

**A SYSTEM THAT
CANNOT SEE ITSELF**

EVIDENCE • MEASUREMENT • REFORM

RESEARCH APPRAISAL • INTERNAL EVIDENCE ONLY

Does the evidence carry the recommendations?

A methodological appraisal of *Misogyny, Online Radicalisation, and Youth Extremism in Australia* (Meger, 2026)

Julian Talbot

11 September 2026

Working paper • v1

DFV Risk

ABSTRACT

This appraisal tests whether the conclusions and policy recommendations in *Misogyny, Online Radicalisation, and Youth Extremism in Australia* follow from the evidence the report itself publishes. It finds that the central empirical claim rests on a regression model that omits age and sex, that shares item content between its predictors and its outcomes, and whose published figure disagrees with its own summary table in four of nine models, including one sign reversal. A headline prevalence figure is drawn from a scale defined under one name and reported under another, after one of its two items was discarded. The qualitative chapters carry no stated method. None of this shows the underlying concern to be unfounded. It shows that two recommendations which would alter Australia's operating definition of violent extremism are not yet supported by a published analysis a reader can check.

9

regression models published, pp.43–44

4

disagree with their own summary table

0

controls for age or sex in the model



dfvrisk.com.au

Three books on domestic and family violence risk — free to download. Scan the code or visit the site.

1. Scope of this appraisal

What is being assessed, and what is not

This appraisal draws only on the report as published (June 2026 web edition, 51 pages). It performs no re-analysis, because the underlying dataset is not available. Every claim below is checkable against a page of the report.

Three things are explicitly outside its scope. It makes no claim about the author's intent; motive is not recoverable from a document and is not needed for the argument. It does not dispute that misogyny and youth radicalisation are real problems in Australia, for which independent evidence is substantial, including sources the report itself cites. And it is not a referee's report on an academic paper, which would be a different and gentler exercise: this report addresses five recommendations to named security, education and regulatory agencies, and the standard applied here is whether a practitioner could act on it.

That last point is why the evidentiary standard matters more than usual here, not less. A weak case for a true proposition is worse than no case at all, because it hands the proposition's opponents a legitimate target.

2. What the report gets right

Five contributions that should survive

- **It fills a real gap.** Attitudinal data from 1,122 Australians aged 13–17, with a matched adult comparison sample of 2,307, collected with ethics approval and guardian consent. Adolescent data on these questions is rare and genuinely difficult to obtain, and the youth–adult comparison is the report's most original design feature.
- **It publishes its item wording.** Pages 14–17 give the actual survey items. This is why an appraisal of this kind is possible at all; a great many comparable reports withhold it, and the decision to publish deserves credit even though it is what makes the problems below visible.
- **It confronts the ethnic attribution hypothesis directly.** Rather than avoiding the most common public deflection, the report tests it and reports that the associations persist under controls for ethnicity, religion and religiosity (pp.24–25). Answering the objection on its own terms is more useful than declining to engage with it.
- **Its most defensible finding is a null.** General social media use shows near-zero association with extremist attitudes ($p = .04-.10$, pp.30–31) against moderate associations for specific content types. Nulls are robust to most of the problems discussed below, and this one argues against blunt screen-time policy in favour of content and recommender design.
- **The teacher data contains the report's single best observation.** The differential-authority pattern, in which male students comply with instructions from male staff and refuse identical instructions from female staff (pp.38, 41), is a behavioural report rather than an attitude score, it is convergent with prior Australian research, and it is the least vulnerable material in the document.

3. Verified internal discrepancies

Six items, each checkable against the published pages

D1 pp.17, 19, 22, 43–44

A headline number comes from a scale that is never defined under that name

Page 17 defines six extremist-attitude scales: Religious, White Supremacist, Economic Reactionary, Left-Wing, Incel and Anti-Feminist. “Right-Wing VE” is not among them, yet it appears in every prevalence and correlation table in the report. The regression output on p.43 labels the same coefficients “Economic Reactionary VE”, and the narrative on p.22 confirms the identity by quoting “economic reactionary ($\rho = .53$)” for the value tabulated as Right-wing VE (.53).

The executive summary states that over 26% of teenage boys agree “with right-wing extremist attitudes” (p.3). That figure, 26.3% on p.19, is the Economic Reactionary scale renamed. Its two items concern the right to defend private property with force and violent resistance to a tax system that supports “freeloaders”. Those propositions may be worth measuring, but they are not what a reader, a journalist or a minister understands by right-wing violent extremism.

