
Development of a Measurement Tool for Courtroom Legal Actor Contributions: A Delphi Study Consulting the Experts

Rhonda Waterworth*

This article describes the development of a qualitative measurement tool – the Legal Actor Contributions Scale (LACS) – designed to measure legal actor contributions (primarily magistrates) in courtroom interactions from a therapeutic perspective. The measure was refined using a Delphi study to collect advice from research participants who are experts in the field of therapeutic change and magistrates’ therapeutic contributions (court craft). Despite adverse research conditions, the LACS measure was successfully developed into a refined form, presented in the article. This measurement scale will be useful for therapeutic and problem-solving court outcomes research, for magistrate court craft self-development, and in mainstreaming the therapeutic jurisprudence movement.

INTRODUCTION

It is widely accepted that legal actors, particularly judges and magistrates, play a significant role in the therapeutic justice process, both in encouraging defendants’ rehabilitation and their resistance to future reoffending.¹ There is a current lack of information about the extent to which legal actors engage in behaviours that have been recognised to have a therapeutic impact and to be a significant variable in the therapeutic outcome for defendants.²

Research into the efficacy of therapeutic practices in courtrooms has generally been hampered at a research design level by the lack of available measures to adequately describe or quantify the therapeutic contribution of the legal actors involved.³ Recent research and trends in the field of therapeutic jurisprudence (TJ) have demonstrated the need for a measurement tool to capture the therapeutic contributions of legal actors, most notably the judiciary.⁴ This article describes the development of such a tool via systematised consultation with experts in the therapeutic change field using the Delphi study methodology, and the incorporation of their feedback into a refined version of the measurement tool. The future potential applications for the measurement tool are then discussed.

* Australian Lawyer and Counselling Psychologist (Australia, France and the UK), PhD candidate in the Faculty of Law, University of Tasmania. The author gratefully acknowledges the assistance of supervisors Adjunct Assoc Prof Terese Henning, Dr Isabelle Bartkowiak-Theron, Dr Mandy Matthewson, and Dr Kimberley Norris. Special thanks to Dr Martine Evans and Magistrate Pauline Spencer for their guidance and support.

¹ R Waterworth, “The Case for Measuring Legal Actor Contributions in Court Proceedings” (2018) 26(1) *Psychiatry, Psychology and Law* 77; Michael S King, *Solution-Focused Judging Bench Book* (AIJA, 2009) 21; Bruce J Winick and David B Wexler, “Drug Treatment Court: Therapeutic Jurisprudence Applied” in Bruce J Winnick and David B Wexler (eds), *Judging in a Therapeutic Key: Therapeutic Jurisprudence and the Courts* (Carolina Academic Press, 2003) 106, 109.

² See Waterworth, n 1.

³ See Waterworth, n 1.

⁴ DB Wexler, “From Theory to Practice and Back Again in Therapeutic Jurisprudence: Now Comes the Hard Part” (2011) 37(1) *Monash University Law Review* 33; DB Wexler, “Moving Forward on Mainstreaming Therapeutic Jurisprudence: An Ongoing Process to Facilitate the Therapeutic Design and Application of the Law” (Arizona Legal Studies Discussion Paper, 2014) 15; E Richardson, P Spencer and D Wexler, “International Framework for Court Excellence and Therapeutic Jurisprudence: Creating Excellent Court and Enhancing Wellbeing” (2016) 25 *JJA* 148.

The Legal Actor Contributions Scale (LACS) is a qualitative measurement scale designed to capture legal actor contributions to interactions in court⁵ – in this case, magistrates in the magistrates court, particularly therapeutic and problem-solving courts.

