

On Good Policy

The elements of an organisational policy document that works

August 2025

Author: Hamish Cameron, CEO AccuFind

Contents

Introduction.....	3
1. Currency and control	3
2. Authority and traceability.....	5
3. Obligation clarity.....	6
4. Decidability and executability	8
5. Findability and structure.....	9
6. Understandability.....	11
7. Workability	13
8. Structured and accessible form	15
9. Instrument fitness	16
10. Implementability and assurance	17
11. Assumed knowledge.....	19
The document is not the unit.....	21
Applying this without stopping everything else	22
Conclusion	22

Introduction

Government runs on documents. Not the legislation, which is written once and amended rarely, but the layer beneath it: the policies, procedures, operating manuals, practice guides, information sheets and forms that translate an Act into something a person at a counter can act on. These documents decide whether a licence is granted, whether a payment is made, whether a child is placed, whether a firearm is seized. They are the point at which law meets a citizen.

They are also, in most agencies, unmanaged. Nobody can say how many there are. Many carry no owner, no version and no review date. Some were written for an organisational structure that no longer exists, under an Act that has since been amended forty times. They accumulate because writing a new procedure is the cheapest available response to any incident, and because no one is ever criticised for having written one.

The obvious question is what a good one looks like. It is asked less often than it should be, and the answers on offer tend to be either a template or a readability score. Neither is an account of quality.

This paper sets out one. It draws on plain language standards, government style authorities, records management standards, the empirical literature on comprehension and document design, safety science on written procedures, the clinical guideline appraisal instruments, and the findings of Australian audit offices and royal commissions. It describes eleven elements of a policy document that works, and a set of properties that belong to a document library rather than to any single document. For each, it gives the evidence, what failure looks like in practice, and how to check.

The eleven are ordered by return. The first two fix the most harm for the least effort. The last two are the hardest and the least often attempted.

A note on the evidence. There is no controlled experimental literature on internal government policy documents as such. The material below is drawn from the adjacent fields where the questions have been studied directly: clinical guidelines, consumer contracts, patient information, government correspondence, and written procedures in process safety and aviation. Where a finding is being carried across from one of those settings, the source is named so the reader can judge the distance.

1. Currency and control

A policy document states who owns it, who approved it, what version it is, and when it must next be looked at. None of those dates has passed.

This is the cheapest element to check and the one most often failed. It is also the one that makes every other element assessable, because a document with no owner has nobody to fix it and a document with no review date has no moment at which anyone is obliged to try.

The Australian National Audit Office found that approximately half of the Australian Federal Police's governance instruments were overdue, or possibly overdue, for review at the time of audit.¹ The Audit Office

¹ Australian National Audit Office, *Australian Federal Police's Use of Statutory Powers*, Auditor-General Report No. 43 of 2020-21 (Canberra: ANAO, 2021), para 3.40. https://www.anao.gov.au/sites/default/files/Auditor-General_Report_2020-21_43.pdf

of New South Wales has reported policies past their scheduled review date in successive annual internal controls reports, together with findings that sixteen per cent of large agencies had not updated financial delegations and twenty six per cent held no delegation for writing off certain debts.² These are not exotic failures. They are the ordinary condition of an unmanaged library.

How fast does a document go out of date? The best quantified evidence comes from clinical guidelines, where the question has been studied directly. Shekelle et al.³ found that about half were outdated within 5.8 years, with a confidence interval of 5.0 to 6.6 years, and that only ninety per cent remained valid at 3.6 years. Working at the level of individual recommendations rather than whole documents, Martínez García et al.⁴ found that twenty five of one hundred and thirteen recommendations, or 22.1 per cent, required updating, with validity falling from ninety two per cent at one year to 77.8 per cent at four. A document that has not been examined in a decade is not merely old. It is, on the arithmetic, wrong in several places.

Standards give the answer to what a control block should contain. ISO 9001:2015 clause 7.5 requires identification, description, review and approval for documented information,⁵ and ISO 15489-1:2016 sets the equivalent requirements for records.⁶ The National Archives of Australia makes identifier, creator and date created mandatory metadata for Commonwealth records.⁷ NSW Police attaches a Document Control Sheet to its guides and procedures carrying title, subject, responsible command, authorisation level, security classification, publication date, version number, review date and record number, together with a modification history that references the decision authorising each operational change.⁸

What failure looks like. A four hundred and eighty nine page manual whose only date evidence is the year on the cover, with no version number, no named owner, and a footer date that is simply the day someone last printed it. An information sheet reviewed last year that still describes the pre-amendment version of the Act, because the review was a date stamp rather than an examination.

How to check.

Every document carries ten extractable fields: unique identifier, instrument type, owner role, approver, approval date, version, effective date, next review date, status, and revision history. Report the percentage of the library complete on each field.

Every document has a next review date, and it has not passed. Report the count, the proportion, and the median months overdue.

² Audit Office of New South Wales, *Internal Controls and Governance* (Sydney: Audit Office of NSW, 2023, 2024 and 2025). <https://www.audit.nsw.gov.au/our-work/reports/internal-controls-and-governance-2024>

³ P.G. Shekelle, E. Ortiz, S. Rhodes, S.C. Morton, M.P. Eccles, J.M. Grimshaw and S.H. Woolf, 'Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated?', *JAMA* 286, no. 12 (2001): 1461-1467.

⁴ L. Martínez García, A.J. Sanabria, E. García Álvarez et al., 'The validity of recommendations from clinical guidelines: a survival analysis', *Canadian Medical Association Journal* 186, no. 16 (2014): 1211-1219. <https://www.cmaj.ca/content/186/16/1211>

⁵ International Organization for Standardization, *ISO 9001:2015 Quality Management Systems: Requirements*, clause 7.5 Documented Information (Geneva: ISO, 2015).

⁶ International Organization for Standardization, *ISO 15489-1:2016 Information and Documentation: Records Management, Part 1: Concepts and Principles* (Geneva: ISO, 2016).

⁷ National Archives of Australia, *Australian Government Recordkeeping Metadata Standard: Minimum Metadata Set* (Canberra: NAA). <https://www.naa.gov.au/information-management/standards/australian-government-recordkeeping-metadata-standard/minimum-metadata-set>

⁸ NSW Police Force, *Procedures for Corporate Policy Documents, Version 1* (Sydney: NSW Police Force). https://www.police.nsw.gov.au/__data/assets/pdf_file/0005/861269/NSWPF_Procedures_for_Corporate_Policy_Documents_V1.pdf

The review interval is set by the rate of change in the underlying law, not by a uniform calendar. Pair the interval with event triggers: a legislative amendment in the document's subject area, a machinery of government change, an incident. Any document citing a provision that has been amended, renumbered or repealed since its last review is due, whatever its stated date says.

The revision history records what changed, when, why, and on whose authority. A new version number with no note of what moved is not traceability.

The named owner is a role that still exists. Restructures orphan documents silently.

2. Authority and traceability

Every requirement is either traceable to a cited provision that actually imposes it, or openly labelled as the agency's own choice.

