Arus Jurnal
Psikologi dan Pendidikan (AJPP)
Website: http://jurnal.ardenjaya.com/index.php/ajpp
Email: jurnal.ajpp@ardenjaya.com

Hate Speech Scale: Measuring Instrument Construction

INFO PENULIS

Lutfia Fitriani
Universitas Islam Negeri Sunan Gunung Djati Bandung
Lutfiafitriani2311@gmail.com

Aulia 'Azizatun Nisaa
Universitas Islam Negeri Sunan Gunung Djati Bandung

Dahlia Nafisa Julyanita
Universitas Islam Negeri Sunan Gunung Djati Bandung

Muhammad Arqan Zahran Putra Yuliandry
Universitas Islam Negeri Sunan Gunung Djati Bandung

Ryanaila Shaka Diba Suryana
Universitas Islam Negeri Sunan Gunung Djati Bandung

Tiara Cahaya Kirani
Universitas Islam Negeri Sunan Gunung Djati Bandung

Tahrir Tahrir
Universitas Islam Negeri Sunan Gunung Djati Bandung

INFO ARTIKEL

ISSN: 2963-8933
Vol. 5, No. 2 Juni 2026
http://jurnal.ardenjaya.com/index.php/ajpp

Saran Penulisan Referensi:

Fitriani, L., Nisaa, A. A., Julyanita, D. N., Yuliandry, M. A. Z. P., Suryana, R. S. D., Kirani, T. C., & Tahrir, T. (2026). Hate Speech Scale: Measuring Instrument Construction. *Arus Jurnal Psikologi dan Pendidikan (AJPP), 5*(2), 1668-1679.

Abstrak

Perkembangan media sosial yang pesat memunculkan fenomena ujaran kebencian yang mengancam kualitas diskursus publik, termasuk di kalangan mahasiswa. Namun, instrumen pengukuran ujaran kebencian yang terstandar dan sesuai dengan konteks lokal Indonesia, khususnya di lingkungan perguruan tinggi Islam, masih terbatas. Penelitian ini bertujuan mengkonstruksi serta menguji validitas dan reliabilitas alat ukur ujaran kebencian berdasarkan teori Bhikhu Parekh (2012) pada mahasiswa UIN Sunan Gunung Djati Bandung. Hipotesis penelitian menyatakan bahwa instrumen yang dikembangkan memiliki validitas isi, validitas konstruk, reliabilitas, dan kecocokan model yang baik. Penelitian menggunakan metode kuantitatif dengan desain terapan. Sebanyak 721 mahasiswa berpartisipasi melalui teknik accidental sampling. Konstruksi alat ukur mengacu pada tiga tahap Loevinger, yaitu validitas substantif, struktural, dan eksternal. Hasil penelitian menunjukkan bahwa seluruh 7 item memiliki validitas isi yang baik (Aiken's V > 0,7), daya beda item tinggi (corrected item-total correlation > 0,3), dan reliabilitas sangat baik (Cronbach's Alpha = 0,945). Analisis CFA mengkonfirmasi tiga dimensi utama, yaitu *Targeted Against Identifiable Group, Stigmatization, dan Legitimizes Hostility*, dengan *loading factor* > 0,94 serta indeks kecocokan model yang baik (RMSEA = 0,073; NFI, NNFI, CFI, IFI ≥ 0,99). Nilai CR sebesar 0,96 dan VE 0,76 memenuhi kriteria. Validitas konvergen ditunjukkan oleh korelasi positif kuat dengan *cyber bullying* (r = 0,748), sedangkan validitas diskriminan ditunjukkan oleh korelasi negatif dengan etika digital, integritas moral, dan kecerdasan moral. Instrumen ini terbukti valid, reliabel, dan layak digunakan untuk mengukur kecenderungan ujaran kebencian pada mahasiswa.

