

CHAPTER IV

CONCLUSION

4.1 Conclusion

This study set out to address two interrelated research questions: first, what the distribution, thematic composition, and temporal trajectory of public sentiment on X (formerly Twitter) toward the allocation of the education budget for the Free Nutritious Meal Program (MBG) in the 2026 State Budget (APBN) look like; and second, how that sentiment reflects the level of public consensus as a dimension of policy legitimacy within Beetham's (1991) framework. The findings are derived from lexicon-based sentiment analysis of 2,037 tweets using the InSet Lexicon (Koto & Rahmaningtyas, 2017) in RStudio, covering a seventeen-month research period from January 2025 to May 2026. The findings from both the automated classification and the supplementary inter-rater reliability (IRR) testing are considered together in the conclusions presented below.

With respect to the first research question, the analysis reveals three interconnected findings concerning the distribution, thematic structure, and temporal development of the discourse. Of the 2,037 tweets collected, 1,360 (66.76%) were classified as negative, 581 (28.52%) as positive, and 69 (3.39%) as neutral. This distributional finding is reinforced by the corpus-level scoring statistics: the aggregate mean sentiment score of -5.405 with a median of -4.0 confirms that negative dominance extends beyond frequency to intensity. The mean score of the Negative category (-10.422) is nearly twice the absolute magnitude of the mean Positive score ($+5.695$), serving as a quantitative marker of the asymmetric opinion strength characterising the critical discourse. Word frequency analysis identified four thematic clusters that provide structural

evidence of how this sentiment is organised. The Fiscal-Contestation cluster, anchored by *anggaran* (2,232 occurrences), *dipotong* (351), and *dialihkan* (220), establishes that public discourse frames the policy primarily as a harmful reduction of education resources. The Attribution of Political Responsibility cluster connects critical evaluation to specific political actors, personalising the accountability dimension of public opposition. The Social Impact on Vulnerable Groups cluster, anchored by *guru* (231), *rakyat* (179), and notably *guru honorer*, personalises the perceived harm to identifiable and sympathetically framed victims. The Evaluation of MBG Programme cluster reveals that public criticism is directed at the financing mechanism of MBG rather than its welfare objectives, a distinction with significant implications for policy communication. Temporally, negative sentiment maintained consistent dominance across all seventeen months of the research period, intensifying rather than moderating as discourse volume grew, from 66.3% negative in February 2026 to 70.0% in April 2026, reaching a peak monthly volume of 745 tweets in March 2026 following the legislative confirmation of 25 February 2026.

With respect to the second research question, the operationalised indicators of public consensus derived from Beetham's (1991) third dimension of legitimacy, namely consent, produce a convergent and clear assessment: the education budget rationalisation policy faces a measurable and persistent public consensus deficit. The 28.52% level of expressed support, while analytically non-negligible as evidence that a meaningful minority actively consents to the policy, falls substantially short of the majority endorsement that would constitute robust consensual legitimacy. Conversely, the 66.76% level of expressed opposition, sustained consistently across all seventeen months and expressed with disproportionate intensity, constitutes the active withdrawal of consent that Beetham (1991) identifies as the primary marker of a legitimacy

deficit along the consent dimension. The policy is, in Beetham's (1991) terms, legally legitimate, formally grounded in Law Number 17 of 2025 on the 2026 APBN and Presidential Regulation Number 118 of 2025, but consensually contested: the expressed consent of the governed, as operationalised through aggregated X user sentiment, has not been secured among the majority of those who engaged in discourse on this policy issue.

A necessary methodological qualification must be stated clearly. The Inter-Rater Reliability (IRR) testing conducted in this study produced a Cohen's Kappa value of $\kappa = 0.0647$ ($z = 1.07$, $p = 0.284$), which falls within the "Slight Agreement" category according to Landis and Koch (1977) and below the minimum threshold of $\kappa \geq 0.61$ established at the outset of the research. This result indicates that the InSet Lexicon's automated classification and the researcher's independent manual judgment are not consistently aligned at the level of individual tweet classification. A closer examination of the disagreement pattern is, however, analytically instructive: of the 30 tweets automatically classified as Positive by the lexicon, 24 were independently assessed as Negative by the researcher upon manual reading. This pattern is consistent with the well-documented failure of word-level lexicon scoring to detect sarcasm and irony, rhetorical strategies particularly prevalent in political discourse on Indonesian social media (Lunando & Purwarianti, 2013; Sykora et al., 2020), in which literally positive language is deployed to communicate criticism or dissent. Crucially, despite this classificatory disagreement at the level of individual tweets, the directional finding remains consistent between both assessments: the automated InSet Lexicon classified 66% of sampled tweets as Negative, while the researcher's independent manual evaluation assessed 73% of the same sample as Negative. This convergence in the dominant sentiment direction preserves the substantive validity of the study's central conclusion, namely that public sentiment toward the education budget