The LACS is, at the time of writing, the only published and readily available behavioural measurement scale able to capture legal actor contributions from an objective, therapeutic perspective.⁶ While some informal observer rating systems have been used in courtroom research in the past to gather data on legal actors,⁷ as well as mutual ratings systems,⁸ the literature base from which these scales have been derived has not been articulated, and the scales themselves have not been shared with the wider research community. Additionally, mutual rating systems are not generally accurate predictors of therapeutic outcomes.⁹

It is fundamentally useful to court outcomes for defendants, court-quality improvement, magistrate training and court outcomes research to be able to capture and describe the therapeutic contributions of magistrates (and, eventually, other legal actors) within a court setting.¹⁰

Overview of LACS Development

The LACS was first published in draft form in 2019. It was grounded in a synthesis of the literature on common important determinants of therapeutic outcomes, as well as how it relates to TJ, procedural justice (PJ) and legitimacy of justice (LJ).¹¹

The first step in the development of the LACS was to pinpoint and define, based on empirical research and the work of experts in therapeutic court craft, the desirable behaviour of magistrates during courtroom interactions with defendants that would have the best chance of facilitating therapeutic change.¹²

The resulting behaviourally anchored description of therapeutic judicial behaviour consists of the following elements, as defined as necessary and sufficient by the literature:¹³

- (1) *Introductions*: The judge maintains good eye contact, warm to neutral emotional tone, calmly and firmly sets boundaries and roles in the courtroom, explains the goals of the hearing, and helps the participant to feel confident to participate in the hearing. The judge takes the time to explain the court processes and how to address the judge. If at all possible, the judge and the court participant create a collaborative definition of goals and take turns speaking.
- (2) *Discussion about the problem*: The judge asks neutral and open questions about the context to the reason they are in court, and includes the court participant's own words in the definition of "the problem". If possible, the judge notices and discusses the strengths that are present in the current situation, despite the reasons for being in court. The judge also notices and discusses the points of view and experiences of other participants to the problem.
- (3) *Sentencing*: The judge gives a summary for sentencing or judgment that includes a collaborative definition of "the problem", and incorporates the parties involved in the summary remarks.

⁵ R Waterworth, "Measuring Legal Actor Contributions in Court: Judges' Roles, Therapeutic Alliance and Therapeutic Change" (2019) 28 JJA 207.

⁶ Waterworth, n 5, 207; Waterworth, n 1.

⁷ See, eg, M Herzog-Evans, "Release and Supervision: Relationships and Support from Classic and Holistic Attorneys" (2016) 1 *International Journal of Therapeutic Jurisprudence* 23.

⁸ Karni Perlman, "It Takes Two for TJ: Correlation Between Bench and Bar Attitudes Toward Therapeutic Jurisprudence: An Israeli Perspective" (2008) 30 *Thomas Jefferson Law Review* 351.

⁹ See, eg, E Marziali and L Alexander, "The Power of the Therapeutic Relationship" (1991) 61 *American Journal of Orthopsychiatry* 383; L Fenton et al, "Perspective Is Everything: The Predictive Validity of Six Working Alliance Instruments" (2001) 10 *Journal of Psychotherapy and Practical Research* 262.

¹⁰ Waterworth, n 1; Waterworth, n 5.

¹¹ Waterworth, n 5, 207.

¹² Waterworth, n 5, 220.

¹³ Waterworth, n 5, 221.

The judge describes the responsibility for action as internal to the court participant and describes the situation and summary to include the participant as having choices over their actions. The judge also highlights context and possible supports available. The summary includes an acknowledgment of the possible experiences of other people who are also involved in the problem situation – for example, the victim of a crime.

- (4) *Judicial Communication Skills*: The judge uses open questions, with active listening skills, and attentive and encouraging body language, as well as non-verbal prompts to encourage the court participant to express themselves well. The judge adapts their language and speed of speaking to the language abilities and comprehension of the court participant, and asks questions to check they have understood. The judge facilitates other legal actors present in the court to do the same, so as to ensure that the court participant understands what is being communicated and the processes behind this. Turn-taking and collaborative dialogue occur during the interaction.
- (5) *Judicial Alliance*: The judge has a neutral to warm emotional tone, has open but authoritative body language, and actively allies themselves with the court participant against “the problem” (ie how to stop offending while using drugs, or how to stop behaving in a violent manner towards family members, or how to separate from a partner in the least damaging way for the children and both partners).