Kata Kunci*:* Ujaran Kebencian, Konstruksi Alat Ukur, Mahasiswa

Abstract

The rapid development of social media has given rise to the phenomenon of hate speech, which threatens the quality of public discourse, including among university students. However, standardized hate speech measurement instruments that are appropriate to the local Indonesian context, particularly within Islamic higher education environments, remain limited. This study aims to construct and test the validity and reliability of a hate speech measurement tool based on Bhikhu Parekh's (2012) theory among students of UIN Sunan Gunung Djati Bandung. The research hypothesis states that the developed instrument possesses good content validity, construct validity, reliability, and model fit. The study employed a quantitative method with an applied design. A total of 721 students participated through accidental sampling. The construction of the measurement instrument followed Loevinger's three stages: substantive, structural, and external validity. The results showed that all 7 items had good content validity (Aiken's V > 0.7), high item discrimination (corrected item-total correlation > 0.3), and excellent reliability (Cronbach's Alpha = 0.945). Confirmatory factor analysis (CFA) confirmed the three main dimensions Targeted Against Identifiable Group, Stigmatization, and Legitimizes Hostility with factor loadings > 0.94 and good model fit indices (RMSEA = 0.073; NFI, NNFI, CFI, IFI ≥ 0.99). The CR value of 0.96 and VE of 0.76 met the criteria. Convergent validity was demonstrated by a strong positive correlation with cyberbullying (r = 0.748), while discriminant validity was shown by negative correlations with digital ethics, moral integrity, and moral intelligence. This instrument is proven to be valid, reliable, and suitable for measuring hate speech tendencies among university students

Keywords: Hate Speech, Instrument Construction, University Students

A. Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang pesat telah mengubah cara manusia berinteraksi. Media sosial kini menjadi ruang publik digital yang digunakan miliaran orang untuk berbagi informasi dan berpendapat. Namun di balik kemudahan tersebut, muncul fenomena yang mengancam kualitas diskursus publik, yakni ujaran kebencian atau *hate speech*. Tontodimamma et al. (2021) mencatat bahwa pertumbuhan eksponensial media sosial membawa serta meningkatnya penyebaran ujaran kebencian dan propaganda berbasis kebencian di berbagai platform daring.

Fenomena ini tidak lagi terbatas pada interaksi tatap muka, melainkan berkembang menjadi masalah sistemik yang memengaruhi individu, kelompok, hingga tatanan sosial secara luas. Bahkan, penelitian Gennaro et al. (2025) menunjukkan bahwa meskipun proporsi pengguna yang memproduksi ujaran kebencian relatif kecil, dampaknya tetap signifikan karena terkonsentrasi dan masif.

Secara konseptual, Bhikhu Parekh (2012) memberikan definisi yang komprehensif dan banyak dirujuk. Menurutnya, ujaran kebencian adalah ekspresi yang mendorong, membangkitkan, atau menghasut kebencian terhadap sekelompok individu yang dibedakan berdasarkan ciri tertentu seperti ras, etnis, jenis kelamin, agama, kebangsaan, dan orientasi seksual.

Parekh (2012) menegaskan tiga ciri penting ujaran kebencian. Pertama, diarahkan terhadap individu atau kelompok berdasarkan atribut normatif yang tidak relevan. Kedua, menstigmatisasi kelompok sasaran dengan mengaitkan kualitas-kualitas yang dianggap sangat tidak diinginkan. Ketiga, menjadikan kelompok sasaran sebagai kehadiran yang tidak diinginkan dan objek permusuhan yang dianggap sah oleh pembuatnya, sehingga dapat melegitimasi diskriminasi.

Dampak ujaran kebencian tidak dapat dipandang remeh. Parekh (2012) menjelaskan bahwa ujaran kebencian menurunkan kualitas debat publik, memperburuk kepekaan moral masyarakat, dan melemahkan budaya saling menghormati. Ujaran kebencian juga melanggar martabat anggota kelompok sasaran dengan menstigmatisasi mereka dan mengabaikan individualitas mereka.

Pada tataran individual, korban ujaran kebencian mengalami tekanan psikologis: merasa gugup di ruang publik, takut mengungkapkan pendapat, dan cenderung terasing dari masyarakat luas. Pada tataran sosial yang lebih luas, kebencian terhadap suatu kelompok berkembang perlahan melalui ucapan-ucapan yang terisolasi, yang secara kumulatif dapat meracuni pikiran kaum muda dan melemahkan norma kesopanan (Parekh, 2012).

Fakta empiris menunjukkan keseriusan persoalan ini. Gennaro et al. (2025) menemukan bahwa 1% pengguna bertanggung jawab atas 46% ujaran kebencian di Twitter Swiss, dan 5% pengguna menghasilkan hingga 83% dari seluruhnya. Pola serupa ditemukan pada surat kabar daring, di mana 1% pengguna memproduksi 56–70% ujaran kebencian.

