

Closing the evidence to practice gap: how can we embed procedural justice principles into complaint responses to prisoners?

Jo Voisey is a Behavioural Scientist in the MoJ Evaluation and Prototyping Hub, Flora Fitzalan Howard, Dr Helen Wakeling and Nicola Cunningham work in HMPPS' Evidence-Based Practice Team, Scott Lane is the Complaints Clerk at HMP Buckley Hall, and Jayne Kirkpatrick is currently the Deputy Governor at HMP Kirkham but was formerly Deputy Governor at HMP Buckley Hall.

This is an exploratory piece of research to test whether a new process and a template had the potential to change staff practices when replying to prisoner complaints. Based on the body of evidence on the benefits of procedural justice (PJ), HMPPS updated their policy in 2019¹ to include the use of these principles in complaint responses. However, we know from a behavioural perspective that changing everyday habitual practices can be difficult.

As part of this research, we developed a tool that enabled us to quantify the amount of procedurally just language in complaint responses and infer simplicity, comprehension and congruence.

We conducted a retrospective one group pre- and post-audit on complaint responses at one Male Adult Category C Prison — HMP Buckley Hall and analysed the text using our newly developed tool, text mining packages and readability measures.

We demonstrated that the approach shows promise in changing behaviour over a sustained period (1 year and 4 months) and is worth testing more rigorously. The proof-of-concept approach has been beneficial: we have evidenced that there is potential to increase the use of PJ within complaint responses, we have identified some unintended consequences and we have tested a tool to quantify the amount of procedurally just language within complaint responses.

The Evidence to Practice Gap

In healthcare, it takes on average 17 years to get Evidence Based Practice (EBP) incorporated into routine practice, and only about half of EBPs ever reach widespread clinical use.² The Evidence-Practice gap is defined as 'The failure of clinicians to adopt proven practices that enhance outcomes for patients.'³

There is a large and robust international evidence base on the importance of PJ to prisoners.⁴ When people feel treated in procedurally just ways it contributes to a host of better outcomes in a relatively simple, swift and economic way. However, there is an absence of research on how to change practices in prison to improve perceptions of PJ. The habitual nature of daily routines means that knowledge of the importance of PJ is likely to be insufficient.

We have looked to other areas, outside of the Criminal Justice System, to learn how we might increase the pace of embedding EBP in prison settings. There has been a lot of work in healthcare to understand the behavioural barriers and facilitators of uptake of EBPs into routine practice. A recent systematic review suggests that interventions should focus on physical and social opportunities and psychological capability (see Table 1).⁵

-
1. HMPPS (2019). Prisoner Complaints Policy Framework. <https://www.gov.uk/government/publications/prisoner-complaints-policy-framework> [accessed June 2022].
 2. Morris, Z., Wooding, S., & Grant, J. (2011) The answer is 17 years, what is the question: understanding time lags in translational research. *Journal of the Royal Society of Medicine*, 104, 510-20.
 3. Free Medical Dictionary by Farlex. Evidence-practice gap | definition of evidence-practice gap by Medical dictionary (thefreedictionary.com) [accessed 29 April 2022].
 4. Fitzalan Howard, F., & Wakeling, H. (2020) People in Prisons' Perceptions of Procedural Justice in England and Wales. *Criminal Justice and Behavior*, 47(12), 1654-1676.
 5. McArthur, C., Bai, Y., Hewston, P., et al. (2021) Barriers and facilitators to implementing evidence-based guidelines in long-term care: a qualitative evidence synthesis. *Implementation Science*, 16, 70.

Table 1: Barriers and facilitators to the implementation of EBP

Behavioural construct	Barriers	Facilitators
Psychological capability: <i>knowledge or psychological skills, strength or stamina to engage in the necessary mental processes</i>	Knowledge gaps	Adequate knowledge and education
Physical opportunity: <i>opportunity afforded by the environment involving time, resources, locations, cues, physical affordance</i>	Time constraints and inadequate staffing Cost and lack of resources Resident complexity Compromised communication and information flow Staff turnover Competing priorities Guideline complexity and associated workload Impractical guidelines	Well-designed strategies, protocols and resources Adequate services, resources and time Innovative environmental modifications
Social opportunity: <i>opportunity afforded by the interpersonal influences, social cues and cultural norms that influence the way that we think about things</i>	Lack of teamwork Lack of organisational support Inconsistent practices Reactive approach	Leadership and champions Support and coordination among staff Involving residents and families Good communication and information flow

There is limited evidence on how to adopt PJ practice in a Justice context. We found one example in a related context: the Police Service in Queensland Australia conducted a randomised control trial to examine the impact of a procedurally just routine random breath test process compared to business as usual practices.⁶ The purpose was to understand whether a defined practice that was built on PJ principles could shape what people think about the police during the encounter and their view on the legitimacy of the police more generally.

Police in the experimental arm were given a simple 'cue card' with reminders to follow a PJ script when completing the random breath test. Citizens in both arms (i.e. the experimental and control groups) were asked to complete a questionnaire using validated measures from legitimacy literature. Citizens who perceived the encounter to be more PJ had more positive views of police, which were related to increases in general perceptions of PJ, legitimacy, satisfaction and cooperation. This research shows that with the use of specific tools it is possible to increase perceptions of PJ.

Prisoner complaints in HMPPS

In the 12 months to 31 March 2021 there were approximately 178,100 complaints from prisoners across the prison estate in England and Wales.⁷ The Prison and Probation Ombudsman (PPO) investigated 1682 complaints. They upheld 30 per cent of the cases. In their annual report, the PPO noted that 'there are still too many examples of careless or policy-non-compliant complaint responses.'