The draft measurement scale was designed to capture this behavioural definition in a qualitative way when observing judicial interactions in courtrooms in real time.¹⁴

Overview of LACS Item Development

The items on the LACS map directly onto the behavioural definition discussed above.¹⁵ The scale design followed an inductive process. The desirable characteristics of the judicial interaction were defined with reference to the relevant literature and then scale items were identified to measure each feature of the construct. The scale is designed to be rated by an observer of the interaction, rather than the participants; this approach was adopted because an “observer-rater perspective” was shown to be the most accurate perspective when rating the relevant elements of behaviour.¹⁶ The scale can also be used by magistrates when watching a recording of their own sitting sessions in court. It is not designed to be used by a participant to the interaction in real time, however, due to the limitations in accuracy this usage would introduce.¹⁷

The scale commences with the “Introductions” section, designed to evaluate how the court interaction begins, with questions about the way in which the judge addresses the defendant, how they explain the court process, the judge’s body language and their emotional tone.

The next section, “Discussion about the problem”, relates to the way in which the judge frames the issue that has brought the defendant to court – primarily, have they asked about the background to the situation, the defendant’s history as it relates to the problem, and have they allowed the defendant to participate in a collaborative definition of the problem (as much as appropriate).

Following this is the section evaluating the “Summary or Sentencing Remarks”, which again looks for aspects of a collaborative dialogue designed to enhance therapeutic change. This can include: a collaborative definition of the problem; incorporation of defendant and context into sentencing; externalisation of the problem but internalisation of a sense of responsibility and self-agency; an explanation of the sentencing reasons; and inclusion of the defendant’s wishes and the victim’s experience (if this is available) into the sentencing narrative. This section also evaluates whether the defendant is signposted to or facilitated in accessing any therapeutic or practical resources that have been identified as potentially useful.

Following on from this, the measure evaluates the overall efficacy of the “Judicial Communication Skills” in use throughout the hearing, noting the types of questions that have been used (open/closed),

¹⁴ Waterworth, n 5, 221.

¹⁵ Waterworth, n 5, 222.

¹⁶ Marziali and Alexander, n 9.

¹⁷ Fenton et al, n 9.

the type of listening used (active/passive) and the percentage of times these have been employed. Other effective communication skills are also noted, such as paraphrasing, asking clarifying questions, validation/minimal encouragers, participation, encouragement for the defendant to participate, turn-taking, adaptation of communication style to defendant's abilities, cultural referencing appropriate to the defendant, and offering appropriate choices.

"Judicial positioning/trust/rapport" is the final section. It measures evident aspects of the relationship alliance between the judge and the defendant, including: the emotional tone of the judge overall; body language overall; judicial positioning overall (allied or adversarial); agreement on goals of the hearing; agreement on tasks to be completed during the hearing; and judicial body language as it relates to an allied or adversarial position within the interaction.

The draft scale was then refined by means of a Delphi study. Opinions, feedback and suggestions for improvement were sought, in a systematic way, from experts in the field of therapeutic change with a view to obtaining expert consensus about the appropriateness and usefulness of the items on the scale. This was to ensure the scale accurately measures therapeutic behaviour in a court context and that it is a usable measurement tool. With regard to the latter, the Delphi Study also sought to answer the question of whether, and to what extent, the tool can be utilised to assess legal actors' therapeutic contributions in mainstream court systems in the sentencing context, specifically in Australian magistrates courts (criminal matters) and problem-solving courts.

A broader aim of the study was to facilitate the development of an effective and acceptable measurement scale to capture and qualitatively describe legal actor contributions more generally from a therapeutic perspective. The refined measurement LACS scale could potentially be used across a wide range of legal contexts, including mainstream criminal courts, problem-solving courts and court diversion lists, in both common law jurisdictions and civil law jurisdictions. The development of the measurement scale has started with an initial focus on magistrates, with the possible future expansion to other legal officers.