Dari sisi akademik, Tontodimamma et al. (2021) mencatat pertumbuhan publikasi ilmiah tentang ujaran kebencian dengan laju tahunan 20,5%. Alkomah dan Ma (2022) juga menegaskan bahwa ujaran kebencian merupakan konten kompleks dan multidimensi yang menargetkan individu atau kelompok, dengan tantangan deteksi yang belum terpecahkan sepenuhnya.

Penelitian terdahulu telah berkembang dari berbagai sudut pandang. Parekh (2012) mengonstruksi argumen filosofis-normatif tentang regulasi hukum terhadap ujaran kebencian. Papcunová et al. (2023) mengoperasionalisasi ujaran kebencian melalui sepuluh indikator terukur, mencakup bahasa seksis, serangan terhadap minoritas, penyangkalan hak asasi manusia, promosi kekerasan, hingga *vulgarisme.*

Tontodimamma et al. (2021) melalui kajian bibliometrik mengidentifikasi tiga kluster penelitian ujaran kebencian: debat kebebasan berekspresi, deteksi otomatis berbasis *machine learning*, serta ujaran kebencian berbasis gender dan *cyberbullying*. Gennaro et al. (2025) mengisi celah dengan meneliti distribusi ujaran kebencian antar pengguna dan efektivitas strategi *counterspeech*.

Dalam penelitian ini, peneliti melakukan kajian literatur terhadap beberapa penelitian terdahulu yang berkaitan dengan pengukuran, identifikasi, dan pengembangan instrumen ujaran kebencian *(hate speech).* Abubakar dkk. (2016) mengembangkan Instrumen Monitoring *Hate Speech* (IM-HT) yang dirancang untuk mengamati, mendeteksi, dan mendokumentasikan ujaran kebencian di masyarakat. Instrumen ini memuat beberapa indikator *hate speech,* yaitu prasangka, penghinaan, pelabelan atau cap negatif, hasutan, dan ancaman. IM-HT dikembangkan sebagai pedoman monitoring untuk mendeteksi potensi terjadinya diskriminasi, *hate crime,* dan konflik sosial.

Penelitian lain dilakukan oleh Yulian dan Findawati (2024) yang mengidentifikasi ujaran kebencian pada media sosial Twitter menggunakan algoritma Naïve Bayes. Penelitian tersebut menggunakan 3.972 data tweet dan menghasilkan tingkat akurasi sebesar 88,9%. Hasil penelitian menunjukkan bahwa ujaran kebencian dapat diidentifikasi dan diklasifikasikan secara otomatis melalui pendekatan komputasional berbasis *machine learning.*

Selanjutnya, Putri dkk. (2024) mengembangkan instrumen *hate speech* yang awalnya terdiri dari 48 item. Setelah melalui proses pengujian, diperoleh 23 item final yang memenuhi

kriteria pengukuran. Instrumen tersebut menunjukkan reliabilitas yang sangat baik dengan nilai Cronbach's Alpha sebesar 0,923 sehingga dinilai memiliki konsistensi internal yang tinggi dalam mengukur *hate speech.* Toussaint dkk. (2020) mengembangkan *Hateful Emotional Responses Scale* (HatERS) untuk mengukur respons emosional kebencian pada individu. Instrumen ini terdiri atas lima item dan dirancang sebagai skala unidimensional. Hasil *Confirmatory Factor Analysis* menunjukkan bahwa seluruh item memiliki muatan faktor yang signifikan dan nilai reliabilitas Cronbach's Alpha sebesar 0,87, sehingga instrumen dinilai valid dan reliabel untuk mengukur respons emosional kebencian.

Selain itu, Mardia, Aisha, dan Dimala (2023) menyusun skala ujaran kebencian berdasarkan konsep yang dikemukakan oleh Parekh. Instrumen tersebut dikembangkan berdasarkan tiga aspek ujaran kebencian, yaitu diarahkan pada individu atau kelompok tertentu, menciptakan stigma, dan memunculkan tindakan diskriminasi. Skala awal terdiri atas 21 item pernyataan, namun setelah dilakukan uji validitas diperoleh 20 item yang valid untuk digunakan. Hasil pengujian reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,912 yang mengindikasikan bahwa instrumen memiliki reliabilitas yang sangat baik.