This is a question of lawfulness, not of drafting taste, which is why it sits ahead of everything about style.

Australian administrative law is settled on the point. An agency's policy cannot displace a statute. Where the two conflict, the law applies and the policy does not. Policy must not prevent a decision maker from exercising a discretion that Parliament conferred, and it must not be applied so inflexibly that the individual case goes unconsidered.⁹ In *Green v Daniels*¹⁰ the departmental criteria applied to an unemployment benefit claim were held invalid because they substituted the department's own test for the one the legislation laid down. The applicant met the statutory conditions. She failed the policy. The policy was the thing that was wrong.

The Royal Commission into the Robodebt Scheme is the modern case. Its findings turned in significant part on internal guidance and public correspondence that asserted a legal position the agency could not sustain, applied by staff who had no practical way of telling which parts of what they had been given were law and which were departmental preference.¹¹

That distinction is the working heart of this element. A great deal of what sits in an operational manual is not legislative at all. Age thresholds, evidentiary standards, form requirements, escalation rules and timeframes are frequently agency decisions, and frequently reasonable ones. The defect is not that they exist. The defect is that they are written in the grammar of legal obligation, with no citation and no signal, so that neither the officer applying them nor the citizen subject to them can tell the difference. An officer who believes a policy rule is a statutory requirement will not exercise a discretion they in fact hold. A citizen who believes the same will not contest a decision they could have contested. The Commonwealth Ombudsman makes the same point in the context of automated decisions, where the distinction between what the law requires and what the agency has chosen to build becomes invisible to everyone downstream.¹²

⁹ Administrative Review Council, *Decision Making: Lawfulness*, Best Practice Guide 1 (Canberra: Commonwealth of Australia, 2007). <https://coat.asn.au/wp-content/uploads/2019/11/ARCBestPracticeGuide1Lawfulness.pdf>

¹⁰ *Green v Daniels* [1977] HCA 18; (1977) 51 ALJR 463; 13 ALR 1.

¹¹ Royal Commission into the Robodebt Scheme, *Report of the Royal Commission into the Robodebt Scheme* (Canberra: Commonwealth of Australia, 2023). https://www.monash.edu/__data/assets/pdf_file/0007/3365503/report_of-the-royal-commission-into-the-robodebt-scheme.pdf

¹² Commonwealth Ombudsman, *Automated Decision-making Better Practice Guide* (Canberra: Commonwealth Ombudsman, 2019). https://www.ombudsman.gov.au/__data/assets/pdf_file/0030/109596/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf

Transcription deserves a note of its own. Copying legislative text into a policy document rather than citing it creates a copy that will not be updated when the provision is. Prescribed forms, statutory notices and warnings that must appear in a specified form are the exception, and they must be reproduced exactly. Everything else should point.

What failure looks like. A procedure stating that sports shooters must attend six competitive events a year where the Regulations require twelve. A collector licence described as issued for life where the national agreement sets a five year maximum. An information sheet stating an age based progression rule that appears nowhere in the Act or the Regulations, written in the same voice as the provisions around it.

How to check.

Report the proportion of obligation bearing statements that carry a resolvable citation. The remainder is candidate agency policy, and should be labelled as such.

Every citation resolves against the current consolidated text. Flag references to provisions that have been repealed, renumbered or amended.

The cited provision actually imposes the stated obligation. Adding a threshold, a timeframe or an evidentiary requirement the provision does not contain is the most common and least visible form of drift.

Classify every requirement as legislative, agency policy, or operational guidance, and check the language against the classification. Agency policy written as “the Act requires” or “it is an offence” is a defect regardless of whether the policy is sound.

Where the document governs a statutory discretion, there is a recorded mechanism for departing from it in an individual case. A policy with no departure route fetters the discretion it was meant to guide.

Maintain a register of places where legislative text is reproduced rather than cited, ranked by whether the source has changed since the document was last reviewed.

3. Obligation clarity

From a sentence alone, a reader can tell whether they must do it, how strongly it binds, and who exactly has to do it.

Requirement language is the most concentrated source of ambiguity in policy writing, and the easiest to fix. Two disciplines carry almost all of the value.

The first is a controlled vocabulary. RFC 2119 is the canonical statement that requirement levels must be explicit and used sparingly,¹³ and the Queensland Government Enterprise Architecture reproduces it almost verbatim as Australian government practice.¹⁴ GOV.UK publishes its own short set, distinguishing what a

¹³ S. Bradner, *Key Words for Use in RFCs to Indicate Requirement Levels*, RFC 2119 (Internet Engineering Task Force, 1997). <https://www.rfc-editor.org/rfc/rfc2119>

¹⁴ Queensland Government, *QGEA Document Governance* (Brisbane: Queensland Government Enterprise Architecture). <https://forgov.qld.gov.au/information-and-communication-technology/queensland-government-enterprise-architecture-qgea/qgea-document-governance>

reader must do from what they need to do and what they can do.¹⁵ The choice of vocabulary matters less than its consistency. What breaks a document is using “must”, “shall”, “will”, “is to”, “is expected to” and “it is important that” interchangeably across a hundred pages, so that a reader learns to treat all of them as advisory.

The second is a named actor. Obligations written in the agentless passive voice have no owner. “Applications are to be assessed within twenty eight days” tells nobody that the twenty eight days are theirs. The Australian Government Style Manual requires active voice on the ground that it makes clear who is doing what,¹⁶ and agency policy frameworks that address the point require responsibilities to be attributed to a position, rank or business unit rather than to the organisation at large.¹⁷

There is some evidence that the strength of written requirements predicts what happens next. WellSAT, an instrument developed to score written organisational policy, codes each provision as absent, weakly stated, or strongly stated.¹⁸ Using it across 151 district policies, Schwartz et al.¹⁹ found that stronger and more comprehensive written policies were associated with greater implementation at school level. The association is not proof of cause, and districts that write firmer policies may simply be districts with more administrative capacity. It is nonetheless the closest thing the field has to evidence that requirement language changes behaviour.

A caution runs the other way. Converting every recommendation into a requirement is not an improvement. It creates fettering exposure where a discretion exists, and it creates unworkable obligations of the kind described in element seven. Over-binding is a defect in the same family as vagueness.

What failure looks like. A single document in which the same act is “must” in the procedure and “should” in the accompanying guide. A responsibilities table listing five roles, three of which never appear in the body of the document, while the body imposes four obligations on roles that never appear in the table.

How to check.

Every obligation uses one of a small published set of modal verbs. Count and list every use outside that set, including bare present tense used as a command.

Every obligation has an identifiable actor. Report the proportion expressed in agentless passive constructions.

The same act carries the same modal throughout the document and across related documents.