METHODOLOGY

How Best to Validate the LACS

Validation of a novel measuring tool can be problematic. New measures are often normed or validated on a comparative basis against a similar or parallel measure, or several measures with overlapping constructs. However, such validation was not possible for the LACS measurement tool. There are (at the time of writing) no other similar measurement tools that operate in a comparative environment and measure the same constructs. Measures of therapeutic alliance already in existence – such as the Vanderbilt Therapeutic Alliance Scale, the Working Alliance Inventory-Observer Form and the California Psychotherapy Alliance Scales – reliably measure some overlapping constructs;¹⁸ however, the contexts in which they are applied are significantly different to that of the LACS and so render redundant any attempts to validate the LACS via those parallel measures.

In the absence of a comparative measurement device against which to assess the validity of the LACS, and given the explorative and collaborative nature of the work at this stage in its development, a Classical Delphi research design methodology was chosen to verify its utility.¹⁹ This approach offers not only the opportunity to refine the measurement tool itself, but also provides a rigorous methodology for consulting with "legal actor" experts in the fields of therapeutic change and therapeutic court craft,

¹⁸ V Tichenor and CE Hill, "A Comparison of Six Measures of Working Alliance" (1989) 26 *Psychotherapy: Theory, Research, Practice, Training* 195; R Di Giuseppe, JJ Linscott and R Jilton, "Developing the Therapeutic Alliance in Children-Adolescent Psychotherapy" (1996) 5(2) *Applied and Preventive Psychology* 85; TJ Tracey and AM Kokotovic, "Factor Structure of the Working Alliance Inventory" (1989) 1 *Psychological Assessment* 207; DE Hartley and HH Strupp, "The Therapeutic Alliance: It's Relationship to Outcome in Brief Psychotherapy" in J Masling (ed), *Empirical Studies in Analytic Theories* (Erlbaum, 1983) 1–38.

¹⁹ Neda Milevska-Kostova and William N Dunn, "Delphi Analysis" in L Zaletel-Kragelj and J Bozиков (eds), *Methods and Tools in Public Health: A Handbook for Teachers, Researchers and Health Professionals* (Hans Jacobs Publishing Co, 2010) 433.

in order to develop the measure and its contents on a collaborative basis, with the added advantage of promoting its use.

RATIONAL FOR USING THE DELPHI STUDY METHODOLOGY

Named after the Delphic oracle,²⁰ the Delphi study methodology involves systematic consultation with experts on an aspect of research that usually involves novel ideas or forecasting of some sort.²¹ In this case, it was used to engage with experts in the therapeutic change field – that is, psychologists with experience of facilitating psychological change and magistrates with practical therapeutic court experience. In a nutshell, this research design secures the input of a “wise crowd”²² in the development of a novel research measure or question.

The Delphi study methodology was also considered appropriate because it has been commonly used in medical, nursing and health services research,²³ where there is an element of forecasting or novel research.²⁴ Despite some criticism,²⁵ Delphi panels can generate more accurate approaches than unaided individual decisions.

Risks and Advantages of the Delphi Methodology

The results of this approach are only as valid as the knowledge and opinions held by the participants.²⁶ As such, this methodology carries with it the risk of participants’ personal subjectivism or lack of knowledge about the specific topic having a deleterious effect. Clearly, at its worst, this effect could derail the research process simply because the participants lack the necessary competence to offer useful opinions or feedback, or if the basic assumptions behind their feedback are erroneous.²⁷ Selection of the participants must be based on careful evaluation of their level of knowledge and expertise.²⁸

There is also the risk that experts may oversimplify complex issues, or view issues in isolation rather than as related events.²⁹ Further, the opinions and suggestions given by the experts need to have a high level of consistency and reliability, otherwise their forecasts may not be reproducible or valid. These concerns point to the need to provide a reverification process among participants of their contributions. In the present study, the responses to each questionnaire were fed back in summarised form to the participants for further reflection. The initial questionnaires in Delphi studies also collect qualitative comments, which are fed back to the participants in a quantitative form through a second questionnaire for further comment.³⁰ The Delphi study is therefore an iterative multistage process designed to achieve consensus of an expert group.³¹

²⁰ Felicity Hasson, Sinead Keeney and Hugh McKenna, “Research Guidelines for the Delphi Survey Technique” (2000) 32(4) *Journal of Advanced Nursing* 1008.