D2 pp.18–19

That scale is reduced to a single item, and the flag pointing at it leads nowhere

Page 18 records that the Economic Reactionary scale failed to cohere, and that “because so many more individual respondents agreed with the first question in this scale than the second, this item was excluded”. The scale has two items. Excluding one leaves one. The p.19 table flags the row “Right-Wing VE*” with an asterisk that has no corresponding footnote anywhere on the page.

D3 pp.43–44

The central regression figure and the table summarising it report different results

The report’s conclusion rests on nine multiple regressions, shown as a coefficient matrix on p.43 and summarised in a table on p.44. Five of the nine models match exactly. Four do not.

Model and coefficient	Figure, p.43	Table, p.44
Religious VE — Misogyny β	−0.053***	+0.049***
Religious VE — VAW β	0.724***	0.738***
Religious VE — Gender equality β	0.026*	0.109***
Radical Violence Sympathy — VAW β	0.593***	0.694***
Radical Violence Sympathy — Gender equality β	−0.016 (n.s.)	0.114***
Radical Intentions — VAW β	0.549***	0.580***
Radical Intentions — Gender equality β	0.003 (n.s.)	0.093***
Radical Intentions — strongest online predictor	Porn 0.093***	Dating 0.042***
Violent Intentions — VAW β	0.440***	0.617***
Violent Intentions — Gender equality β	−0.003 (n.s.)	0.100***

Online dating is reported as the strongest online predictor for Radical Intentions on p.44 at $\beta = .042^{***}$; on p.43 the same cell reads 0.011 and is not significant.

The sign reversal is the most consequential. On the report's own figure, misogyny has a small negative independent association with religious violent extremism once attitudes to violence against women are held constant. The facing page reports it as positive and significant, and the narrative describes misogyny as a crosscutting predictor across every ideological formation. Three further coefficients that are not significant in the figure appear at $p < .001$ in the table. A reader has no way to determine which set of numbers is the analysis.

D4 p.43

Age and sex are not in the model

The figure caption states that the models control for “politico-religiosity and online behaviours”. Age and sex are not among the controls. The report's own descriptive chapters establish that age and sex are the dominant sources of variation in both the predictors and the outcomes: adolescent boys score highest on very nearly every measure in the document. A model that omits them cannot distinguish the claim that attitudes to violence against women predict extremist attitudes from the observation that young males score above threshold on both sets of items.

D5 p.44

A coefficient is out by a factor of ten

The summary table gives pornography $\beta = 0.38^{**}$ for Left-Wing VE. The figure shows 0.038^{**} . On the table as printed, this is the second-largest online coefficient in the report.

D6 pp.7, 13, 17–21, 22–23, 43–44

Smaller inconsistencies, which matter cumulatively rather than individually

The adult sample is 2,307 on p.7 and 2,305 on p.13. One scale appears as “White Supremacist VE” (pp.17, 19, 44), “White ethnat. VE” (pp.20–21) and “White Ethnonationalist VE” (p.43). The strongest misogyny correlation is reported as $\rho = .62$ on p.22 and $\rho = .61$ on p.23, two paragraphs apart. Individually these are copy-editing faults. Together with D1 to D5 they indicate that the tables were not reconciled against the analysis before publication.

4. Design limitations

Issues of method rather than of transcription

4.1 The predictors and the outcomes measure overlapping content

The report's strongest reported relationship is between attitudes permissive of violence against women and anti-feminist violent extremism. The first item of the anti-feminist scale is “Feminism is damaging to our society and should be resisted with violence if necessary” (p.17), which is itself an endorsement of gender-based violence. Predictor and outcome share content by construction. Standardised betas of .52 to .75 sustained across nine outcomes, including ideologically opposed ones, is the signature of one latent construct measured repeatedly rather than of nine distinct predictions.

The report states that “exploratory factor analysis found high alpha scores for each of these scales” (p.18). Factor analysis does not produce alpha; they are different procedures answering different questions. No reliability coefficient, factor loading, factor count or fit statistic appears anywhere in the report. Four of the nine scales have two items, on

which alpha is not meaningfully interpretable. The question that actually matters here, whether the gender-attitude items and the extremism items load on distinguishable factors, requires a confirmatory analysis that is not reported.