¹⁵ Government Digital Service, *GOV.UK Content and Publishing Guidance: Use Clear Language*. <https://guidance.publishing.service.gov.uk/writing-to-gov-uk-standards/writing-guidelines/clear-language/>

¹⁶ Australian Government Style Manual, *Plain Language and Word Choice* (Canberra: Australian Public Service Commission). <https://www.stylemanual.gov.au/writing-and-designing-content/clear-language-and-writing-style/plain-language-and-word-choice>

¹⁷ NSW Health, *Policy Development and Publication*, Policy Directive PD2022_047 (Sydney: NSW Ministry of Health, 2022). https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2022_047.pdf. See also Queensland Department of Education, *Policy and Procedure Development and Improvement Cycle Procedure*. <https://ppr.qed.qld.gov.au/pp/policy-and-procedure-development-and-improvement-cycle-procedure>

¹⁸ M.B. Schwartz, E. Piekarz-Porter, M.A. Read et al., ‘Wellness School Assessment Tool version 3.0: an updated quantitative measure of written school wellness policies’, *Preventing Chronic Disease* 17 (2020): 190373. https://www.cdc.gov/pcd/issues/2020/19_0373.htm

¹⁹ M.B. Schwartz, K.E. Henderson, J. Falbe et al., ‘Strength and comprehensiveness of district school wellness policies predict policy implementation at the school level’, *Journal of School Health* 82, no. 6 (2012): 262-267.

A document declared advisory contains no mandatory language, and a document declared mandatory does not bury binding requirements inside advisory prose.

Every role named in the responsibilities section appears in the body, and every role that bears an obligation in the body appears in the responsibilities section.

4. Decidability and executability

A competent officer can tell exactly when the rule applies, and exactly what to do, without ringing someone.

These two properties come from GLIA, the GuideLine Implementability Appraisal, an instrument built specifically to work out why a written recommendation fails to be acted on.²⁰ Decidability asks whether a reader can consistently determine that the conditions triggering a requirement have been met. Executability asks whether the required action is stated specifically enough that different people would do the same thing. They are the formal names for the two complaints every front line officer has about the manual.

The failure patterns are consistent. Conditions expressed in prose with two or more connectives, so that resolving a case requires reading the sentence three times. Enumerated conditions with no statement of what happens when none of them applies. Evaluative terms carrying decision weight with no elaboration anywhere in the document. Instructions built on abstract verbs, where the officer is told to ensure, manage, monitor or address something with no accompanying step, artefact or deadline. Specificity of recommendations is also one of the four AGREE II items that Hatakeyama et al.²¹ found, across 206 guidelines, together explained 72.1 per cent of the variance in overall quality judgements.

A distinction matters here, and getting it wrong produces bad advice. Some evaluative language in a regulatory scheme is deliberate and legally freighted. “Fit and proper” and “as soon as practicable” are statutory terms of art with decades of case law construing them. They are undefined in the legislation because their content comes from authority, not from a definitions clause, and a policy document that tried to define them exhaustively would be fettering the discretion they preserve. The right treatment is to identify where they sit and name the considerations that bear on them.

Agency-invented evaluative language is a different matter. “Adequate storage”, “a satisfactory standard”, “reasonable steps”, used at a decision point with no stated considerations, no examples and no escalation path, is not preserved discretion. It is an unfinished sentence, and it is the reason two officers reach different outcomes on the same facts.

Presentation carries much of the load. Rules with several conditions and exceptions are far easier to apply from a table than from a paragraph. The Australian Government Style Manual treats sequential structure as the correct choice for step by step instruction and hierarchical structure for reference lookup, and warns that

²⁰ R.N. Shiffman, J. Dixon, C. Brandt et al., ‘The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation’, *BMC Medical Informatics and Decision Making* 5 (2005): article 23.
<https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-5-23>

²¹ Y. Hatakeyama, K. Seto, R. Amin, T. Kitazawa, S. Fujita, K. Matsumoto and T. Hasegawa, ‘The structure of the quality of clinical practice guidelines with the items and overall assessment in AGREE II: a regression analysis’, *BMC Health Services Research* 19 (2019): article 788.
<https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-019-4532-0>

unconventional or inconsistent structures make people work harder to find and understand content.²² A procedure written as continuous narrative prose is a structural defect, not a stylistic preference.

What failure looks like. A storage requirement expressed as a single sentence containing “unless”, “except where” and “provided that”, governing four categories of firearm with two exceptions, none of which is enumerated. A procedure instructing an officer to escalate matters of concern to the appropriate delegate, where neither the threshold nor the delegate is identified anywhere in the document.

How to check.

For every requirement, fill four slots: who does what, under what condition, and by when. Report requirements with an empty slot.

Every enumerated set of conditions states the outcome when none is met.

Any sentence containing two or more conditional connectives, or any exception to an exception, is presented as a table, a decision structure, or numbered conditions.

Separate statutory terms of art from agency-invented evaluative language. Report the first as a map of where judgement lives in the scheme. Report the second, where it carries decision weight and has no stated considerations, as a defect.

Directive sentences whose main verb is abstract carry an accompanying concrete step, actor, artefact or timeframe.

Procedures use numbered sequential steps. Reference material uses a hierarchy. Neither uses continuous prose to carry a process.

5. Findability and structure

A reader arriving with a question reaches the governing clause quickly, and that clause makes sense on its own.

Nobody reads a policy document from front to back. Readers arrive with a question, enter at a point of need, extract an answer and leave. Every structural decision should be made for that reader rather than for a hypothetical one who begins at page one.

The evidence on this is unambiguous and slightly bleak. Bakos, Marotta-Wurgler and Trossen²³, tracking 48,154 monthly visitors across ninety retail software sites, found that only one or two shoppers per thousand accessed the licence agreement at all, and most of those who did read a small fraction of it. Nielsen²⁴, reporting testing across forty two organisations, found average task success on intranets at seventy four per cent, essentially unchanged from seventy five per cent a decade earlier and below the roughly eighty per

²² Australian Government Style Manual, *Types of Structure* (Canberra: Australian Public Service Commission). <https://www.stylemanual.gov.au/structuring-content/types-structure>

²³ Y. Bakos, F. Marotta-Wurgler and D.R. Trossen, 'Does anyone read the fine print? Consumer attention to standard-form contracts', *Journal of Legal Studies* 43, no. 1 (2014): 1-35. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1443256

²⁴ J. Nielsen, 'Intranet users stuck at low productivity', Nielsen Norman Group, 3 December 2012. <https://www.nngroup.com/articles/intranet-users-stuck-low-productivity/>

cent typical of public websites. A decade of investment in search and portal design moved nothing, because findability is a property of how content is written and organised rather than of the layer placed on top of it.

The counter-evidence is equally clear, and it comes from the rewriting studies collected by Kimble²⁵. When Federal Express restructured its operations manuals, the time taken to find information fell from five minutes to 3.6 minutes and, more importantly, the success rate rose from fifty three per cent to eighty per cent. Nearly half the searches that had previously failed began to succeed. When the Federal Communications Commission rewrote its marine radio rules, mean scores on a comprehension test rose from 10.66 out of twenty to 16.85, and the time taken to answer fell from 2.97 minutes to 1.62. When the United States Veterans Administration rewrote a standard letter, the proportion of recipients who failed to understand it fell from fifty six per cent to eleven per cent, and inbound calls on a related form letter fell from 1,128 per 750 letters sent to 192 per 710.