Berdasarkan berbagai penelitian terdahulu tersebut, dapat diketahui bahwa *hate speech* dan perilaku kebencian telah diteliti melalui berbagai pendekatan, mulai dari pengembangan instrumen psikologis, instrumen monitoring, penggunaan instrumen dalam penelitian psikologi, hingga pendekatan klasifikasi berbasis teknologi. Hal ini menunjukkan bahwa penelitian mengenai *hate speech* telah berkembang dalam berbagai konteks dan metode pengukuran.

Meskipun penelitian tentang konstruksi alat ukur ujaran kebencian telah berkembang pesat, belum ditemukan instrumen yang secara khusus dikonstruksi dan diuji pada mahasiswa UIN Sunan Gunung Djati Bandung. Sebagian besar penelitian terdahulu dilakukan pada populasi umum, pengguna media sosial, atau kelompok lain yang memiliki karakteristik berbeda dengan mahasiswa.

Oleh karena itu, penelitian ini memfokuskan pada konstruksi alat ukur ujaran kebencian pada mahasiswa UIN Sunan Gunung Djati Bandung. Hal ini dilakukan untuk memperoleh instrumen yang sesuai dengan karakteristik mahasiswa sehingga mampu mengukur kecenderungan ujaran kebencian secara lebih akurat dan kontekstual. Sebagian besar penelitian berfokus pada platform berskala besar dengan konteks Barat dan data pengguna umum yang heterogen. Kajian tentang ujaran kebencian di kalangan mahasiswa pada institusi pendidikan tinggi Islam di Indonesia masih sangat terbatas.

Mahasiswa UIN Sunan Gunung Djati Bandung merupakan kelompok yang unik karena berada di persimpangan nilai-nilai keislaman, konteks multikulturalisme Indonesia, dan penggunaan media sosial yang intensif sebagai generasi *digital native.* Belum ada penelitian yang secara khusus memetakan persepsi dan perilaku ujaran kebencian pada kelompok ini menggunakan kerangka indikator yang terstandarisasi.

Urgensi penelitian ini bertumpu pada dua hal. Pertama, Indonesia sebagai negara dengan pengguna media sosial terbesar di dunia menghadapi tantangan serius terkait penyebaran ujaran kebencian yang dapat mengancam persatuan sosial. Kedua, kampus sebagai ruang intelektual seharusnya menjadi wadah diskursus yang sehat, namun tidak serta-merta imun terhadap fenomena ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mendeskripsikan dan menganalisis ujaran kebencian di kalangan mahasiswa UIN Sunan Gunung Djati Bandung menggunakan indikator dan dimensi yang dikembangkan dari kerangka teoritis Bhikhu Parekh (2012), yakni stigmatisasi kelompok sasaran, penghasutan kebencian, dan legitimasi diskriminasi terhadap kelompok tertentu.

Selain kajian konseptual mengenai ujaran kebencian, sejumlah penelitian juga telah mengembangkan instrumen untuk mengukur *hate speech* secara empiris dan terstandarisasi. Salah satu instrumen yang cukup banyak digunakan adalah *Hate Behaviours Scale* (HBS) yang dikembangkan oleh Vergani et al. (2023). Instrumen ini disusun berdasarkan teori prasangka Allport yang memandang ekspresi kebencian sebagai suatu spektrum perilaku, mulai dari antilocution hingga bentuk kekerasan yang lebih ekstrem. HBS terdiri atas 12 item yang terbagi ke dalam tiga dimensi, yaitu *discrimination, defensive violence,* dan *belligerent violence.* Hasil pengujian menunjukkan bahwa instrumen ini memiliki reliabilitas dan validitas yang baik serta mampu memprediksi berbagai perilaku kebencian dalam kehidupan nyata.

Instrumen lain dikembangkan oleh Sachdeva et al. (2022) melalui *Measuring Hate Speech Corpus.* Berbeda dengan HBS yang berfokus pada individu, instrumen ini dirancang untuk menilai tingkat ujaran kebencian dalam konten media sosial. Pengukuran dilakukan melalui sepuluh indikator, seperti sentimen negatif, penghinaan, dehumanisasi, kekerasan, hingga

genosida. Setiap konten ditempatkan pada suatu spektrum *hate speech* menggunakan pendekatan *Rasch Measurement Theory.* Instrumen ini banyak digunakan dalam penelitian yang berkaitan dengan deteksi otomatis ujaran kebencian dan pengembangan model *machine learning.*

Meskipun kedua instrumen tersebut memberikan kontribusi penting dalam pengukuran ujaran kebencian, terdapat perbedaan mendasar dengan alat ukur yang dikembangkan dalam penelitian ini. HBS berfokus pada kecenderungan perilaku individu untuk melakukan tindakan kebencian, sedangkan *Measuring Hate Speech Corpus* berfokus pada penilaian tingkat kebencian dalam suatu konten digital. Sementara itu, alat ukur yang dikembangkan dalam penelitian ini berlandaskan teori Bhikhu Parekh (2012) dan dirancang untuk mengukur kecenderungan ujaran kebencian berdasarkan struktur konseptual *hate speech* itu sendiri.