rationalisation policy is predominantly and substantially negative, even as it identifies the automated tool's systematic overestimation of positive sentiment in sarcasm-laden contexts. The low Kappa value is therefore treated as a documented methodological limitation that qualifies the precision of the automated classification at the individual-tweet level, without overturning the aggregate directional conclusion that constitutes the study's primary empirical finding.

The implication of this validation finding for the conclusions reached under both research questions must be stated explicitly. For the first research question, the distributional proportions reported in this study (66.76 percent negative, 28.52 percent positive, 3.39 percent neutral) should be read as approximate rather than precise measures of the composition of the discourse. The category most affected is Positive: because lexicon misclassification is concentrated in that category, its reported share represents an upper bound on expressed support, and the actual proportion of genuinely supportive discourse is likely lower than the figure reported here. The Negative proportion is correspondingly likely to be understated. For the second research question, this asymmetry has a direct and favourable consequence: the consensus deficit assessed through Beetham's (1991) framework is, if anything, larger than the automated classification indicates. The conclusion that the policy lacks majority consensual support is therefore robust to the measurement error the validation procedure identified. What is not robust is the precision of any individual percentage, and no claim in this study should be read as resting on the exact magnitude of the positive or neutral shares. This qualification does not diminish the study's contribution; it defines it. The research establishes the direction, persistence, and thematic structure of public sentiment toward the education budget allocation for MBG with a degree of confidence that survives the

documented instrument limitations. It does not establish, and does not claim to establish, the distribution of that sentiment to the decimal point.

At the theoretical level, this study demonstrates that the integration of Lippmann's (1922) public opinion theory, Habermas's (1989) public sphere theory, Beetham's (1991) legitimacy framework, and Dryzek's (2000) discursive legitimacy theory constitutes a coherent and analytically productive architecture for the Government Science study of public sentiment toward contested policy decisions in the digital era. Lippmann's framework explains how public opinion is formed, not through direct policy analysis, but through pseudo-environments constructed from mediated language, as evidenced by the corpus's lexical preference for loss-framing terms such as *dipotong* and *dialihkan* over neutral alternatives. Habermas's framework establishes where that opinion is expressed and why it carries democratic weight: X satisfies three empirical criteria of a digital public sphere in this context, namely dialogical scale (521,689 retweets; 1,351,930 likes), reflexive responsiveness to policy events (discourse surged 276% following the February 2026 legislative confirmation), and sustained evaluative plurality (both critical and supportive discourses maintained presence across all seventeen months). Beetham's framework specifies what the aggregated expression of that opinion means for governance, revealing a measurable legitimacy gap between the policy's formal legal authority and its expressed public acceptance. Dryzek's framework illuminates how that deficit is discursively structured: three of the four dominant thematic clusters, constituting a discursive constellation of loss discourse, accountability discourse, and welfare discourse, are misaligned with the government's policy narrative and intensify rather than resolve over time. Taken together, these findings confirm that the education budget rationalisation policy for MBG in the 2026 APBN confronts a substantive and analytically documentable problem of public legitimacy, one carrying implications not only

for this specific policy, but for the broader question of how contested budgetary decisions ought to be governed and communicated in Indonesia's digitally mediated democratic context.

4.2 Recommendations

4.2.1 Recommendations for the Government and Policymakers

The findings of this study carry four specific and empirically grounded implications for the government's management of the MBG programme and its associated budgetary arrangement. These recommendations are derived directly from the thematic, quantitative, and temporal findings of the analysis, consistent with Dryzek's (2000) argument that discursive legitimacy is rebuilt through genuine policy responsiveness to the prevailing constellation of public discourses.