Those gains came from structure, ordering and heading design, not from vocabulary substitution. The Australian Government Style Manual sets out the rules concretely enough to check: exactly one top level heading, no more than three or four heading levels, no skipped levels, no orphan subheadings, no heading longer than seventy characters, and keywords in the first two or three words of every heading. It advises against question headings on the ground that they push the main idea to the end, and it treats deep hierarchies as a signal that a document is trying to do too much.²⁶

Two further properties matter and are less often discussed. The first is addressability. Every substantive provision needs a stable number that can be cited, linked to and referred to in training material, and cross references need to point at those numbers rather than at “the section above” or a page number that changes on reformatting. This is the principle behind the naming conventions used for legislation itself, where every provision carries a persistent identifier independent of how the document is rendered.²⁷ The second is self-containment. A reader who lands in the middle of a document has not read what precedes it, so a section that opens with “this requirement” or “the said officer” without saying which requirement or which officer has failed before it starts.

What failure looks like. A manual with a fourteen page table of contents in which the entries are called General, Other, Background and Further Information. A cross reference reading “see the relevant procedure” where no such procedure exists. A section whose obligations apply only to a class of licence named four pages earlier.

How to check.

Heading hierarchy: one top level heading, three or four levels maximum, no skipped levels, no orphan subheadings, nothing over seventy characters.

Heading informativeness: keywords in the first two or three words, no generic labels, headings that predict what sits beneath them.

²⁵ J. Kimble, ‘Writing for dollars, writing to please’, *Scribes Journal of Legal Writing* 6 (1996-97): 1-38. <https://www.editorsoftware.com/wp-content/uploads/2021/03/kimble-writing-for-dollars-plain-english.pdf>

²⁶ Australian Government Style Manual, *Headings* (Canberra: Australian Public Service Commission). <https://www.stylemanual.gov.au/structuring-content/headings>

²⁷ OASIS, *Akoma Ntoso Naming Convention Version 1.0* (2018). <https://unsceb-hlcm.github.io/part1/index-13.html>

Front-loading: each section states its rule or requirement before its background.

Addressability: every substantive provision numbered, every internal cross reference resolving to a numbered target.

Self-containment: split the document at heading boundaries and read each section cold. Flag unresolved pronouns and definite references whose antecedent sits outside the section, and backward references with no numbered target.

Structure type fitness: the organising structure matches the reader task the section serves.

6. Understandability

Sentences a reader can hold in working memory, one term for one concept, and no arithmetic required to determine an obligation.

This element is the one most often reduced to a number, and the number does not carry the meaning.

Mac et al.²⁸ gave 2,235 Australian adults the same health information written at four different reading grade levels in a four arm randomised trial. Mean knowledge scores were 9.0, 9.1, 8.9 and 9.1 out of fourteen. There were no statistically significant differences, and no differences in perceived ease, acceptability or trustworthiness either. Lowering the grade level, on its own, achieved nothing.

That result is easier to accept once you see what the formulas measure. Flesch-Kincaid, SMOG and their relatives count syllables, words and sentence lengths. As Schriver²⁹ observes in her account of the continuum from text-focused to reader-focused evaluation, a passage scores identically whether its sentences run forwards or backwards. Jarrett and Redish³⁰ set out the practical consequences at length. A formula cannot see whether a document answers the reader's question, whether the answer is findable, or whether the rule is stated before three paragraphs of history. Readability scores are useful as a triage signal for finding the worst material in a large library. They are not a verdict.

What does change comprehension has been identified precisely. Martínez, Mollica and Gibson³¹, analysing a corpus of roughly 225 million words, found that contract-style writing overuses four features: low frequency jargon, centre-embedded clauses, passive voice, and non-standard capitalisation used for emphasis. An accompanying experiment with 184 participants found that those features reduced recall and comprehension, with centre-embedding, the insertion of material between a subject and its verb, the worst offender. The follow up tested practising lawyers on the same material.³² They comprehended legalese worse

²⁸ O. Mac, J. Ayre, K. McCaffery, F. Boroumand, K. Bell and D.M. Muscat, 'The Readability Study: a randomised trial of health information written at different grade reading levels', *Journal of General Internal Medicine* 40, no. 8 (2025): 1820-1828. doi:10.1007/s11606-024-09200-z

²⁹ K.A. Schriver, 'Evaluating text quality: the continuum from text-focused to reader-focused methods', *IEEE Transactions on Professional Communication* 32, no. 4 (1989): 238-255. <https://www.karenschriverassociates.com/wp-content/uploads/2020/03/9-Schriver-Evaluating-Text-Quality-Text-to-Reader-focused-Methods.pdf>

³⁰ C. Jarrett and J. Redish, *Readability Formulas: Seven Reasons to Avoid Them and What to Do Instead* (Effortmark, 2019). <https://www.effortmark.co.uk/readability-formulas-seven-reasons-to-avoid-them-and-what-to-do-instead/>

³¹ E. Martínez, F. Mollica and E. Gibson, 'Poor writing, not specialized concepts, drives processing difficulty in legal language', *Cognition* 224 (2022): 105070. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4048460

³² E. Martínez, F. Mollica and E. Gibson, 'Even lawyers do not like legalese', *Proceedings of the National Academy of Sciences* 120, no. 23 (2023): e2302672120. doi:10.1073/pnas.2302672120

as well, rated simplified contracts as equally enforceable, and preferred them. The mechanism is a limit on working memory rather than the difficulty of the underlying concepts, which is why professional expertise does not confer immunity.

Plain language is not the same thing as informality, and the two should not be confused. Linos et al.³³, across six studies with a combined sample of 67,632 including randomised field experiments in live government programmes, found that formal government letters outperformed informal ones on real compliance behaviour, with relative improvements of twenty five per cent on business self-certification, forty five per cent on emergency medical transport enrolment and twenty eight per cent on tax credit engagement. More than eighty four per cent of expert forecasters had predicted the opposite. The formal condition in those experiments varied several features together, so the finding bears on register rather than on sentence structure, but the direction is clear enough: an instrument that carries legal obligation should sound like one. The work is to simplify its structure, not to make it chatty.

The scale of the plain language task is also routinely underestimated. The International Plain Language Federation, which co-developed ISO 24495-1, reports that only about seven per cent of the standard's roughly 485 discrete ideas concern word choice.³⁴ The other ninety three per cent addresses identifying readers and their purpose, selecting content, structuring the document, designing headings, and evaluating the result with real readers. The standard's four governing principles are framed as reader outcomes rather than text properties: readers get what they need, can easily find what they need, can easily understand what they find, and can easily use the information.³⁵

Terminology discipline sits inside this element and is worth separate attention. One concept should have one term, used identically across the document and across the documents around it. Synonym drift, where a licence holder becomes a permit holder becomes the registered person, forces the reader to work out whether three things are being described or one. Every specialised term should be defined once, at or before first substantive use, and every acronym expanded on first use in each document, because readers enter documents individually rather than reading the suite in order.³⁶

What failure looks like. A definition section defining eleven terms, six of which never appear again, while four terms used throughout the document are defined nowhere. An eligibility rule requiring the reader to calculate a period from an event that is itself defined by reference to another document.