²¹ Milevska-Kostova and Dunn, n 19.

²² J Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few* (Abacus, 2004).

²³ See, eg, PL Williams and C Webb, “The Delphi Technique: An Adaptive Research Tool” (1994) 61(4) *British Journal of Occupational Therapy* 153; S Kirk, C Carlisle and KA Luker, “The Changing Academic Role of the Nurse Teacher in the United Kingdom” (1996) 24 *Journal of Advanced Nursing* 1054; JME Gibson, “Using the Delphi to Identify the Content and Context of Nurses Continuing Professional Development Needs” (1998) 7 *Journal of Clinical Nursing* 451.

²⁴ Milevska-Kostova and Dunn, n 19; O Kuusi, “Expertise in the Future Use of Generic Technologies” (Government Institute for Economic Research Finland, 1999) 59.

²⁵ Milevska-Kostova and Dunn, n 19.

²⁶ J Martino, “The Precision of Delphi Estimates” (1970) 1 *Technological Forecasting and Social Change* 293.

²⁷ Milevska-Kostova and Dunn, n 19.

²⁸ S Makridakis and SC Wheelwright, *Forecasting Methods for Management* (Wiley, 5th ed, 1989).

²⁹ Milevska-Kostova and Dunn, n 19.

³⁰ Hasson, Keeney and McKenna, n 20, 1009.

³¹ HP McKenna, “The Delphi Technique: A Worthwhile Approach for Nursing?” (1994) 19 *Journal of Advanced Nursing* 1221; MR Lynn, EL Laman and SP Englehardt, “Nursing Administration Research Priorities: A National Delphi Study” (1998) 28(5) *Journal of Nursing Administration* 7.

This methodology is particularly useful when evaluating a research question in a context where there is considerable existing knowledge, using simple questions within defined project objectives.³² It is also useful in complex, multi-dimensional modelling situations, when supported by additional data from more than one source.³³

Procedure for the Delphi Study: Participant Selection and Recruitment

The literature indicates that anywhere between four and 3,000 panelists may be appropriate in conducting Delphi studies.³⁴ On a practical level, the pool of experts in the present field is not, objectively speaking, particularly large. Additionally, because of the constraints of ethics approval, the participants were further narrowed to Australian participants only. Bearing in mind these limitations, while also ensuring that sufficient data could be obtained, the initial aim was to recruit at least five participants from each of the two occupational categories included in the study, giving a total of 10 potential participants.

Two groups of experts were originally recruited:

- Group 1: five judges/magistrates/other court workers who work in a range of court settings with some expertise in a therapeutic court intervention.
- Group 2: five psychologists with experience in facilitating therapeutic change.

These occupational categories were chosen as Group 1 professionals were assumed to be experts in court craft and court interactions, and Group 2 professionals were assumed to be experts in therapeutic change and the interactional requirements that best facilitate therapeutic change. These groups were later combined due to a high dropout rate in survey return, the reasons for which are discussed below.

The judges and psychologists contacted were from a range of jurisdictions, geographical locations and levels of therapeutic expertise across Australia. This diversity meets the criterion for engaging a “wise crowd” as proposed by Surowiecki,³⁵ which is necessary to ensure the effectiveness of the Delphi model. Additionally, inclusion of the legal community of experts at such a fundamental stage of the development of the measure is likely to facilitate acceptance of the end product by the same community.

Recruitment for both groups occurred via the TJ listserv group, the professional body for psychology in Australia and informal professional networks. An initial letter of invitation was sent to the contact email address for the professional bodies involved and to specific therapeutic courts, as well as those meeting the “expert” eligibility criteria identified from the listserv group and informal professional networks within Australia. A subsequent follow-up letter with more information and a consent form was then sent to participants who indicated interest in participating in the study. The recruitment letters and information sheet are included in Appendix 1. As noted above, initially, five psychologists and five magistrates were approached. The draft measurement tool and the initial survey sent out for the first round of surveys are contained in Appendix 2 and 3 respectively.