Although not well-tested empirically, internal reviews, external reports (such as from the PPO) and anecdotal reports from within HMPPS suggest that the complaints system is one process in prisons which is perceived to be procedurally unjust. In 2019, HMPPS updated the complaints policy, and based on international and national evidence, some of the additions now required PJ principles to be reflected in all responses to prisoners. Specifically, regarding PJ, the policy was updated to contain explanations as to why PJ is important in prison settings, with examples of what this would look like in practice. The policy also

6. Mazerolle, L., Antrobus, E., & Bennett, S. (2013) Shaping citizen perceptions of police legitimacy: a randomised field trial of procedural justice. *Criminology*, 51(1), 33.

7. Prisons and Probation Ombudsman (2021). Annual Report 2020/21.

contained an introductory section on the new complaints response form to prompt people to consider PJ as they were responding.

Procedural Justice

PJ is the extent that people perceive that they are treated in a just and fair manner by people in authority. The theory argues that if people experience process and procedures to be fair and just then they are more likely to view the law and authority figures to be legitimate.⁸ In turn, this leads to increased compliance and commitment to obey the law.⁹ PJ works even if the outcome of the decision or process is not in the person's favour, as although outcomes are obviously important, the literature suggests that it is the perception of the process that matters more.

PJ perceptions comprise four principles:

- ❑ **Voice** — people need an opportunity to give their side of the story and to feel that they have been listened to and their concerns heard.
- ❑ **Neutrality** — decisions are made from a starting point of neutrality and that rules are interpreted and applied consistently and transparently.
- ❑ **Respect** — people feel that they are treated courteously and with respect by authority and that their rights and dignity are respected.
- ❑ **Trustworthy Motives** — people need to trust those in authority and believe that they act in everybody's best interests.

Evidence on Procedural Justice in a Prison Context

Previous studies, both in England and Wales and internationally, have linked PJ perceptions to a series of important outcomes for prisoners, including mental health and wellbeing, misconduct and violence, and

reoffending after release. Some of these studies have used particularly robust research designs (enabling causal conclusions to be drawn), and collectively the similarity of findings across studies implies that we can be confident in our understanding.

In summary, prisoner perceptions of procedurally unjust treatment have been associated with depression, anxiety and distress, and self-reported mental health symptoms, self-harm and attempted suicide, and vice versa.^{10 11 12 13 14} Similarly, such perceptions are associated with significantly higher rates of self-reported and officially recorded rule-breaking, including assaults.^{15 16 17} And finally, the one study available investigating recidivism outcomes, reported that prisoners who feel treated fairly and respectfully in prison are significantly less likely to be reconvicted within 18 months of release.¹⁹

Incorporating PJ in Complaint Responses

Based on recommendations by the HMPPS Evidence-Based Practice Team and requirements in the updated Complaints policy framework, the Complaints Clerk and Deputy Governor at HMP Buckley Hall designed a new process and complaint template ('the prototype') to improve their local practices. It had four main components (see below and Figure 1).

1. Reflection workshop for senior staff and those responding to complaints run by the Deputy Governor on PJ including good/bad examples. In addition, the complaint responses being written in a PJ way, the expectation was set that each prisoner is to be spoken to before a reply is issued.
2. Complaints Clerk sends reminder checklist and template with the complaint to the responder to make it easy to follow the new process. The template provides a letter response format with specific entries which should be tailored for each response, intended

8. Lind, E., & Tyler, T. (1988) *The social psychology of procedural justice*. Plenum Press.
9. Tyler, T (1990) *Why people obey the law*. Yale University Press.
10. Gover, A., MacKenzie, D., & Armstrong, G. (2000) Importation and deprivation explanations of juveniles' adjustment to correctional facilities. *International Journal of Offender Therapy and Comparative Criminology*, 44, 450-466
11. Liebling, A., Durie, L., Stiles, A., & Tait, S. (2005). Revisiting prison suicide: the role of fairness and distress. In A. Liebling & S. Maruna (Eds.) *The effects of imprisonment*. Cullompton, Devon: Willan Publishing
12. Slotboom, A., Kruttschnitt, C., Bijleveld, C., & Menting, B. (2011). Psychological wellbeing of Dutch incarcerated women: importation or deprivation? *Punishment and Society*, 13, 176-197
13. Beijersbergen, K., Dirkzwager, A., Eichelsheim, V., et al (2014). Procedural justice and prisoners' mental health problems: a longitudinal study. *Criminal Behaviour and Mental Health*, 24, 100-112
14. See footnote 4: Fitzalan Howard, F., & Wakeling, H. (2020)
15. Butler, M. & Maruna, S. (2009). The impact of disrespect on prisoners' aggression: outcomes of experimentally inducing violence-supportive cognitions. *Psychology, Crime & Law*, 15, 256-250
16. See footnote 4: Fitzalan Howard, F., & Wakeling, H. (2020).
17. Beijersbergen, K. A., Dirkzwager, A. J. E., Eichelsheim, V. I., & Van der Lann, P. H. (2015). Procedural Justice, anger, and prisoners' misconduct. *Criminal Justice and behaviour*, 42, 196-218
18. Bieri, D. M. (2013). Procedural justice and prison violence: examining complaints among federal inmates 2000-2007. *Psychology, Public Policy and Law*, 19, 15-29
19. Beijersbergen KA, Birkzwager AJE & Nieuwebeerta P (2016) Reoffending after release: does procedural justice during imprisonment matter? *Criminal behaviour and Mental Health*, 43, 63-82

to demonstrate the four principles of PJ. The template is a mix of suggested content (some of which is meant to be adapted depending on the context) and cues for specific types of content to be included, such as '[include here evidence considered]'.

3. Complaints Clerk quality assures proposed replies and escalates to Deputy Governor if deemed not to be of high enough quality.
4. If needed, Deputy Governor has coaching conversation with staff to improve the PJ content of their complaint response before it is issued.