How to check.

Average sentence length against a fifteen word target, with every sentence over twenty five words listed and a split proposed.³⁷ Exclude quoted statutory text, which cannot lawfully be edited.

³³ E. Linos, J. Lasky-Fink, C. Larkin, L. Moore and E. Kirkman, 'The formality effect', *Nature Human Behaviour* 8, no. 2 (2024): 300-310. doi:10.1038/s41562-023-01761-z

³⁴ International Plain Language Federation, *The ISO Plain Language Standard: Part 1 and Other Parts* (2025). <https://www.iplfederation.org/the-iso-plain-language-standard-part1-and-part2/>

³⁵ International Organization for Standardization, *ISO 24495-1:2023 Plain Language, Part 1: Governing Principles and Guidelines* (Geneva: ISO, 2023). <https://www.iso.org/standard/78907.html>

³⁶ Australian Government Style Manual n.d.a, n 16.

³⁷ Australian Government Style Manual, *Sentences* (Canberra: Australian Public Service Commission). <https://www.stylemanual.gov.au/writing-and-designing-content/clear-language-and-writing-style/sentences>. See also Government Digital Service, 'Sentence length: why 25 words is our limit', Inside GOV.UK blog, 4 August 2014. <https://insidegovuk.blog.gov.uk/2014/08/04/sentence-length-why-25-words-is-our-limit/>

Centre-embedded clauses, sentences with more than one level of subordination, and blocks of capitals used for emphasis.

Every acronym expanded on first use in each document. Every defined term defined once, before or at first use, and used consistently after.

Near-synonymous terms clustered across the document set, with clusters flagged where two or more terms denote the same entity or obligation.

Syntactic complexity and register formality scored independently, so that neither is used to fix the other.

Any obligation requiring the reader to combine two numbers, or expressed as a proportion of an undefined base.

7. Workability

The document describes work that can actually be performed, in the time and with the resources available.

This is the element that separates a document which is correct from a document which is used, and it is the one most consistently absent from written quality frameworks.

The evidence comes from process safety and from healthcare, where the consequences of a procedure that cannot be followed are visible enough to have been studied. Hale, Borys and Else³⁸, reviewing the literature on safety rules and procedures for the Institution of Occupational Safety and Health, collect the relevant surveys. In one study of four hundred chemical industry operators and managers, forty per cent said procedures were unworkable in practice, sixty two per cent said the job could not be done in the time available if procedures were followed to the letter, forty eight per cent found them too restrictive, and forty four per cent said there were simply too many. In a separate industrial survey, fifty seven per cent of workers were unaware that written procedures for their job existed at all. Dutch railway workers reported at ninety five per cent that keeping to the rules meant the work could never be finished on time, seventy nine per cent that there were too many rules, and seventy seven per cent that the rules were sometimes contradictory.

The consequences show up in incident data. A Dutch analysis of loss of containment incidents, also reported in that review, found that half were procedure related, and the breakdown is the interesting part: ten per cent involved no procedure or an unclear one, twelve per cent involved a procedure that was wrong, and twenty eight per cent involved the correct procedure not being followed. Roughly two fifths of procedure related incidents traced to a defect in the document rather than to indiscipline. The UK Health and Safety Executive reaches the same conclusion in its guidance on procedures, which treats an unusable procedure as a management failure rather than a compliance failure.³⁹

Volume is part of the same problem. Hale, Borys and Else⁴⁰ report that the sheer size of rule books in complex industries acts as a barrier to their use, and that increases in the number of procedures have been

³⁸ A. Hale, D. Borys and D. Else, *Management of Safety Rules and Procedures: A Review of the Literature* (Leicester: Institution of Occupational Safety and Health, 2012). <https://iosh.com/media/rbohgc2k/iosh-research-safety-rules-and-procedures-literature-review-full-report.pdf>

³⁹ Health and Safety Executive, *Human Factors: Procedures* (Bootle: HSE). <https://www.hse.gov.uk/humanfactors/topics/procedures.htm>

⁴⁰ Hale, Borys and Else, n 38.

associated with worse rather than better safety performance. Adding a rule is the cheapest visible response to an incident and the one most likely to be taken, which is why document sets grow in one direction only. A suite whose size is never measured and never reduced is a suite getting steadily less usable. Degani and Wiener⁴¹, writing on flight deck procedure design, make the related point that procedures which do not match how the work is actually done invite improvisation, and that the feedback path from operator to procedure author is the control that keeps the two aligned.

Publication is not implementation. Runciman et al.⁴² found that Australian patients received care in line with guideline recommendations at fifty seven per cent of 35,573 eligible encounters, with a confidence interval of fifty four to sixty per cent, despite the guidelines being expert-endorsed and widely disseminated. Compliance ranged from thirteen per cent for alcohol dependence to ninety per cent for coronary artery disease. The variation between conditions, using the same guidelines and the same clinicians, points at the documents and the systems around them rather than at willingness. Cabana et al.⁴³ supply the framework that separates the reasons: a clinician may be unaware of a recommendation, unfamiliar with it, in disagreement with it, lacking confidence in performing it, doubtful that it will work, or blocked by circumstances outside their control. Only the first two are solved by publishing.

What failure looks like. A procedure requiring nine mandatory checks within a two hour window that also contains four hours of prescribed waiting. A step requiring approval from a committee dissolved in the last restructure. A form referenced by name in three places and available in none of them.

How to check.

Flag steps requiring a form, system, role, delegation or approval that no longer exists.

Flag sequential dependencies that are circular, and steps requiring information the actor cannot hold at that point in the process.

Compare the number of mandatory steps against the timeframe the document itself specifies.

Measure suite volume: total documents, total pages, and total distinct obligations, tracked over time.

Require every proposed new document to name what it replaces, and report net obligations added each review cycle.

Ask a sample of the people who do the work when they last performed the task exactly as written, on a voluntary and de-identified basis, framed as a test of the document. Where practice and text diverge, treat it as a finding against the document until the document has been ruled out.

⁴¹ A. Degani and E.L. Wiener, *On the Design of Flight-Deck Procedures*, NASA Contractor Report 177642 (Moffett Field, CA: NASA Ames Research Center, 1994). <https://skybrary.aero/sites/default/files/bookshelf/1567.pdf>

⁴² W.B. Runciman, T.D. Hunt, N.A. Hannaford et al., 'CareTrack: assessing the appropriateness of health care delivery in Australia', *Medical Journal of Australia* 197, no. 2 (2012): 100-105. <https://www.mja.com.au/journal/2012/197/2/caretrack-assessing-appropriateness-health-care-delivery-australia>

⁴³ M.D. Cabana, C.S. Rand, N.R. Powe et al., 'Why don't physicians follow clinical practice guidelines? A framework for improvement', *JAMA* 282, no. 15 (1999): 1458-1465.

8. Structured and accessible form

The authoritative version is structured content rather than an image of a document.