4.2 Collinearity is not diagnosed

The four gender-attitude scales correlate with each other and with the outcome set at $\rho \geq .6$ (p.22). When strongly collinear predictors enter one model, one absorbs the shared variance and the others collapse toward zero. That is precisely the published pattern: VAW betas of .44 to .75, everything else between .00 and .23. The report reads the collapse substantively, concluding that the link between online behaviour and extremism “is substantially explained by misogynistic attitudes and VAW attitudes that online behaviours cultivate” (p.42). A variance inflation diagnostic is the minimum required before that reading; none is reported.

4.3 Causal language on a single cross-section

The data are one wave, October to November 2025. The report describes misogyny as a “training ground” (p.10), a “gateway” and “entry point” (p.23) and a “pipeline” (p.41); says online environments “cultivate” attitudes (p.32); and states as a key finding that “online environments amplify but do not create extremist orientation” (p.3). None of these are recoverable from a single cross-section. The mediation claim in particular, that online behaviour produces gender attitudes which produce extremism, is observationally identical to the reverse ordering and to a common third cause.

4.4 The prevalence figures are not population estimates

Agreement is defined as 5 or above on a seven-point scale (p.19). On a standard agree–disagree scale, point 5 is slight agreement, so the headline percentages count mild assent as endorsement. No confidence interval appears anywhere in the report. The adult sample is weighted to ABS benchmarks; no weighting is stated for the youth sample, which was deliberately oversampled and recruited through paid panel guardians. Data cleaning is described as removal of bots, duplicates, speeders and straight-liners plus one attention check (p.8), with no screening for mischievous responding, which is the standard concern in adolescent survey research. The pattern most characteristic of that artefact, boys aged 13–17 scoring above every adult group on all nine extremism scales simultaneously, including mutually exclusive ideologies, is present throughout. The executive summary nonetheless presents “roughly 1/3 of teenaged boys” as a national figure.

4.5 A face-validity failure is carried into the headline

Twelve per cent of boys aged 13–17 report using online dating at least once a day, and roughly a third report some use (p.29). Dating platforms are restricted to adults. Online dating is then reported as the most consistent online predictor of extremist attitudes across the regression set and named in the executive summary as one of three environments that are “not the ones we typically assume” (p.3). A variable a large share of the youth sample cannot plausibly be answering accurately should not carry a headline claim without a stated treatment.

4.6 The manfluencer measure does not isolate the manosphere

Boys were asked how regularly they watch “influencers whose content focuses on masculinity, wealth, fitness, discipline, self-improvement and relationships” (p.29). Eighty-eight per cent said they do. That wording captures mainstream fitness and self-improvement content alongside anything else. Recommendation 5, which would restrict algorithmic recommendation of “misogynistic, manosphere, or violent extremist” content to under-18s (p.50), rests substantially on this variable. Girls were asked a non-parallel question and only girls were asked it, so the two cannot be compared or jointly controlled.

4.7 The qualitative chapters have no stated method

Chapter 4 reports roughly 3,000 comments with sentiment splits given as approximate ranges, “~55–60%” and “~85–90%” (p.34), and cross-platform tables scored “High”, “Moderate” and “Very high” with no operational definitions, no counts and no inter-rater reliability. The teacher survey of about 500 respondents reports no sampling frame, no recruitment method, no response rate and no fieldwork date. On the social media minimum-age policy, the report states that the largest cluster of teachers saw no meaningful impact and then quotes only positive responses (p.40).

The report also states that the gendered pattern of hostile comments “reinforces the validity of the study’s findings as a live sample” (p.36), having conceded on the same page that comment sections are not representative of the general population. A self-selected hostile sample cannot validate a population estimate, and the report is right the second time.

4.8 Two pages of ethnic-group data are published without interpretation

Pages 20 and 21 carry ethnic-group breakdowns and regression coefficients under a single caption line and no interpretive text at all. Cell sizes include n = 34, n = 37 and n = 57, across which estimates range from 3% to 38%. The p.21 figure shows significant positive coefficients for Aboriginal and Torres Strait Islander respondents on all four gender-attitude measures, controlling for age, gender and religiosity. Chapter 4, discussing the teacher data, states that “the data does not support an interpretation of misogyny as a culturally specific phenomenon” (p.39). Whatever the correct reading of those small subsamples, the report does not reconcile the two anywhere.