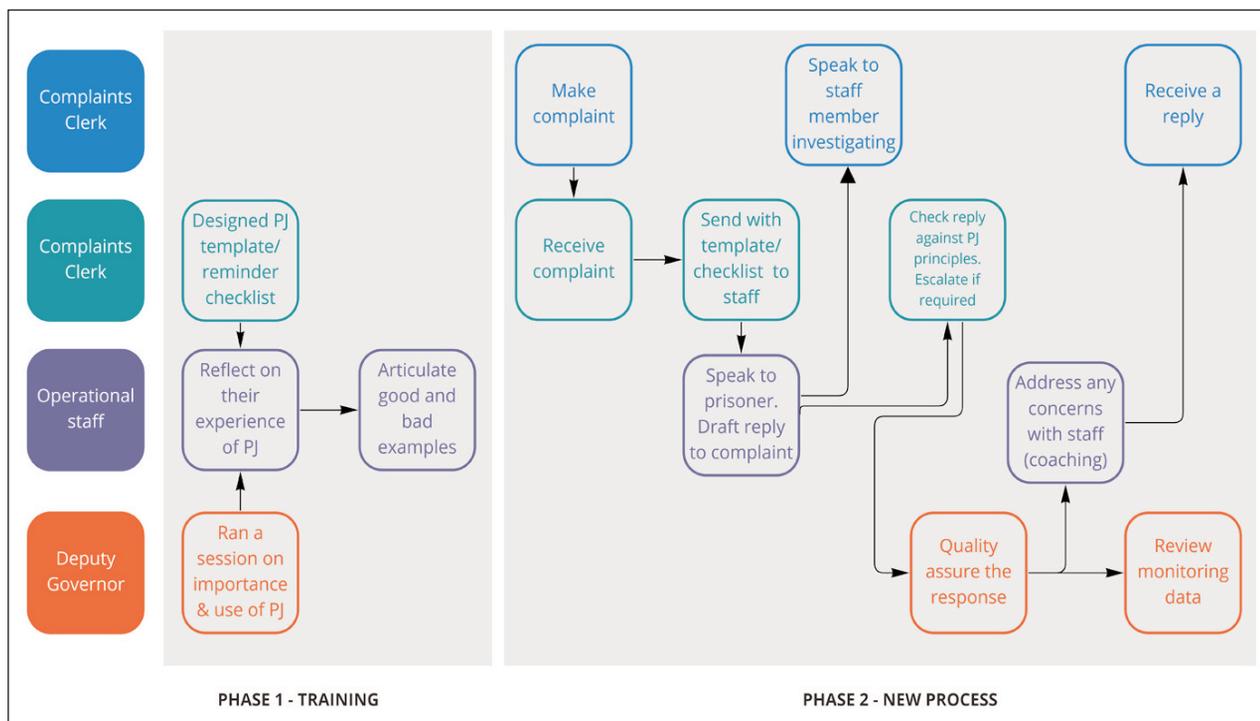


Figure 1: The prototype — Process and Template designed to incorporate PJ into complaints process.

How change might happen

Even with the best intentions, individuals can struggle to change. Previous research into incorporating PJ principles into disciplinary adjudications²⁰ highlight some of the problems²¹:

1. People lacking self-awareness:

'What struck me on the training was that everybody thinks they're really good already. We all think we're already doing it, don't we, and we can't be'

2. Making more abstract concepts applicable to real life:

'...where you'd drawn out examples of good and bad [practice], I found that the most

powerful bit really because you can understand it yourself.... and you can really identify with that.'

3. Remembering to do it:

'[we need] a visual reminder... a piece of paper that gets stuck to a desk'

In the field of psychology, it is believed that many of our daily actions are controlled by habits and impulses driven by short-term rewards.²² We perform routine tasks on autopilot and as the trigger is sub-conscious, it can be hard to execute a different action, even when intended.

However, in the field of sociology, habit is viewed slightly differently. It is defined as a series of social practices rather than discrete behaviours.²³ In social

20. Disciplinary adjudications is a process in English and Welsh prisons for responding to more serious alleged misconduct. For more information, see PSI 05/2018 <https://www.justice.gov.uk/downloads/offenders/psipso/psi-2018/psi-05-2018-prisoner-discipline-procedures-adjudications.pdf>

21. Fitzalan Howard, F. & Wakeling, H. (2021). The experience of delivering 'Rehabilitative Adjudications' in English Prisons. *Psychology, Crime & Law* <https://doi.org/10.1080/1068316X.2020.1850726>

22. Gardner, B. (2015). A review and analysis of the use of habit in understanding, predicting and influencing health-related behaviour. *Health Psychol Rev*, 9(3), 277-295.