The way a document is encoded determines who and what can read it. Headings applied as bold body text are not headings. A table built with tab characters is not a table. A scanned page is not text at all.

The accessibility case is settled and, for Australian agencies, carries legal weight. WCAG 2.2 success criterion 1.3.1 requires that information, structure and relationships conveyed through presentation be programmatically determinable, including the relationship of each table cell to its row and column headers.⁴⁴ The obligation reaches government documents through the Disability Discrimination Act 1992 and the Commonwealth's digital service requirements.⁴⁵ A licensing eligibility table that a screen reader cannot parse is not a formatting preference. It is a barrier to a class of applicants.

The same structural work has a second payoff. Content that a screen reader can navigate is content that search, indexing and automated retrieval can navigate. Studies of document question answering find structure-aware chunking outperforming fixed-size chunking by a wide margin, and adding structured metadata to content lifts retrieval precision materially.⁴⁶ The Government Digital Service, whose estate has accumulated on the order of two hundred thousand PDFs, has published its own account of why the format is the wrong default: PDFs are harder to find and search, do not adapt to the device, behave unpredictably with assistive technology, and are updated far less often than the pages around them.⁴⁷ Its later work on serving that estate through an assistant found retrieval failing where pages were too long or poorly structured, which is the same defect surfacing in a different channel.⁴⁸

Format policy follows from this. The Australian Government Content Guide treats HTML as the default and requires a demonstrated user need before publishing as PDF.⁴⁹ Where a PDF must exist, it should be tagged, its text layer intact, its language set and its tables real. Automated extraction of tables from PDFs remains unreliable enough that decision content held only in that form should be treated as at risk.⁵⁰

What failure looks like. A four hundred and eighty nine page manual delivered as a scan, in which the storage requirements table is an image, so that no search will ever return it and no screen reader will ever read it. A fee schedule maintained in two formats that have drifted apart, with no indication of which is authoritative.

⁴⁴ W3C, *Understanding Success Criterion 1.3.1: Info and Relationships (WCAG 2.2)*. <https://www.w3.org/WAI/WCAG22/Understanding/info-and-relationships.html>

⁴⁵ W3C Web Accessibility Initiative, *Web Accessibility Laws and Policies: Australia* (2025), <https://www.w3.org/WAI/policies/australia>; and Digital Transformation Agency, *Digital Service Standard, Criterion 3: Leave No One Behind*, <https://www.digital.gov.au/policy/digital-experience/digital-service-standard/criterion-3>

⁴⁶ *Vision-Guided Chunking Is All You Need: Enhancing RAG with Multimodal Document Understanding*, arXiv preprint 2506.16035 (2025), <https://arxiv.org/html/2506.16035>; and *A Systematic Framework for Enterprise Knowledge Retrieval: Leveraging LLM-Generated Metadata to Enhance RAG Systems*, arXiv preprint 2512.05411 (2026), <https://arxiv.org/html/2512.05411v1>

⁴⁷ Government Digital Service, 'Why GOV.UK content should be published in HTML and not PDF', GDS blog, 16 July 2018. <https://gds.blog.gov.uk/2018/07/16/why-gov-uk-content-should-be-published-in-html-and-not-pdf/>

⁴⁸ Government Digital Service, 'Developing GOV.UK Chat: our data science and AI engineering journey', Inside GOV.UK blog, 15 May 2026. <https://insidegovuk.blog.gov.uk/2026/05/15/developing-gov-uk-chat-our-data-science-and-ai-engineering-journey/>

⁴⁹ Australian Government, *Australian Government Content Guide: Types of Content* (Canberra: Digital Transformation Agency). <https://guides.service.gov.au/content-guide/types-of-content/>

⁵⁰ *Benchmarking PDF Parsers on Table Extraction with LLM-based Semantic Evaluation*, arXiv preprint 2603.18652 (2026). <https://arxiv.org/html/2603.18652v1>

How to check.

Headings use heading styles. Lists use list features. Tables are table objects.

Table header rows and columns are identified, no cells are merged or blank, headers repeat across page breaks, and no table carrying decision content exists as an image.

No scanned or image-only documents. PDFs are tagged, the text layer is present, and document language is set.

Link text is meaningful out of context. No meaning is carried by colour alone. Contrast meets the required ratio.

Inventory the library by format and report the proportion available only as PDF, together with whether a user need was recorded for each.

9. Instrument fitness

The document is one declared kind of thing, it behaves like that kind of thing, and it should exist at all.

Readers cannot tell whether a document binds them unless it says what it is. A policy states a position and the mandatory minimum. A procedure states the sequence of steps. A standard states a measurable requirement. A guideline advises and may be departed from with reason. Agencies that publish this taxonomy, and attach a compliance consequence to each tier, give readers a way of calibrating everything they open. NSW Health treats policy directives as mandatory and guidelines as departures permitted only where sound reasons exist,⁵¹ and comparable frameworks operate in Queensland and in the university sector.⁵²

The taxonomy should be small. Published frameworks range from three types to fourteen, and there is no evidence that a richer taxonomy produces better governance. A set that staff cannot recite is worse than a coarse one they can. What matters is that every document declares its type, that the type comes from the published set, and that the content matches the claim. The information typing principle in technical communication makes the same argument from a different direction: mixing content types in a single unit loses the reader's focus and prevents reuse.⁵³

Content mismatch is common and diagnostic. A document labelled policy that consists mostly of numbered steps is a procedure with the wrong cover. A document labelled guideline that imposes mandatory requirements is binding readers who have been told they have discretion. A single document that combines organisational position, process detail, historical background and transcribed legislation is doing four jobs and doing none of them where a reader would look for it.

⁵¹ NSW Health, *Policy Development and Publication*, Policy Directive PD2022_047 (Sydney: NSW Ministry of Health, 2022). https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2022_047.pdf

⁵² Queensland Health, *Department of Health Policy Framework*, <https://www.health.qld.gov.au/system-governance/policies-standards/framework>; and Western Sydney University, *Policy Framework*, <https://policies.westernsydney.edu.au/document/view.current.php?id=121>

⁵³ OASIS DITA Technical Committee, *DITA 1.3 Architectural Specification: Information Typing* (2018). <https://docs.oasis-open.org/dita/dita/v1.3/os/part1-base/archSpec/base/information-typing.html>

Instrument fitness also raises the question a policy suite rarely asks of itself, which is whether a written procedure is the right control at all. Some tasks are better handled by training, by a form that cannot be completed incorrectly, or by a system that will not permit the wrong action. Procedures written for tasks that did not need them are a substantial part of what makes a document set unusable.

What failure looks like. A four hundred and eighty nine page document titled as a manual, containing organisational position, operational procedure, transcribed legislation and historical commentary, none of it separable and all of it carrying the same apparent authority.

How to check.

The instrument type is stated on the face of the document and comes from the agency's published set.

The document states whether compliance is mandatory and what departing from it requires.

Content type matches the declared type. Count obligation sentences, numbered imperative steps and advisory statements, and flag documents whose mix contradicts their label.

Every guideline references the policy or standard it supports. Orphaned advisory documents belong to nothing.