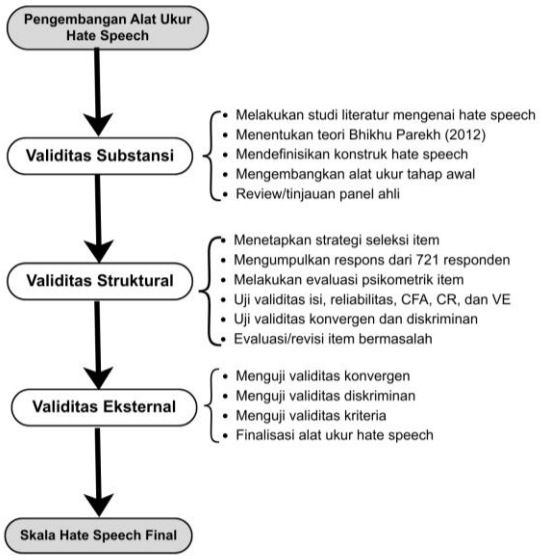
Perbedaan lainnya terletak pada landasan teoritis yang digunakan. HBS berakar pada teori prasangka dalam psikologi sosial, sedangkan *Measuring Hate Speech Corpus* lebih menekankan pendekatan empiris berbasis penilaian kolektif terhadap konten digital. Sebaliknya, alat ukur yang dikembangkan dalam penelitian ini berangkat dari kerangka filsafat politik normatif yang secara eksplisit mendefinisikan apa yang dimaksud dengan ujaran kebencian. Berdasarkan teori Parekh (2012), hate speech dipahami melalui tiga dimensi utama, yaitu penargetan terhadap kelompok berdasarkan atribut identitas tertentu *(targeted against identifiable group),* stigmatisasi kelompok sasaran *(stigmatization),* dan legitimasi permusuhan terhadap kelompok tersebut *(legitimizes hostility).*

Selain itu, instrumen-instrumen sebelumnya umumnya dikembangkan dan diuji dalam konteks masyarakat Barat.

Adapun alat ukur dalam penelitian ini dirancang dan divalidasi pada mahasiswa Indonesia, khususnya di lingkungan perguruan tinggi Islam. Dengan demikian, instrumen ini diharapkan lebih mampu merepresentasikan dinamika sosial, budaya, dan keagamaan yang berkembang dalam konteks lokal Indonesia. Di samping itu, tiga dimensi yang ditawarkan Parekh memberikan struktur konseptual yang relatif ringkas namun tetap mampu menangkap esensi ujaran kebencian, sehingga lebih operasional untuk dikembangkan menjadi item-item pengukuran pada populasi mahasiswa.

B. Metodologi

Metode kuantitatif dengan desain penelitian terapan digunakan sebagai metode dalam penelitian ini. Jumlah responden adalah 721 dari mahasiswa UIN Sunan Gunung Djati Bandung. Langkah-langkah yang diambil dalam penelitian dalam proses konstruksi alat ukur ini mengacu pada Loevinger yang terdiri dari tiga tahap utama, yaitu validitas substantif, validitas struktural dan validitas eksternal (Loevinger, 1957).
Tahapan konstruksi alat ukur *hate speech* dalam penelitian ini adalah:



Tahap pertama adalah validitas substansi, yaitu pengembangan alat ukur berdasarkan teori dan temuan empiris yang relevan. Pada tahap ini dilakukan kajian literatur mengenai *hate speech* dengan mengacu pada teori Parekh (2012) yang terdiri atas dimensi *Targeted Against Identifiable Group, Stigmatization,* dan *Legitimizes Hostility.* Selanjutnya dirumuskan definisi

konseptual dan operasional konstruk, disusun indikator, *blueprint*, dan item instrumen, serta dilakukan penilaian oleh panel ahli untuk mengevaluasi kesesuaian definisi, indikator, dan item yang dikembangkan.