The subsequent return rate of surveys was 50%, resulting in an actual participation pool of five experts in total – three psychologists and two magistrates. This rate appears acceptable when referencing the literature with similar methodology.³⁶ The reason for the low return rate is hypothesised to have been

³² M Adler and E Ziglio, *Gazing into the Oracle* (Jessica Kingsley Publishers, 1996); E Cornish, *The Study of the Future: An Introduction to Understanding the Art and Science for Understanding and Shaping Tomorrow's World* (World Future Society, 1977).

³³ European Commission, *Evaluating Socio Economic Development. Sourcebook 2: Methods & Techniques, Delphi Method* (2005) 110.

³⁴ Shane R Brady, “Utilizing and Adapting the Delphi Method for Use in Qualitative Research” (2015) *International Journal of Qualitative Methods* 1; Anthony F Jorm, “Using the Delphi Expert Consensus Method in Mental Health Research” (2015) 49(10) *Australian and New Zealand Journal of Psychiatry* 887; HA Linstone, *The Delphi Technique. Handbook of Futures Research* (Greenwood, 1978) 271–300; F Hassan, S Keeney and H McKenna, “Research Guidelines for the Delphi Survey Technique” (2000) 32 *Journal of Advanced Nursing* 1008.

³⁵ Surowiecki, n 22.

³⁶ See, eg, Hanna Tolonen et al, “Differences in Participation Rates and Lessons Learned About Recruitment of Participants – The European Health Examination Survey Pilot Project” (2015) 43 *Scandinavian Journal of Public Health* 212.

due to an extremely difficult and unusually prolonged bushfire season during the Australian summer of 2019–2020, which affected many of the communities where the prospective participants lived and worked, thereby making research participation potentially of low priority. Results from the five returned surveys were analysed for median, mode and mean responses. Most items were well endorsed; however, some were identified as being too similar to each other, or less relevant. The participant's qualitative comments and feedback were incorporated directly into the design of the scale. Items that had been identified as too close in meaning were removed, the likert rating scale throughout was changed from a seven-point scale to a five-point scale to give it a higher discriminative value, and several items were reworded to be more precise in meaning (see discussion below).

The survey questionnaire was also revised, with items on the survey form that were no longer included in the measurement tool being removed. This revised measurement tool was then returned to participants for a second round of surveys. Survey participants also received individualised feedback as an attached word document on how their survey responses compared to the (anonymised) group responses on each item.

The second round of surveys was also unfortunately timed due to the worldwide COVID-19 outbreak that occurred from early 2020 onwards – an event that may explain why, of the five surveys sent out, only one was returned. An alternate explanation, one endorsed by the respondent who returned their survey, is that survey participants did not see the need to add any extra feedback. The returned survey endorsed the items as they were, without suggesting any further changes.

RESULTS

The feedback from the participants in this study was invaluable, and the results of the survey are included in Appendix 3. The changes that were made to the first LACS measurement tool were:

- (1) The deletion of items that appeared to overlap and be measured elsewhere (1e–1h, 2a, 2b).
- (2) Changing all likert scales to a five-point scale (1i, 5a–5d).
- (3) The addition of an extra item (3hi).
- (4) Slight rewording of some scale items to be more precise (see discussion below).

The revised version of the measure is included in Appendix 4.