5. Consequences for the four key findings

Tested against the material above

Key finding (p.3)	Status on the report’s own evidence
Misogyny is a robust, crosscutting predictor of violent extremism	Supported as a bivariate association. Not established as independent of age, sex, item overlap or response style.
VAW attitudes are the dominant, proximal predictor of extremism	The most exposed claim. Rests on the model affected by D3, D4, 4.1 and 4.2 together.
Online environments amplify but do not create extremist orientation	A causal and temporal claim a single cross-section cannot support in either direction. The underlying null for general social media use is nonetheless the report’s most defensible result.
Teachers are at the front line	Supported by the qualitative material, subject to the absence of stated method at 4.7.

6. Foreseeable effects if the recommendations are adopted as they stand

Risk assessment, not attribution of motive

6.1 A screening instrument with no discriminating power

Recommendation 2 proposes validated screening tools for gender-based attitudes for use in counter-extremism risk assessment by practitioners, police and social workers (p.47). The report’s own thresholds classify between 26% and 35% of adolescent boys as positive. The report also states that Australia has roughly 200 active participants in countering violent extremism programs, two-thirds of them aged 10 to 24 (p.5).

A screen that returns positive on a quarter of a demographic cohort, deployed to identify a national population of that order, has a positive predictive value indistinguishable from zero. Each genuine case would arrive accompanied by an extraordinary number of false positives, every one of them a child placed in contact with a security process. The report

reports no sensitivity, no specificity, no positive predictive value and no false-positive cost anywhere in its 51 pages. For anyone who would have to operate the instrument, this is the most consequential gap in the document.

6.2 A definitional shift from conduct to opinion

Recommendation 1 would amend the national definition of violent extremism to include “the use or endorsement of violence motivated by hostility toward women and/or feminist movements” (p.46). The word “endorsement” moves the threshold from what a person does to what a person agrees with on a survey. On the report’s own figures, that is 23% of Gen Z males. Recommendation 3 states that “the solution is not to securitise misogyny” (p.48) while Recommendation 1 directs ASIO to establish a dedicated analytical stream. The tension between the two is not addressed, and it is the first thing an unsympathetic reader will find.

6.3 The intervention may transmit the grievance it describes

Chapter 4 documents, using Kimmel’s aggrieved-entitlement framework, that the dominant public response to the preliminary findings was the claim that boys are being demonised (p.34). Those findings were released through News Corp mastheads and Triple J (p.34) as a statement about what a third of Australian boys believe. The report identifies the grievance narrative as a mechanism of radicalisation, and then supplies that mechanism with a highly quotable input. Whether the net effect of the communication strategy is preventive or iatrogenic is an empirical question the report raises and does not test.

6.4 Contested findings stated with uncontested confidence

The report records that around 65% of public comment on its preliminary findings was hostile and that commenters specifically attacked the methodology (p.34). A final report carrying a relabelled scale behind a headline number, an unexplained asterisk and a figure that contradicts its own summary table gives that criticism a target which is legitimate rather than merely reactive. The predictable result is not that the critics are persuaded, but that the general reader concludes the whole area is contested and disengages. Where the underlying problem is real, that is a net loss.

6.5 Material that travels badly

Two pages of uninterpreted ethnic-group breakdowns with cells as small as 34 respondents, and verbatim teacher quotations attributing misogyny to named national and ethnic groups (p.39), sit inside a report whose Chapter 4 correctly identifies ethnic deflection as a principal bad-faith response to this research. Publishing the first without interpretation makes the second materially harder to resist. The report’s own analysis is the best available answer to that deflection; leaving it unwritten on pages 20 and 21 forfeits it.

7. What would settle the question

Six remediations, none of them expensive

- **Publish the analysis dataset, model specifications and code**, or an appendix carrying them. Four of the nine published models currently cannot be reconciled with the figure on the facing page.
- **Re-run the nine models with age and sex as controls**, and report variance inflation factors or condition indices for the gender-attitude block.
- **Report reliability coefficients and factor loadings**, and run a confirmatory factor analysis testing whether the gender-attitude items and the extremism items load on distinguishable factors. The central claim turns on this and on nothing else.