Tahap kedua adalah validitas struktural yang berfokus pada hubungan antar item dalam mengukur konstruk. Pada tahap ini dilakukan penetapan teknik pemilihan item, pengumpulan data dari 721 mahasiswa UIN Sunan Gunung Djati Bandung, serta analisis kualitas psikometrik instrumen. Analisis meliputi uji validitas isi, daya beda item, validitas konstruk, reliabilitas, *Confirmatory Factor Analysis* (CFA), *Construct Reliability* (CR), *Variance Extracted* (VE), validitas konvergen, validitas diskriminan, dan evaluasi kecocokan model. Hasil analisis menunjukkan bahwa seluruh item memenuhi kriteria sehingga tidak diperlukan revisi item.

Tahap ketiga adalah validitas eksternal. Pada tahap ini dilakukan pengujian instrumen menggunakan alat ukur pembanding melalui uji validitas konvergen, diskriminan, dan kriteria. Setelah seluruh tahapan pengujian terpenuhi, alat ukur *hate speech* ditetapkan sebagai instrumen final penelitian

C. Hasil dan Pembahasan

Hasil

Definisi Konsep, Dimensi, dan Indikator

Hate speech merupakan bentuk ekspresi yang mengandung permusuhan, kebencian, atau penghinaan terhadap individu maupun kelompok tertentu dalam bentuk perkataan, baik yang bersifat menyinggung, kasar, penuh amarah, maupun menghina. Selain itu, *hate speech* juga mencakup tindakan yang mendorong, membangkitkan, atau menghasut kebencian terhadap sekelompok individu yang dibedakan berdasarkan karakteristik tertentu, seperti ras, etnis, jenis kelamin, agama, kebangsaan, dan orientasi seksual. Dengan demikian, *hate speech* tidak hanya berfungsi sebagai ekspresi sikap negatif, tetapi juga berpotensi memperkuat prasangka, diskriminasi, dan permusuhan terhadap kelompok sasaran.

Berdasarkan definisinya, Peneliti menyatakan bahwa ada 3 dimensi *hate speech*, yaitu

1. *Targeted Against Identifiable Group*

Dimensi ini mengacu pada kecenderungan individu untuk mengarahkan ekspresi kebencian kepada individu atau kelompok berdasarkan karakteristik identitas tertentu yang sebenarnya tidak relevan dengan perilaku atau tindakan yang dilakukan oleh kelompok tersebut. Karakteristik tersebut meliputi ras, etnis, jenis kelamin, agama, kebangsaan, maupun orientasi seksual.

2. *Stigmatization*

Dimensi *stigmatization* mengacu pada kecenderungan memberikan stigma, label negatif, stereotip, atau karakteristik yang tidak diinginkan kepada kelompok sasaran. Stigma tersebut dapat diberikan secara eksplisit maupun implisit melalui pernyataan yang menggambarkan kelompok tertentu sebagai kelompok yang buruk, berbahaya, tidak bermoral, atau tidak layak diterima dalam masyarakat.

3. *Legitimizes Hostility*

Dimensi *legitimizes hostility* mengacu pada kecenderungan memandang kelompok sasaran sebagai pihak yang pantas menerima perlakuan negatif atau menjadi objek permusuhan yang sah. Pandangan ini muncul karena kelompok tersebut dianggap memiliki karakteristik negatif sehingga keberadaannya dipersepsikan sebagai sesuatu yang tidak diinginkan.

Berdasarkan ketiga dimensi *hate speech* yang sudah diidentifikasi, peneliti menyusun tujuh indikator yang merepresentasikan aspek-aspek penting dari masing-masing dimensi. Kemudian, setiap indikator dikembangkan ke dalam satu item pernyataan sehingga setiap item merepresentasikan ke dalam satu item pernyataan seperti dalam tabel 1 berikut.