First, a targeted counter-framing communication strategy is needed. The Fiscal-Contestation cluster is the most prominent in the corpus, anchored by *dipotong* (cut) and *dialihkan* (diverted), indicating that the public's dominant mental model frames this policy as a harmful loss of education resources. The government's framing of MBG as educational welfare delivery failed to penetrate this loss-dominant discourse across the research period, as evidenced by the intensification of negative sentiment from 66.3% in February 2026 to 70.0% in April 2026. An effective communication strategy should not merely reassert the legitimacy of the budgetary arrangement but should directly and repeatedly address the loss perception, providing transparent, accessible, and comparative data clarifying what proportion of the education budget remains available for core educational activities after the MBG allocation, how the arrangement compares with school meal programmes in Japan (137 years) and Brazil (71 years), and what measurable nutritional and educational outcomes are targeted in return for the Rp223.55

trillion investment. Critically, the strategy must specifically address the guru honorer narrative: the prominent appearance of contract teachers in the corpus's vocabulary of social harm indicates that this group has become the personalised face of the policy's perceived victims. Concrete, publicly communicated data on the effects of the budgetary arrangement on teacher welfare and school operational grants (BOS) would directly target the most affectively resonant opposition framing in the corpus.

Second, the government should publish a formal disaggregated breakdown of the education budget. The current arrangement, in which Rp223.55 trillion (29.07% of the total education budget) is classified under the education envelope without clear delineation from core educational expenditures, produces a structural informational ambiguity that fuels the loss discourse and provides the basis for the Constitutional Court petition. A publicly accessible, disaggregated budget presentation that clearly differentiates MBG operational costs from allocations to teacher welfare, school operational grants, higher education, and educational infrastructure would reduce the informational foundation of the dominant critical pseudo-environment. This is not merely an administrative measure; it constitutes a direct application of Lippmann's (1922) insight that the quality of public opinion is determined by the quality of the mediated information available to citizens rather than by their direct engagement with policy documents.

Third, structured mechanisms for public deliberation should be established or strengthened for contested budgetary decisions of this magnitude. The temporal analysis reveals that critical public opinion on X had already emerged as early as mid-2025, several months prior to the formal legislative confirmation of February 2026, indicating that the opposition represented a gradually consolidating discursive formation rather than a sudden post-confirmation reaction. Earlier and more structured engagement

with this emerging discourse, through parliamentary public hearings, formally constituted multi-stakeholder dialogue processes, or accessible digital deliberation channels convened prior to the final budget classification being enacted, would have allowed the government to identify and address the core concerns that subsequently consolidated into the dominant critical framing documented in this study. Within Habermas's (1989) normative framework, this represents the fulfilment of the public sphere's democratic function: not as a space in which citizens merely react to completed policy decisions, but as a space whose discursive outputs genuinely inform the formulation of decisions yet to be made.

Fourth, the Constitutional Court petition should be treated as a governance signal rather than merely as a legal matter. The petition's core argument, that the inclusion of MBG funding within the constitutionally mandated education budget dilutes the guarantee of adequate education financing, is precisely the concern that dominates the most frequently expressed and most intensely scored negative discourse on X. The government's substantive response to the petition, whether through legislative clarification, the establishment of protected spending floors for core educational programmes, or strengthened mechanisms ensuring that the MBG classification does not reduce aggregate spending on teacher welfare and school operations, would directly address the Accountability discourse cluster identified in Section 3.2.2. A transparent and constructive institutional response would also provide the empirical basis for a more effective counter-narrative on X, as it would offer concrete governance evidence that the dominant public concerns have been heard, acknowledged, and acted upon, which constitutes the foundational precondition for rebuilding the consensual legitimacy that this study identifies as structurally weakened.

4.2.2 Recommendations for Future Research

This study makes a methodologically and theoretically grounded contribution to the study of public policy legitimacy in the digital era, but its findings are bounded by constraints that define a productive agenda for future research.