One document, one job. Flag documents combining principle, process, background and transcribed legislation.

10. Implementability and assurance

Someone can act on it, someone can check that it was done, and someone has watched a real reader try.

This is empirically the weakest property of guidance documents, and knowing that is useful. Dijkers et al.⁵⁴, reviewing 544 rehabilitation clinical practice guidelines appraised with AGREE II, found that the Applicability domain, which covers practical tools, resource implications and monitoring criteria, averaged thirty four out of one hundred against seventy two for Scope and Purpose. Only thirty six per cent of the guidelines were recommended for use without modification, forty four per cent with modification, and twenty per cent not at all. Documents are good at saying what they are for and poor at saying how anyone would know they had been followed. Hatakeyama et al.⁵⁵ reach a compatible conclusion from the other side: the provision of advice and tools for implementation is one of the four AGREE II items that carry most of the weight in an overall quality judgement.

Three things close that gap.

The first is tools. A document that requires an action should supply or link what the reader needs to perform it: the form, the checklist, the template, the decision aid, the worked example.⁵⁶ A requirement whose

⁵⁴ M.P. Dijkers, S.L. Murphy, J. Krellman et al., 'Quality of rehabilitation clinical practice guidelines: an overview study of AGREE II appraisals', *Archives of Physical Medicine and Rehabilitation* 101, no. 9 (2020): 1643-1655. [https://www.archives-pmr.org/article/S0003-9993\(20\)30266-5/fulltext](https://www.archives-pmr.org/article/S0003-9993(20)30266-5/fulltext)

⁵⁵ Hatakeyama et al., n 21.

⁵⁶ M.C. Brouwers, M.E. Kho, G.P. Browman et al., 'Development of the AGREE II, part 1: performance, usefulness and areas for improvement', *Canadian Medical Association Journal* 182, no. 10 (2010): 1045-1052. <https://www.cmaj.ca/content/182/10/1045>

associated form is described but not provided has offloaded the work of finding it onto every reader, every time.

The second is monitoring. The document should state an observable indicator of compliance, a role responsible for checking it, and a frequency or trigger. Most policy documents contain no monitoring clause at all, which means non-compliance can persist indefinitely without anyone being obliged to notice.

The third is reader testing, and it is the element that no amount of analysis substitutes for. ISO 24495-1 makes evaluation with readers a governing principle,⁵⁷ and the practice is well established in the one field that regulates it: medicines regulators require patient information leaflets to be user tested before approval, using a defined protocol and a stated pass mark.⁵⁸ The Therapeutic Goods Administration applied the same method to Australian consumer medicine information and found that the changes which moved documents to the pass benchmark were structural rather than lexical.⁵⁹ Almost no organisational policy document has ever been tested on the people who must use it. The minimum viable version costs an afternoon: write down the ten questions the document exists to answer, give three people who must use it one section each, and ask them to say in their own words what they must do. Any divergence is a finding. Record the test date in the register alongside the review date so that testing carries the same governance visibility as review.

A feedback channel is the fourth thing, and the single best diagnostic of whether a policy suite is alive. Does the document tell the reader how to raise a problem with it, and can the agency point to a change made in the last twelve months because someone did?

What failure looks like. A procedure that has governed a licensing function for twenty six years and has never been read aloud to a licensing officer to see whether they can follow it.

How to check.

The document supplies or links the forms, checklists, templates and worked examples its requirements depend on.

The document states an observable compliance indicator, a responsible role, and a checking frequency or trigger.

The document names a route for readers to report a defect, and the agency can evidence at least one change that came from that route.

For each high-use document, a record exists of a non-author attempting to answer its core questions using it, with a stated pass mark and evidence of revision where it failed.

⁵⁷ ISO 2023, n 35.

⁵⁸ Medicines and Healthcare products Regulatory Agency, *User Testing Guidance on Patient Information Leaflets* (London: MHRA, 2019). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/808259/User_Testing_Guidance_on_Patient_Information_Leaflets_for_Parallel_Importers.pdf

⁵⁹ Therapeutic Goods Administration, *Usability Evaluation of Consumer Medicine Information Documents* (Canberra: TGA, 2019). <https://www.tga.gov.au/sites/default/files/improved-consumer-medicine-information-template-report.pdf>

11. Assumed knowledge

The corpus is usable by the reader it actually has, not the reader who has been travelling with it since the beginning.

Every set of guidance encodes an implicit model of its reader, and the model is rarely written down. When a framework develops over a decade, with foundational concepts held in one document and applied guidance spread across twenty more, that model quietly becomes *someone who has followed the whole thing as it was built*. That reader exists. They are the people who wrote it, plus a small number of specialists who work with the framework every week.

The reader who actually arrives is different. They did a one day induction three years ago. They meet the situation the framework governs a handful of times a year. They open a document at the moment they need it, under time pressure, often with a client in front of them. For that reader, a corpus organised as a body of knowledge to be acquired is close to unusable, and not because any individual document is badly written. It is because acting on chapter fourteen requires knowledge that was distributed across a separate foundation document, two earlier chapters, and a definitional shift that happened between one release and the next.

This is the reason the element sits here rather than earlier. Everything above assesses a document against a reader arriving at it. This one assesses the whole corpus against a reader's working life, and it is the point at which document quality stops being a property of documents.

The pattern has several recognisable symptoms.

Foundational material is held in its own document, which every applied document assumes and none restates. The foundation document is read once, at induction, and never again.

Terminology evolves across releases, so the same term carries slightly different content depending on the vintage of the chapter a reader happens to open. Nothing in either chapter says so.

Parallel document sets are maintained for different populations, with much of the guidance restated and the audience terms swapped. A practitioner who works across both must hold two nearly identical mental copies and know at every moment which one governs. Two chapters may share more than half their content and still differ in ways that matter.

Documents are organised by professional role or by framework structure rather than by the situation a reader is in, so someone with a problem must first work out which role they are occupying before they can find the guidance for it.

Nothing states the entry cost. No document says what a reader is assumed to know before opening it, or where to acquire it.

The supporting evidence is indirect but consistent. Cabana et al.⁶⁰ separate the reasons a practitioner does not follow guidance, and the first two, being unaware of a recommendation and being unfamiliar with it, are knowledge distribution failures rather than motivation failures. Hale, Borys and Else⁶¹ report an industrial

⁶⁰ Cabana et al., n 43.

⁶¹ Hale, Borys and Else, n 38.

survey in which fifty seven per cent of workers did not know that written procedures for their job existed at all. Carroll⁶², in the work that established minimalist instruction, found that people do not read documentation in order to learn a system; they read to act, they start in the middle, and they abandon material that requires them to build up context first. Bakos, Marotta-Wurgler and Trossen⁶³ show what happens to material that a reader is not compelled to open.

The fix is not to copy the foundation into every document, which recreates the duplication problem described in the next section. It is layering. Each applied document carries inline the minimum knowledge required to make the decision it governs, with a pointer to the fuller treatment for the reader who wants it. The test of the layer is whether an occasional user can act correctly without leaving the document, not whether the document is complete.