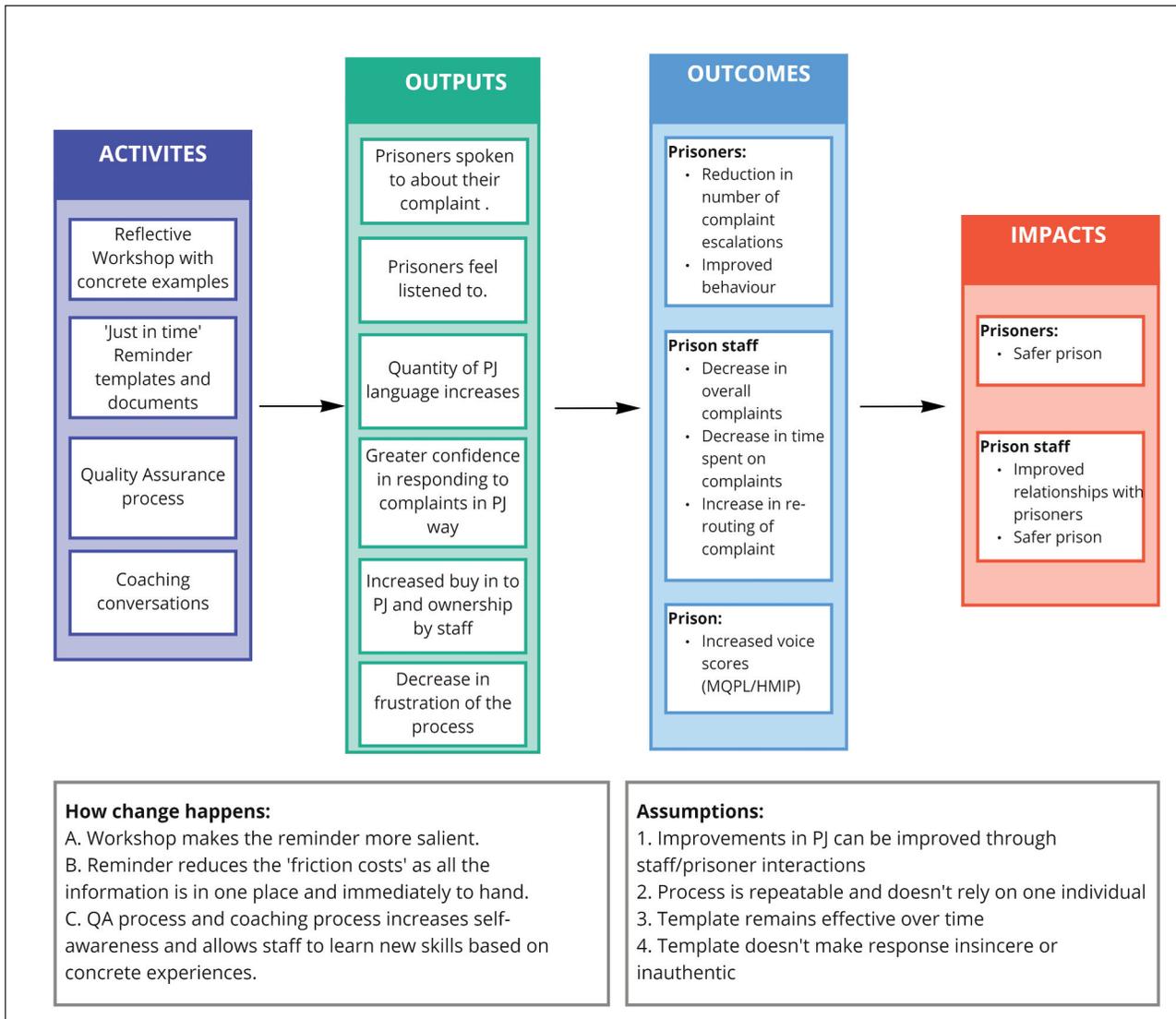
23. Berger, P. L., & Luckmann, T. (1966). *The social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Anchor Books

practice theory, the habit is the whole practice which means that to change routine it is essential to tackle the elements that sustain them. This approach tends to less on individual motivations and beliefs and more on making structural changes that make it easier for the individual to adopt the 'new' routine.

The barriers and facilitators identified in Table 1 from healthcare research appear to be in line with

sociology's view of changing habits: that it is the interplay between structures and processes to provide the physical and social opportunity which are important rather than individual motivations and beliefs. We used this definition as the overarching theory for how change might happen through a focus on process to facilitate required behaviours to occur.

Figure 2: Our Theory of Change



As a behavioural strategy, sending the template and checklist to the staff member with the complaint is a 'just in time' reminder. Reminders are a well-established 'nudge' strategy in healthcare to help people overcome procrastination and change their behaviour.²⁴ The hypothesis was that the workshop on PJ would increase staff buy-in to the new process and this would make the reminder more salient. We also hypothesised that it would reduce 'friction costs' by making it easier for the individual to complete the

new task; they would not have to remember where the template is or what they were supposed to do — all the information was there in one place just when they needed it. It was envisaged then that the workshop and template would help to overcome the barriers of time constraints, communication and information flow, guideline complexity and associated workload, impractical guideline, lack of organisational support, inconsistent practices, and a reactive approach.

24. Kwan, Y. H., Cheng, T. Y., & Yoon, S. et al. (2020). A systematic review of nudge theories and strategies used to influence adult health behaviour and outcome in diabetes management. *Diabetes and Metabolism*, 46(6), 450-460.

The Quality Assurance (QA) process and coaching practice is based on the premise of Kolb's cycle and that people learn through experience.²⁵ We hypothesised that the QA and coaching would enable the individual staff members to reflect on their concrete experience and form and test new concepts based upon this reflection. We anticipated that this would facilitate EBP into action through improving personal knowledge and education, use of leadership and champions, and support and coordination among staff.

Aim of Evaluation

The HMPPS Evidence-Based Practice Team and the MoJ Evaluation and Prototyping Hub joined up with HMP Buckley Hall to learn from the implementation of their prototype.

Our overarching aim was to understand whether the prototype showed any promise in changing practice in responses to complaints. We wanted to test some of the assumptions that underpin our model as well as seeing if PJ language in complaint responses increased.

Primary Question

1. Did the template change the amount of PJ language in written responses?

Secondary Questions

2. How frequently were four PJ Principles incorporated into written responses?
3. How authentic/genuine did the responses feel?
4. Did the impact of the template degrade over time?

Method

This was a retrospective *audit* that measured the content of written responses before and after the new

prototype was introduced. This meant that we could only work with pre-existing data that was available, and that there was no comparison group. Whilst there are limitations with this method, the purpose of this work was to demonstrate proof of concept and to decide whether it is worth taking to a more rigorous method of evaluation. We also wanted to develop and test our data collection tool (see below). This quick retrospective audit allowed us to test this tool with real data and make improvements, so that we could decide whether the tool was reliable enough to be used for further research. The primary outcome was the quantity of PJ content. To measure this, we developed a tool and used summative content analysis²⁶ to rate the amount of PJ content in complaint responses.