First, and most urgently, future research should adopt machine learning or deep learning classification approaches to address the sarcasm and irony limitations documented in this study. The Cohen's Kappa result of $\kappa = 0.0647$ provides empirical quantification of the extent to which the InSet Lexicon systematically misclassifies sarcastic positive language as genuine positive sentiment, a limitation that produced a 24/30 misclassification rate for positive-labelled tweets in the IRR sample. This finding is not merely a methodological note; it constitutes an empirically grounded argument for the adoption of contextual classification approaches. Transformer-based models fine-tuned on Indonesian social media text, such as IndoBERT (Wilie et al., 2020) or other pre-trained models trained on Indonesian corpora, are significantly better equipped to capture contextual meaning, negation, irony, and figurative expression that word-level lexicon scoring systematically fails to detect. A direct comparative study applying both InSet Lexicon-based and IndoBERT-based classification to the same 2,037-tweet corpus used in this study would generate an empirically grounded assessment of the performance differential between the two approaches in this specific domain, namely politically charged Indonesian Twitter discourse on contested budgetary policy, and would itself constitute a methodologically valuable contribution to Indonesian computational social science.

Second, future research should extend the platform scope beyond X (formerly Twitter). This study's limitation to X is theoretically justified by the

platform's publicly accessible, text-based, and searchable architecture, which is specifically compatible with Habermasian public sphere criteria and with the InSet Lexicon's Indonesian-language coverage. However, the distributional pattern of public opinion on X may not accurately reflect the sentiment landscape across other platforms with different demographic profiles and content norms. Indonesian users engage extensively with political content across Instagram, TikTok, YouTube comment sections, and Facebook, each of which hosts distinct communities, content formats, and rhetorical conventions that may produce meaningfully different sentiment distributions on the same policy issue. Future multi-platform research applying a unified analytical framework and consistent keyword parameters would provide a more comprehensive mapping of how Indonesia's digital public sphere, understood as a multi-platform ecosystem rather than a single-platform arena, collectively responds to contested budgetary decisions.

Third, future research should operationalise the full three-dimensional legitimacy framework proposed by Beetham (1991). This study operationalises only the third dimension, consent or public consensus, as this is the only dimension measurable through sentiment analysis of naturally occurring social media text. The first dimension, legality, is currently being contested through the Constitutional Court petition, the outcome of which constitutes a critical pending development that future research should incorporate as a case event. The second dimension, normative justifiability, that is, whether the values and beliefs underlying the policy are genuinely shared between governors and the governed, requires qualitative methodological approaches beyond sentiment analysis, including in-depth interviews with key stakeholders (educators, contract teachers, policymakers, and civil society organisations), structured expert surveys, or mixed-method research designs triangulating computational sentiment findings with

deliberative or interpretive inquiry. A complete three-dimensional empirical legitimacy assessment would provide a substantially more comprehensive and governance-actionable analytical foundation than the consent-dimension analysis that computational methods alone can support, and would constitute a meaningful methodological advance in Government Science research on policy legitimacy in Indonesia.

Fourth, a longitudinal research design should be employed to track how discursive legitimacy evolves in response to specific governance events. The present study's temporal analysis demonstrates that the discursive legitimacy deficit intensified across the research period, a finding consistent with a governance dynamic in which the government's counter-narrative failed to reshape the prevailing discursive constellation. A longitudinal design extending data collection beyond May 2026, systematically coding specific events, including the Constitutional Court ruling, potential legislative amendments, and BGN's communication responses, against corresponding changes in sentiment distribution and thematic framing, would enable researchers to empirically test Dryzek's (2000) proposition that responsive governance can shift the discursive constellation of the public sphere, or conversely to document the conditions under which a discursive legitimacy deficit persists and deepens despite formal policy continuity. The MBG education budget case, with its combination of high public salience, ongoing legal contestation, and traceable policy events, represents a tractable and analytically productive case for longitudinal legitimacy tracking.

Fifth, the analytical model developed in this study should be replicated comparatively across other contested policy decisions in Indonesia. The integration of lexicon-based sentiment analysis with a four-layered theoretical framework, encompassing Lippmann (pseudo-environment), Habermas (digital public sphere), Beetham (policy legitimacy), and Dryzek (discursive

legitimacy), constitutes a replicable analytical architecture for Government Science research on contested public policy. Comparative applications to adjacent policy controversies, such as the 2025 general budget efficiency policy, regional transfer fund adjustments, or social protection reallocations, would test the generalisability of the findings across different policy domains and public engagement contexts. Such comparative work would progressively clarify which features of the discursive legitimacy deficit documented in this study are specific to the education budget rationalisation case and which reflect more generalised patterns in how Indonesian citizens digitally express evaluative orientations toward contested government decisions, thereby contributing to the development of a systematic, theoretically grounded tradition of computational legitimacy analysis within Government Science in Indonesia.