DISCUSSION

At present, the LACS is a qualitative measurement scale designed to capture the contributions of legal actors (particularly magistrates) within a courtroom setting (particularly therapeutic and problem-solving courts). Prior to the LACS, there was no published measure of therapeutic contributions from an objective perspective. However, there are many compelling reasons to have such a measure available to the judiciary, researchers and decision-makers.³⁷ As discussed above, the constructs captured by the measurement scale, and the items covering these constructs, map directly onto the therapeutic change literature, incorporating the available research on TJ, PJ, LJ, sociological theory and psychological therapy outcomes as to what constitutes desirable judicial behaviour that has the best chance of facilitating therapeutic change.³⁸

However, it is useful to consider whether, from a methodological perspective, the dropout rate for this study undermines its generalisability. Recent research has found that a 50% return rate (or lower) is quite common in the social sciences, and does not preclude the study continuing, however, there is a counter argument that the Delphi technique in this situation was potentially flawed as the usual Delphi iterative process could not be engaged, leading to a lack of confidence in reliability and validity.³⁹ Interestingly,

³⁷ Waterworth, n 1; Waterworth, n 5.

³⁸ Summarised in Waterworth, n 5, 208–218.

³⁹ For example, participation rates ranged from 16% to 57% for men and 31% to 74% for women of the studies, as reviewed by Tolonen et al, n 36. , for information regarding the countering perspective regarding response rate, please refer to R Boulkedid et al, "Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review" (2011) 6(6) PLoS ONE e20476, doi:10.1371/journal.pone.0020476; M Niederberger and J Spranger, "Delphi Technique in Health Sciences: A Map", *Frontiers in Public Health* (22 September 2020) <<https://doi.org/10.3389/fpubh.2020.00457>>

the first round of respondents gave very similar (nearly unanimous) responses to questions, which could be viewed as either unanimous agreement, or a problem with homogeneity of respondents. The participants' lack of response to the second round could be interpreted in various ways: for example, potentially as not needing to add anything further, or potentially discontinuing research participation due to extreme role pressure that judicial officers experienced due to the local conditions at that time.

It is a truism that two heads are better than one; in this case, even though the participation numbers were low due to unforeseen and unprecedented world events, five heads are still definitely better than one.

Commentary provided by participants on the tool via the survey was particularly useful in gaining insight into how best to improve the measure. First, participants observed that several items overlapped in the constructs they were designed to measure 1e–1h:

- 1e: The judge addresses the defendant by name.
- 1f: The judge explains to the defendant how to address the judge.
- 1g: The judge explains the goals of the hearing: defines goal for defendant, for court and for society.
- 1h: The judge explains the rules for participating in the court using language the defendant can understand.

These items seemed to be adequately covered in items 1a and 1b:

- 1a: Establishes context: The judge explains to the defendant how the court works and the aims of the hearing.
- 1b: The judge addresses the defendant personally.

Similarly, items 2a “The judge asks about the defendant’s background” and 2b “The judge asks about the defendant” appeared to be adequately and more precisely dealt with by items 2c “The judge asks about the context of the defendant’s life and the reason they are in court” and 2d “The judge includes the defendant’s words into statements and questions”. Items 2a and 2b were removed from the revised LACS based on this feedback.

Feedback from the participants also led to structural changes in the measurement scale, whereby all likert scales were changed from a seven-point scale to a five-point scale for ease of use and scoring (items 1i, 5a–5d). A five-point likert scale provides a clearer differentiation between poor and excellent performance on the scale, and therefore has higher discriminative value; it can also more easily be converted into a percentage rating.

Feedback by survey participants for items 3b and 3c recommended separating out an offender from their actions, while encouraging them to accept responsibility for their behaviour – this is a core therapeutic change technique common to multiple strands of therapy, particularly Narrative Therapy.⁴⁰ This feedback also led to the addition of an extra item – 3hi “The judge builds a coherent story or narrative of the context of the crime” – to aid in the measurement of this style of narrative intervention.

Item 4 feedback recommended not letting TJ court craft (or culture, prior trauma or socioeconomic status) in any way justify criminal behaviour, nor allowing a defendant to deliberately cause distress to others in court – again a key component of maintaining safety and boundaries within a hearing setting.⁴¹

Additionally, item 4j received the useful commentary that providing too many choices for defendants could be unhelpful or overwhelming in a court setting. Instead, the item was reworded to reflect the desirability of offering fixed choices, which is a technique effective for those under stress as it facilitates their decision-making. Fixed choices offered within useful parameters by those with (expert) knowledge about the desirable behavioural options available can be particularly useful within a courtroom environment.⁴²

⁴⁰ See, eg, P Jenkins, *Invitations to Responsibility: The Therapeutic Engagement of Men Who Are Violent and Abusive* (Dulwich Centre Publications, 1990); King, n 1, 21.