PJ Content

The first step was to explore usage of PJ principles. We took each principle and developed criteria for what we would expect to see in a complaint response if that principle was being demonstrated (see Table 2). We created two tiers for each principle as we felt that some PJ criteria were fundamental to any response (tier 1) whilst some criteria were additional (tier 2). Given that it would be impossible (and not particularly useful) to expect each complaint response to meet every criteria, we scored 1 point if the response demonstrated any of the criteria in each tier. This meant that a response could score a maximum of 2 points, one for a tier 1 demonstration, and one for a tier 2 demonstration.

Table Two: Defining the Procedural Justice Criteria in our Tool

	Voice	Neutrality	Respect	Trustworthy Motives
Tier 1	Actively spoke to the individual	Use facts and evidence to make decision Use causal language to demonstrate outcome	Used please and thank you Apologised if late Felt personal vs a 'stock' reply	Showed ownership Used perspective taking Demonstrated empathy
Tier 2	Demonstrate active listening	Balance of probability in decision making Clear about the process	Used non stigmatising language	Demonstrated an understanding of the impact of the outcome Signposted to further information
Overall	Infer if it was: 1. Simple language with no jargon 2. Easy to understand (comprehension) 3. Congruent throughout			

25. Sims, R. R. (1983). Kolb's Experiential learning Theory: A framework for Assessing Person-Job Interaction. *Academy of Management Review*, 8, 501-508.

26. Hsieh, H. F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277-1288.

The second step was to infer meaning and interpretation of content. We wanted to infer whether it was simple and comprehensible but also whether the content was congruent. This last inference was included due to our concerns before we started, that a template can make responses come across as disingenuous and inauthentic. For example, when we know someone who is meant to be helping us is reading verbatim from a script it can backfire and create dissatisfaction. For this part of the tool, we devised an overall score with a point for simplicity, a point for comprehension, and the ability to take away a point if the content was 'jarring' and incongruent. For example, a point might be taken away if the response said 'I hope this has resolved your issue' when it was clear from the content of the original complaint that this wasn't the case.

Altogether, the overall score for each response could range from -1 to 10 and had both a quantitative and qualitative element in the overall score.

Sentiment

Further, we used sentiment analysis of the text to understand the content of responses (using R²⁷). This enables words to be 'tagged' as to whether they are negative or positive, score how negative or positive the words are (i.e. their strength of positivity or negativity), and tag the opinion or emotion associated with the word.

We used the change in the amount of negative to positive words as a proxy for neutrality and were able to analyse the change in trustworthy words, litigious words and uncertainty words. These were proxies for respect and trustworthy motives. This analysis was descriptive so that we could understand and learn what changes in content the new prototype delivered.

Readability

Finally, as the most recent data published by the Ministry of Justice shows that 57 per cent of adult prisoners taking initial assessments had literacy levels below those expected of an 11-year-old,²⁸ we used

readability measures²⁹ to understand what changes happened in terms of sentence length, percentage of difficult words and the reading age of the text. Readability refers to the ease in which a passage of written text can be understood. It is often used in assessing the suitability of a text for an audience. The Flesch-Kincaid Grade Level heuristic was used to measure reading age and it indicates that the text can be read by the average student in the specified grade level (please note that the grade level is US School Grade).³⁰

Procedure

PJ Content

We randomly sampled every 15th complaint response in the period of 14 January 2019 — 10 May 2019 ('the pre period') and every 15th complaint response after the new prototype had been introduced between 15 May 2019 — 30 March 2020 ('the post period'). Each complaint response was scored by two raters with an understanding of PJ. Each rater was 'blind' to the others' score until all ratings had been completed. Raters were randomly allocated their letters using a random number generation in Excel.

If the scores from two raters were within 1 point of each other, the mean was taken as the final PJ score for that letter. Seventy five percent of the scores

fell within this variance. If the scores were greater than 1, the letters were discussed at a moderator meeting and after debate and discussion, agreed by all four raters. These discussions were captured and used to develop the tool and the protocol to increase consistency in its use in any future evaluations. The developments consisted of giving concrete examples and better descriptions of what constituted a criterion being met.

As discussed above one of the outcomes of this proof-of-concept stage has been to develop the tool for future evaluations or to enable prisons to assess their own complaint response content. For tool reliability we needed to meet three criteria:

One of the outcomes of this proof-of-concept stage has been to develop the tool for future evaluations or to enable prisons to assess their own complaint response content.

27. A free to use source code which allows you interrogate data.

28. Ministry of Justice (2021). Prison Education Statistics April 2019 to March 2020. Official Statistics Bulletin.

29. Readability Analyzer (datayze.com)

30. Flesch Reading Ease and the Flesch Kincaid Grade Level – Readable [accessed 13 May 2022]

- ❑ Stability — the tendency for coders to consistently code the same data in the same way
- ❑ Reproducibility — to classify categories in the same way
- ❑ Accuracy — the extent to which the classification corresponds to a norm statistically

We demonstrated reasonable reliability, but we would want to better this in any future work using our tool through updates to the criteria based on what we learnt (see later).

Sentiment Analysis

An analyst with experience of text mining, imported all text in the responses pre and post into R. They tagged the words using different dictionaries to enable multiple analyses.³¹ Pre and post periods were compared.

Readability Analysis

Text was imported into Readability Analyzer (datayze.com) and the different metrics captured. Pre and post periods were compared for sentence length, percentage of difficult words and school grade (reading age).

