judges*. Melbourne, Victoria: Australian Institute of Judicial Administration Incorporated.
- Ogloff, J. R. P., & Rose, V. G. (2005). Comprehension of judicial instructions. In N. Brewer & K. Williams (Eds.), *Psychology and law* (pp. 407–444). New York, NY: Guildford Press.
- Prager, I., Deckelbaum, G., & Cutler, B. L. (1989). Improving juror understanding for intervening causation instructions. *Forensic Reports*, 3, 187–193.
- Rose, V. G., & Ogloff, J. R. P. (2001). The comprehensibility of jury instructions: Validating a method. *Law and Human Behavior*, 25, 409–431
- Semmler, C., & Brewer, N. (2002). Using a flowchart to improve comprehension of jury instructions. *Psychiatry, Psychology and the Law*, 9, 262–270.
- Severance, L. J., Greene, E., & Loftus, E. F. (1984). Toward criminal jury instructions that jurors can understand. *The Journal of Criminal Law and Criminology*, 75, 198–233.
- Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to understand and apply instructions. *Law and Society Review*, 17, 153–198.
- Smith, A. E., & Haney, C. (2011). Getting to the point: Attempting to improve juror comprehension of capital penalty phase instructions. *Law and Human Behavior*, 35, 339–350. doi:10.1007/s10979-010-9246-0
- Steele, W. W., & Thornburg, E. G. (1988). *Jury instructions: A persistent failure to communicate*. *North Carolina Law Review*, 67, 77–119.
- Thomas, C. (2010). Are juries fair? Ministry of Justice Research Series, 1/10, Ministry of Justice, United Kingdom.
- Thomas, C. (2013). Avoiding the perfect storm of juror contempt. *Criminal Law Review*, 6, 483–503.
- Tiersma, P., & Curtis, M. (2008). Testing the comprehensibility of jury instructions: California's old and new instructions on circumstantial evidence. *Journal of Court Innovation*, 1, 231–261.
- Victorian Law Reform Commission. (2009). *Jury directions: Final report*. Melbourne, Victoria: Victorian Law Reform Commission.
- Weinberg, M. (2012). *Peter Brett memorial lecture*. Melbourne, Australia: Melbourne University Law School.
- Wiener, R. L., Hurt, L. E., Thomas, S. L., Sadler, M. S., Bauer, G. A., & Sargent, T. M. (1998). The role of declarative and procedural knowledge in capital murder sentencing. *Journal of Applied Social Psychology*, 28, 124. doi:10.1111/j.15591816.1998.tb01696.x
- Wiener, R. L., Rogers, M., Winter, R., Hurt, L., Hackney, A., Kadela, K., ... Morasco, B. (2004). Guided jury discretion in capital murder cases. *Psychology, Public Policy and Law*, 10, 516–576.