- **Restore the scale names defined on p.17, or state the mapping**, and recompute the right-wing extremism headline from a scale that measures right-wing extremism.
- **Attach confidence intervals to every prevalence figure**, state the youth weighting, and report a sensitivity analysis for mischievous responding.
- **Before any screening instrument is built under Recommendation 2**, publish an indicative sensitivity, specificity and positive predictive value analysis at the proposed threshold.

8. Conclusion

The concern behind this report is real, the adolescent sample is a genuine contribution, and several of its findings deserve to survive, in particular the null result for general social media use and the teacher observations on differential authority. Nothing here suggests otherwise.

But the report's headline claims are stated at a level of confidence its published analysis does not support, and the two security recommendations would move a national counter-terrorism definition and build a screening instrument on the strength of a model whose own figure and summary table do not agree. The appropriate response is not to dismiss the work. It is to hold the recommendations until the appendix is published, and to say so plainly, because the alternative is that the case for a real problem is made on evidence that will not hold, in front of an audience already looking for a reason to dismiss it.

Annex A — SWOT

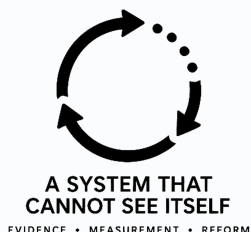
Strengths and weaknesses internal to the report; opportunities and threats external to it

Strengths	Weaknesses
• Rare 13–17 sample with a matched adult comparison, ethically obtained • Item wording published in full • Ethnic and religious attribution tested rather than avoided • Robust null on general social media use • Teacher data behaviourally grounded and convergent with prior Australian work • Recommendations operational and addressed to named agencies	• Central model omits age and sex • Figure and summary table disagree on four of nine models, including a sign reversal • Predictor and outcome items overlap; no reliability, loading or collinearity statistics • Headline scale relabelled and reduced to one item • Agreement threshold set at slight assent; no confidence intervals • Qualitative chapters carry no stated method • Youth online dating data not face-valid, yet carries a headline claim

Opportunities	Threats
• Lands in a live policy window alongside the 2025 counter-terrorism strategy and the under-16 social media measure • The content-not-platform finding is a genuine corrective to screen-time framing • Stage-one design invites a second wave that would resolve the causal question outright • Teacher findings can support professional development funding independently of the security framing • Recommendation 3's inter-agency cluster is the cheapest and most defensible of the five	• Base-rate failure if Recommendation 2 is implemented as a screen • Definitional creep from conduct to opinion under Recommendation 1 • Methodological soft spots handed to critics the report shows are already mobilised • Backfire through the aggrieved-entitlement channel the report itself documents • Single-author, unreplicated and not peer-reviewed at the point of policy adoption • Uninterpreted ethnic data fuelling the deflection the report opposes

About DFV Risk

Evidence • Measurement • Reform



Scan to visit
dfvrisk.com.au

DFV Risk publishes research and public-policy analysis on domestic and family violence response systems in Australia. Its central argument is that a system which does not measure its own performance cannot reform itself, and that the first task of reform is therefore measurement. The same standard is applied to research: a finding that cannot be checked cannot be built on.

Julian Talbot is a risk management specialist and lead architect of the Security Risk Management Body of Knowledge (SRMBOK).

Three books, free to download:

A System That Cannot See Itself — Summary

The short-form edition: the findings, the gaps and the recommendations in brief.

A System That Cannot See Itself — the full book

Domestic and family violence in the Australian Capital Territory, 2016–2026, and how to fix it.

Risk Management Companion Guide

The practitioner companion: applying the framework to real risk decisions.

All three are available at dfvrisk.com.au

Source. All page references are to Meger, S. (2026). *Misogyny, Online Radicalisation, and Youth Extremism in Australia*. Research Report. The University of Melbourne, June 2026 (web edition, 51 pp.), produced with funding from the Office of National Intelligence under the National Intelligence Postdoctoral Fellowship scheme, project NIPG202412.

Method. Desk appraisal of the published document only. No re-analysis was performed; the underlying dataset is not publicly available. Coefficient comparisons at D3 and D5 were read from the published figure on p.43 at high resolution and checked against the table on p.44.

Standing. This is a methodological appraisal, not a rebuttal of the report's subject matter. Misogyny and youth radicalisation are serious problems in Australia on evidence independent of this report.