No	Indikator	Item
1	Menyapa individu di ruang publik dengan sebutan yang meremehkan identitas sosialnya (ras, etnis)	Saya menyapa individu di ruang publik dengan sebutan yang meremehkan identitas sosialnya (ras, etnis)
2	Menggunakan kata-kata kasar pada suatu kelompok yang mengarah pada ciri fisik bawaan	Saya menggunakan kata-kata kasar pada suatu kelompok yang mengarah pada ciri fisik bawaan

No	Indikator	Item
3	Menyatakan secara terbuka suatu kelompok memiliki sifat buruk yang dianggap sebagai identitas sosialnya.	Saya terus terang menyebut suatu kelompok punya sifat buruk yang sudah jadi ciri khas mereka
4	Menggunakan candaan untuk menampilkan kesan kelompok tertentu tidak dapat dipercaya	Saya memakai candaan untuk membentuk rasa tidak percaya orang lain terhadap kelompok tertentu.
5	Melabeli suatu kelompok dengan sifat negatif yang dianggap melekat secara permanen.	Saya melabeli suatu kelompok dengan sifat negatif yang dianggap melekat secara permanen
6	Membenarkan tindakan diskriminasi terhadap kelompok yang dianggap tidak diinginkan.	Saya membenarkan tindakan diskriminasi terhadap kelompok yang dianggap tidak diinginkan
7	Mengabaikan kehadiran suatu kelompok karena identitas sosialnya berbeda	Saya mengabaikan kehadiran suatu kelompok karena identitas sosialnya berbeda

Tabel 1 Item Alat Ukur Hate Speech

Analisis Kasus

a. Uji Validitas Konten Aiken's

Setelah penulisan item, langkah selanjutnya adalah melakukan uji validitas isi dengan cara menyebarkan item-item tersebut kepada 6 orang panel ahli untuk menilai apakah setiap item mampu mengukur indikator yang ingin diukur atau tidak. Uji validitas isi mengacu pada teknik yang dirumuskan oleh Aiken's. Dengan menggunakan rumus validitas isi Aiken's, peneliti menyimpulkan hasil uji validitas isi sebagai berikut

No	*Statement*	*Value*	*Conclusion*
1	Saya menyapa individu di ruang publik dengan sebutan yang meremehkan identitas sosialnya (ras, etnis)	0,9	Valid
2	Saya menggunakan kata-kata kasar pada suatu kelompok yang mengarah pada ciri fisik bawaan	0,8	Valid
3	Saya terus terang menyebut suatu kelompok punya sifat buruk yang sudah jadi ciri khas mereka	0,9	Valid
4	Saya memakai candaan untuk membentuk rasa tidak percaya orang lain terhadap kelompok tertentu.	0,8	Valid
5	Saya melabeli suatu kelompok dengan sifat negatif yang dianggap melekat secara permanen	0,8	Valid
6	Saya membenarkan tindakan diskriminasi terhadap kelompok yang dianggap tidak diinginkan	0,9	Valid
7	Saya mengabaikan kehadiran suatu kelompok karena identitas sosialnya berbeda	0,9	Valid

Tabel 2 Hasil Uji Validitas Konten *Aiken's*

Hasil uji validitas isi menunjukkan bahwa semua item memiliki nilai V di atas 0,7, melebihi batas minimum, sehingga validitas isi dapat diterima. Dengan demikian, alat ukur *Hate Speech* dapat diterima dan layak digunakan untuk mengukur variabel yang diukur.

b. Uji Beda Item (Scale)

No	*Statement*	*Corrected item-total correlation*	*Conclusion*
1	Saya menyapa individu di ruang publik dengan sebutan yang meremehkan identitas sosialnya (ras, etnis)	0,750	*Accepted (High Correlation)*
2	Saya menggunakan kata-kata kasar pada suatu kelompok yang mengarah pada ciri fisik bawaan	0,848	*Accepted (High Correlation)*
3	Saya terus terang menyebut suatu kelompok punya sifat buruk yang sudah jadi ciri khas mereka	0,811	*Accepted (High Correlation)*
4	Saya memakai candaan untuk membentuk rasa tidak percaya orang lain terhadap kelompok tertentu.	0,833	*Accepted (High Correlation)*
5	Saya melabeli suatu kelompok dengan sifat negatif yang dianggap melekat secara permanen	0,816	*Accepted (High Correlation)*
6	Saya membenarkan tindakan diskriminasi terhadap kelompok yang dianggap tidak diinginkan	0,845	*Accepted (High Correlation)*
7	Saya mengabaikan kehadiran suatu kelompok karena identitas sosialnya berbeda	0,823	*Accepted (High Correlation)*