Findings

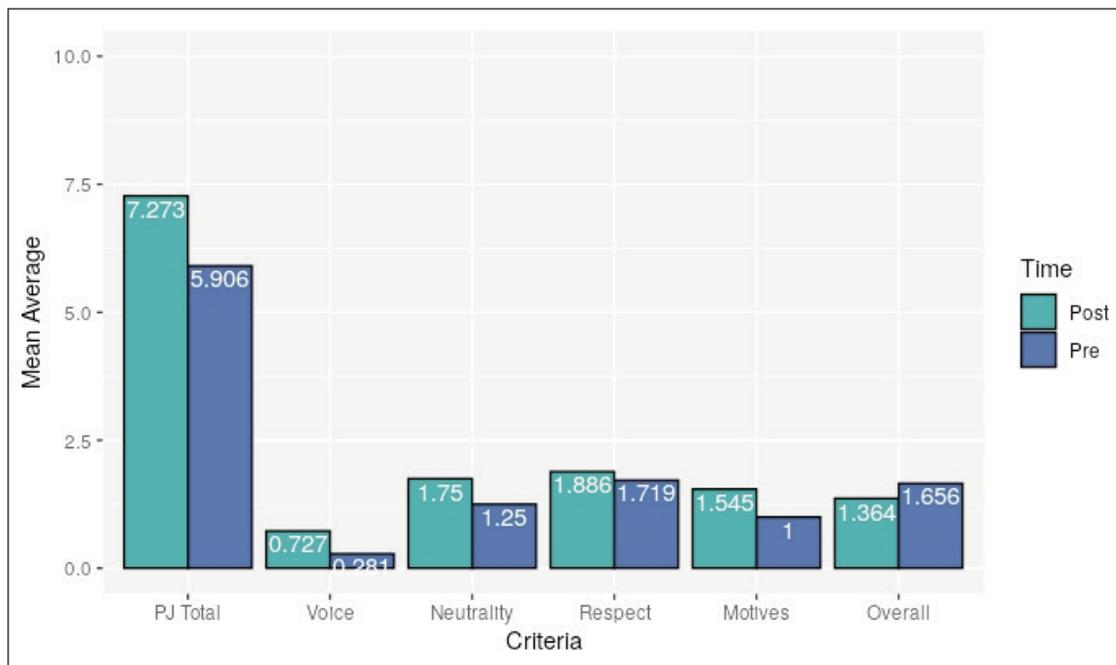
Procedural Justice Content

Our overall aim was to understand if this prototype showed promise — we had a very small sample size so it wasn't powered to detect statistically significant changes. The following results are descriptive and enable us to understand the potential and direction of any change and establish baseline levels.

We found that overall PJ content increased from an average score of 5.91 (out of 10) to 7.27 (see Figure 3). This is an increase of 23 percent. In addition, all the individual PJ principle scores increased although there was variation in baseline levels. For example, despite intention to speak to every prisoner prior to replying to the complaint and reflecting this back in the written complaint response, we found the least evidence of this principle being applied in practice.

The only score to decline was the overall score which comprised simplicity, comprehension and congruence. This is because we inferred a consistent incongruence between the content of the reply and signposting for further support. This was rarely tailored to the circumstances and felt consistently inauthentic to the four people reviewing the content. For example, a prisoner who had complained about not receiving a pair of trainers they purchased was signposted to get further support from a listener, a member of unit staff, Healthcare and Drug and Alcohol recovery services, the Independent Monitoring Board or a Governor.

Figure 3: Difference in PJ content in the pre and post period.



Note: PJ total has a maximum score of 10 and is made up of adding together voice, neutrality, respect, motives and overall, each of which have a maximum score of 2.

31. The dictionaries used were affinn, bing, Loughran and Nrc Function reference • textdata (emilhvifeldt.github.io) [accessed 13 May 2022]

Sentiment Analysis

We used overall sentiment change between positive and negative words as a proxy marker for neutrality (Figure 4.1). This involves taking away the number of negative words from positive words to create a polarity score. In the pre period the polarity was -16 and in the post period was -2. This means that the post period language was more balanced. In other analysis, we looked at the strength (valence) of these positive and negative words. There was a huge increase in the number of positive words with a small to medium positive valence. These are words like: fair, gain, need, hopeful, better, confidence, determined, encourage, progress and appreciates.

In Figure 4.2, we examined the change in emotional content of words in pre and post responses. The largest increase was in the proportion of 'trust' words (such as responsible, team, system, authority, understanding) and the largest decrease was in 'sadness' words (such as late, unfortunate, error, unacceptable, unable).

Finally, we used a different type of dictionary to demonstrate the change in opinionated words. The findings mirrored the previous changes discussed in positive and negative words but also showed a decrease in the proportion of 'uncertainty' words (such as doubt, confusion, risk, presuming) and litigious words (such as adjudication, claim, appeals and regulations).