What failure looks like. A framework of twenty applied chapters plus a separate foundation guide, published across two release years, in which two parallel sets restate the same collaboration guidance for different populations, and a practitioner who encounters the situation twice a year must know which of four documents governs before they can start.

How to check.

State the intended reader and how often they encounter this material. Where the answer is less than monthly, the corpus must teach at the point of use rather than assume prior acquisition.

For each applied document, list what a reader must already know to act on it, and where that knowledge currently lives. The number of other documents they must have read is the entry cost, and it should be reported.

Every document states what it assumes and where to get it. A prerequisite that is not declared is a prerequisite the reader discovers by failing.

Track defined terms across releases. Where a term's content has shifted between vintages, both documents say so.

Where parallel sets exist for different audiences, identify which is authoritative for shared content, and how a reader working across both is meant to tell them apart.

Provide at least one entry path organised by situation rather than by role or framework structure.

Run the occasional user test. Take someone who completed the training two years ago, give them a real case, and observe. Count the documents they open, the points at which they stop, and whether they reach the right answer. This is the same method as the reader testing in element ten, aimed at the corpus rather than the page.

⁶² J.M. Carroll, *The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill* (Cambridge, MA: MIT Press, 1990).

⁶³ Bakos, Marotta-Wurgler and Trossen, n 23.

The document is not the unit

Every element above assesses one document. A significant share of the damage lives between documents, where no reader and no reviewer will find it, because finding it requires holding the whole library in mind at once.

Single source of truth. Each requirement is stated authoritatively in one place, and other documents cross reference rather than restate it. Where the same requirement appears in six documents, five of them will eventually be wrong, and the amendment that makes them wrong will update only the one whose owner happened to hear about it. Clusters of near-identical requirement statements where the copies already differ are confirmed defects. Clusters where they still agree are defects waiting for the next amendment.

Non-contradiction. No two documents impose different obligations on the same actor under the same conditions. The Audit Office of New South Wales has made this an explicit sector recommendation, requiring agencies to ensure their policies do not conflict with one another.⁶⁴ The failure is rarely dramatic. It is a fee stated as two different amounts, a timeframe stated as twenty eight days in one place and one month in another, a category described as requiring approval here and notification there.

The register and the shadow library. A central register records every document actually in circulation, with owner, approval date, version, review date and status.⁶⁵ The measure that matters is the gap between what the register holds and what staff are actually using from shared drives, intranets and inboxes. An agency cannot govern what it has not counted.

Retirement. Superseded documents are formally rescinded with a recorded status, and no current document cites, incorporates or cross references a rescinded one.⁶⁶ A live library should contain only live instruments.

Coherence. Every procedure implements a stated policy, and every policy has an implementing procedure. Procedures that trace to nothing, and policies that nobody implements, are both signs of a set that accreted incident by incident rather than being designed.

Consistency of outcome. The suite is specific enough that officers applying it to like cases reach like outcomes. Where outcome data exists, unexplained variation between officers, regions or periods is the observable symptom of everything above. Firearms licensing in England and Wales supplied a striking illustration, with inspected forces administering the same statute recording average certificate processing times ranging from five days to one hundred and sixty five, under guidance that existed and was not consistently followed.⁶⁷

⁶⁴ Audit Office of New South Wales, *Internal Controls and Governance 2024* (Sydney: Audit Office of NSW, 2024). <https://www.audit.nsw.gov.au/our-work/reports/internal-controls-and-governance-2024>

⁶⁵ Audit Office of NSW 2024, n 64.

⁶⁶ NSW Health 2022, n 51.

⁶⁷ HM Inspectorate of Constabulary, *Targeting the Risk: An Inspection of the Efficiency and Effectiveness of Firearms Licensing in Police Forces in England and Wales* (London: HMIC, 2015). <https://hmicfrs.justiceinspectorates.gov.uk/news/news-feed/firearms-licensing-missing-the-mark>

Applying this without stopping everything else

A library assessed against eleven elements will produce findings in the thousands, and a report of that size does not get read. Sequence matters more than completeness.

The evidence on appraisal cost is instructive. A full AGREE II appraisal takes forty five to sixty minutes per document and requires between two and five trained reviewers to reach acceptable reliability,⁶⁸ which makes it a project rather than a routine for any library of consequence. Rapid instruments exist for exactly this reason and take three to five minutes per document.⁶⁹ And as noted above, four AGREE II items alone explained 72.1 per cent of the variance in overall quality judgements across 206 guidelines.⁷⁰ Depth has sharply diminishing returns. A short assessment that actually runs across the whole library is worth more than a thorough one that runs across nothing.

A workable order:

Count and date everything. Register coverage and overdue exposure. Two numbers, obtainable in a fortnight, and defensible on their own. Audit offices repeatedly find that the missing control is not better documents but a regular report to senior management on which documents are overdue.⁷¹

Fix the decisions that carry legal consequence. Obligation-level citation for eligibility, refusal, revocation, seizure, fees and statutory timeframes. Not the whole corpus, and not the parts where being wrong costs nothing.

Resolve duplicates and contradictions. Each resolution permanently reduces future maintenance, because there is one fewer copy to keep in step.

Retire what is dead. A free reduction in the size of the suite, and it improves every other measure.

Then structure and language, beginning with the documents that carry the most decisions rather than the ones that read worst.

Two habits make the difference between a quality programme and a report. Cap the number of findings per document and rank them, so that the first page is the twenty things worth doing rather than the first twenty things the assessment happened to find. And exclude quoted legislative text from every language check, because it cannot lawfully be edited and it will otherwise dominate every list.

Conclusion

The elements in this paper are not a style guide. They are a description of what has to be true for a written instrument to do the job an agency wrote it for: to be current, lawful, clear about what it requires and of

⁶⁸ Brouwers et al. 2010, n 56.

⁶⁹ International Centre for Allied Health Evidence, *iCAHE Guideline Quality Checklist* (Adelaide: University of South Australia, 2011). https://implementationcentral.com/docs/iCAHE_Guideline_Quality_Checklist.pdf

⁷⁰ Y. Hatakeyama, K. Seto, R. Amin, T. Kitazawa, S. Fujita, K. Matsumoto and T. Hasegawa, 'The structure of the quality of clinical practice guidelines with the items and overall assessment in AGREE II: a regression analysis', *BMC Health Services Research* 19 (2019): article 788. <https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-019-4532-0>

⁷¹ Audit Office of NSW 2023, n 2.

whom, answerable to a reader in a hurry, readable by the people and the systems that must use it, honest about what kind of document it is, capable of being followed by someone doing the work, and usable by a reader who meets it occasionally rather than daily.

Most policy libraries fail on the first two, which are also the cheapest to fix. Very few agencies have ever tested a document on a reader, which is the one thing that would tell them whether any of the rest of it worked. Fewer still have asked what their corpus assumes its reader already knows. Between those facts sits most of the available improvement.

The measure of a good policy document is not that it is well written. It is that a person who needs an answer gets the right one, and a person who has to act knows what to do.