⁴¹ King, n 1.

⁴² See, eg, SS Iyengar and MR Lepper, “When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?” (2000) 79(6) *Journal of Personality and Social Psychology* 995; B Scheibehenne, R Greifeneder and PM Todd, “Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload” (2010) 37(3) *Journal of Consumer Research* 409; AG Miller, “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information” (1956) 63(2) *Psychological Review* 81.

Generally, feedback was positive about the possibility of the measure being available to the court community, and its potential positive application. Some feedback went beyond the scope of this study and the LACS in its current form, and could not therefore be incorporated within the refinement of the measure. For example, it was suggested that different types of hearings might be taken into account (ie different versions of the LACS for different types of hearings or courts), that the length of the hearing be measured and that an assessment of the behaviour of the defendant be included. Two respondents also suggested non-specific changes to the measure's format so as to promote its use.

The revised version of the measure with the suggestions included is available in Appendix 4.

The LACS has evolved and developed via consultation with those with the necessary and relevant knowledge. The benefit of this process has been twofold: first, to improve the measure; and second, to make sure that it is regarded as useful to the group of people most likely to use it. Given the low survey return rate, it would be worthwhile repeating the Delphi study with a larger pool of participants in Australia, and also in other jurisdictions, to further refine and validate it, and to ensure that the scale is refined with feedback from across different legal systems.

The scale design followed an inductive process, whereby the desirable characteristics of the judicial interaction were defined by reference to the literature and then scale items were designed to measure the resulting constructs via observable behaviour and communication. The next stage in the development of the measurement tool is to trial its use with magistrates in court and in training sessions, as well as utilising it in an independent research pilot study. However, this is beyond the scope of the current Delphi study. In the meantime, the LACS could be included in training for magistrates in therapeutic court craft, and used as a self-evaluation tool when revisiting recordings of court sessions.

Further thought could be given to ease of user application, and to developing a user instruction guide for each item in the measure, to ensure user consistency in its application. The measure could also undergo further qualitative refinement, and could be studied to develop its psychometric properties and enable the generation of quantitative data. To develop into a more psychometrically robust measurement tool, the current measure would need to progress through the stages required to develop these psychometric properties, designed to determine the discriminant and convergent validity, the inter-rater reliability and factor/cluster analysis of the data obtained from the scale.⁴³ The mechanics of these steps are beyond the scope of the discussion here, and await further exploration in future research.

Pilot testing linking the scale measures to court participants' outcome measures would be a further useful stage, as would the process of test validation and norming over multiple instances of application.

CONCLUSION

This article has charted the refinement of the first literature-derived, behavioural measurement scale for legal actors within a courtroom setting, from an objective, therapeutic perspective. The present study is the first step in a potentially long process to develop the scale's psychometric properties.

After vetting (and significant encouragement) from an expert portion of the therapeutic change community, the LACS is now available as a measurement tool for incorporation into court outcomes research design, for magistrate court craft development and for further studies aiming to link measurements of court craft to therapeutic outcomes.

There is considerable scope for this scale to be used in its current form as a qualitative measurement tool to support the mainstreaming of the TJ movement. It could also be used to inform professional development planning for the judiciary in consultation with the relevant professional bodies.

Ultimately, this Delphi study should encourage use of the LACS to capture and measure hitherto hidden therapeutic contributions of magistrates (and other legal actors) in court settings, and the therapeutic impact of their court craft. This will necessarily make a useful contribution to advancing the interests of those passing through the court system and in need of therapeutic change.

⁴³ Paul E Spector, *Summated Rating Scale Construction: An Introduction Sage University Papers Series. Quantitative Applications in the Social Sciences; No. 07-082* (Sage Publications Inc, 1992).