Figure 4.1: Changes in polarity of words in the pre and post period

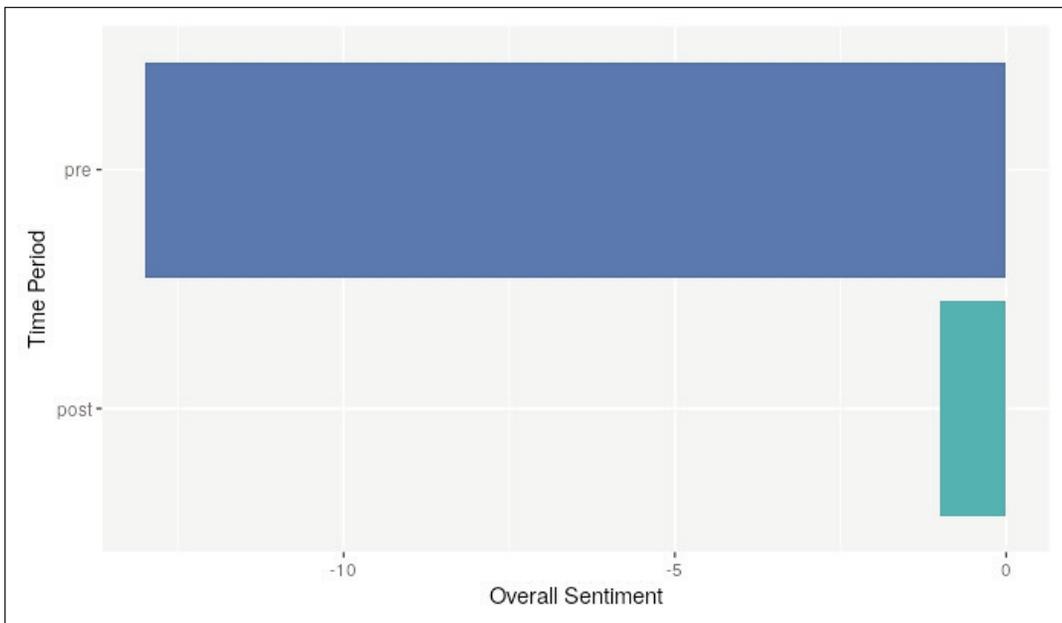
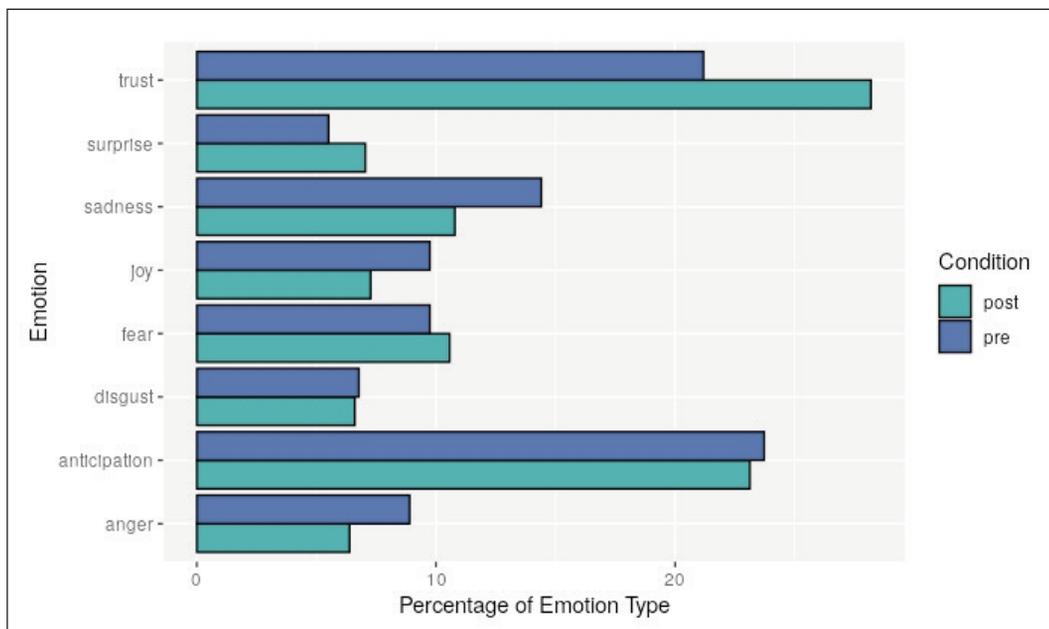


Figure 4.2: Changes in the emotional content of words in the pre and post period



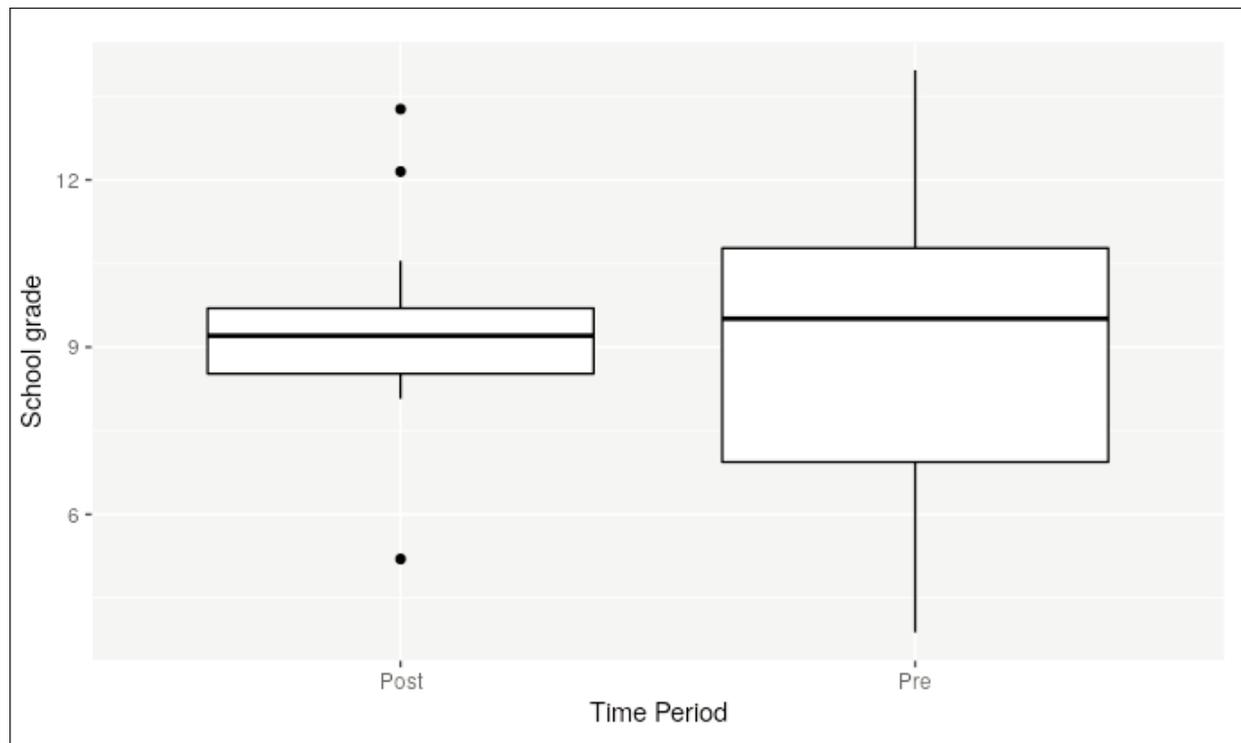
Readability

Whilst the overall average (denoted by the thick line in the middle of each box in Figure 5) in school grade didn't change very much at all between pre and post phases, there was a decrease in the variance of distribution of school grade which means that responses in the post period were more consistent (denoted by the size of the boxes and length of the

lines from the boxes being smaller). The average school grade was 9 (US system) which equates to 14-15 year old reading age.

We also looked at sentence length and percentage of difficult words. It should be noted that there was a slight increase in the average number of difficult words used, which is a concern and reading age is still too high. This is something that we discuss later.

Figure 5: Change in reading age



Discussion

This study expands on the work done in HMPPS to develop PJ practice to improve outcomes. Additionally, and perhaps more importantly, the work informs broader operational understanding of how we might develop and test different techniques to change practice and close the evidence to practice gap.

1. Did the template change the amount of PJ language in written responses?

Our primary question was to understand whether the new prototype could change the quantity of PJ language in complaint responses. We have been able to confirm that changes in content appear to be changing in the desired direction with a positive increase of over 20 per cent in PJ content. We were aware that inter-rater variability was a bit lower than we would have liked. We have updated the criteria for some principles to make it more reflective of what we learned through this exploratory study.

For example, in the neutrality principle we had four criteria:

- Tier 1: Use facts and evidence to decide
- Tier 1: Use causal language to demonstrate the outcome
- Tier 2: Balance of probability in decision making
- Tier 2: Clear about the process

We found that often we were scoring the same content for 'using facts and evidence' and 'being clear about the process' which meant that we were double counting for one piece of evidence. We felt that this didn't accurately differentiate response content. Following moderation discussions, we have changed this principle to have three criteria:

- Tier 1: Clear about the process
- Tier 1: Use causal language to demonstrate the outcome with facts and evidence
- Tier 2: Balance of probability in decision making

We also used specific examples from the complaint responses to update the protocol for using the tool so

that we could give concrete examples of the content/wording that would correspond to each criteria to reduce misunderstanding or individual interpretation by future users.

2. How frequently were the four PJ Principles incorporated into written responses?

There appeared to be a positive change in frequency of each PJ principle although there was some variance between them from the baseline levels. Our R analysis showed that there was an increase in trust words, which links to the principle of trustworthy motives and the overall neutrality of positive/negative words.

The lowest overall score was for voice, both in the pre period and post period. Whilst it has improved, this is a bit of a concern given the emphasis to speak to each prisoner prior to responding within the process. This needs some further investigation to understand if there are any other barriers which have stopped this from happening.

3. How authentic/genuine did the response feel?

There was a slight unintended consequence with the signposting of information in the prototype which was not tailored to the individual response/case. This meant that people were referred to inappropriate or unsuitable additional support. This demonstrates the importance of testing even seemingly simple templates. We would therefore need to update the response template prior to any further implementation of the prototype.

We also found that whilst there was more consistency in readability scores in the post period responses, sentence length was too high, the average percentage of difficult words slightly increased and the reading age was 14-15 years. Again, all of these can be addressed by updating the prototype prior to any further implementation.

4. Did the impact of the template degrade over time?

We examined complaint responses over a period of 1 year and 4 months, from January 2019 to April 2020. There was no noticeable tapering off with the PJ score in the post period. This was one of the assumptions that we wanted to test in our theory of change.

Limitations

The biggest limitation is that there was no 'counterfactual' included in this study, and so we

cannot say with any certainty that these changes wouldn't have taken place without the prototype process.

HMP Buckley Hall has done a lot of work on PJ principles in other areas and scores highly on PJ and on staff/prisoner relationships.³² Again, we cannot say whether this type of approach would result in a similar change in a prison with a different culture, a different category/type of prison or a prison of a different size.

What next?

Given the promising findings, our intention is to test a revised prototype in a different Male Adult Category C prison to see whether we can replicate our findings elsewhere. With this second study, we will be able to test other pathways in the theory of change to try to understand more about how change is happening, with whom and why. We also intend to test the impact more rigorously with a more robust methodology that uses a comparison group who don't have access to the prototype. This will mean that we can be more certain of the causes of any impact. We will also utilise a larger sample in future testing, to ensure we can test for statistical differences in outcomes between the two groups.

If this is successful, we would then want to test scale-ability to see if the findings replicate in different contexts (different sizes/types/categories of prisons) that represent the broader organisation.

From a practical perspective, if the PJ tool proves to be reliable then this would be made available to prisons to monitor how well they adhere to complaints policy. The policy requires quality assurance of complaints each year and this tool could be used by prison staff for that purpose.

We also want to extend our understanding of whether we can 'bundle' multiple PJ practices together and whether this has positive spill over effects on to other everyday conversations. For example, we could look to compare introducing a process on complaints and cell searches in combination and test whether there is any difference on PJ scores compared to just implementing a process on complaints.

The ultimate goal is to improve security and safety in prisons. There is more chance of influencing these outcomes if we bundle PJ practices together but it is very difficult to disentangle the impact of PJ on these outcomes as they are influenced by many factors. However, in the longer term this is something that we would aim to do.

32. HMIP (2019). Report of an unannounced inspection of Buckley Hall by HM Chief Inspector of prisons. Retrieved from Buckley-Hall-Web-2019.pdf (justiceinspectorates.gov.uk)