Tabel 3 Hasil Uji Beda Item

Berdasarkan hasil uji di atas, semua item memiliki nilai r > 3. Hal ini berarti bahwa item-item *hate speech* mampu untuk membedakan jawaban yang cukup baik sesuai dengan yang dipersyaratkan. Oleh karena itu, semua item tersebut dapat membedakan individu yang memiliki karakteristik yang hendak diukur. Selanjutnya, item-item tersebut dapat digunakan untuk melakukan pengukuran.

c. Uji Validitas Multidimensi (Korelasi)

Setelah menguji validitas isi, peneliti membagikan kuesioner kepada 721 responden. Untuk menilai konsistensi internal dan validitas, analisis korelasi item-total dilakukan dengan menggunakan analisis korelasi *Pearson Product Moment*. Evaluasi keselarasan antara setiap dimensi dengan konstruk keseluruhan dilakukan dengan mengkorelasikan skor total dimensi terhadap skor total instrumen. Adapun hasil dari uji korelasi *item-total* tersebut dipaparkan sebagai berikut:

No	*Dimensional*	*Pearson Correlation*	*Conclusion*
1	*Targeted Against Identifiable Group*	0,906	Accepted (*High Correlation*)
2	*Stigmatization*	0,960	Accepted (*High Correlation*)
3	*Legitimizes Hostility*	0,932	Accepted (*High Correlation*)

Tabel 4 Hasil Uji Validitas Multi Dimensi

Melalui uji konsistensi internal menggunakan analisis korelasi *item-total*, diketahui bahwa ketiga dimensi memiliki nilai korelasi di atas 0,60. Hasil ini mengindikasikan adanya hubungan atau korelasi yang kuat antar-dimensi tersebut.

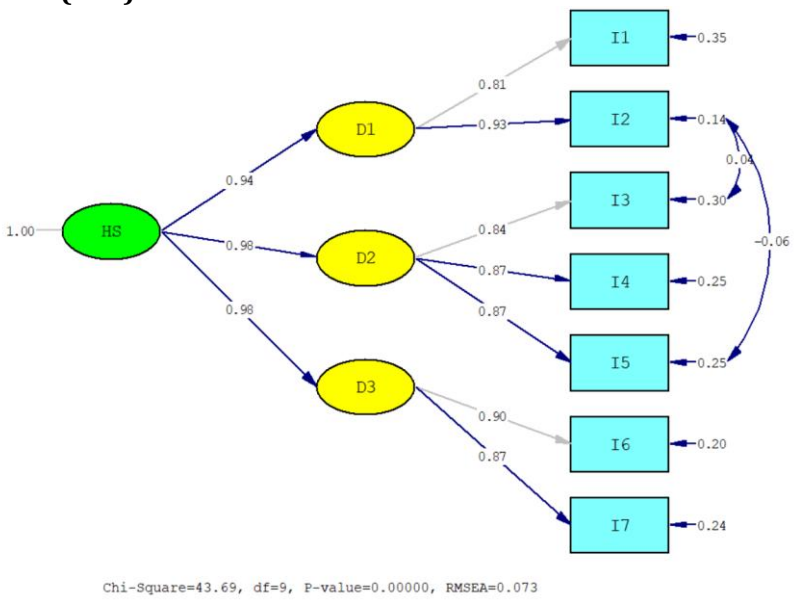
d. Uji Reliabilitas Alat Ukur

Uji reliabilitas dilakukan dengan tujuan untuk mengetahui sejauh mana konsistensi, stabilitas, dan keandalan alat ukur yang digunakan dalam penelitian ini ketika dilakukan pengukuran berulang. Untuk memperoleh nilai reliabilitas tersebut, peneliti menggunakan pendekatan koefisien *Cronbanch's Alpha*. Pengujian ini dipilih karena mampu mengevaluasi konsistensi internal dari item-item berskala multi-respons (seperti skala Likert). Sebuah instrumen atau alat ukur dikatakan andal dan memenuhi kriteria reliabilitas yang baik apabila nilai *Cronbanch's Alpha* yang diperoleh lebih besar dari 0,60 atau 0,70.

Berdasarkan hasil uji reliabilitas, diperoleh nilai *Cronbach's Alpha* sebesar 0,945. Hasil ini mengindikasikan bahwa alat ukur *hate speech* memiliki tingkat reliabilitas yang tinggi sehingga terbukti konsisten dan andal untuk mengukur konsep *hate speech*.

Analisis Modern

a. Uji Validitas Item (CFA)



Gambar 1 Hasil Output Lisrel CFA

Kemudian untuk mengetahui kesesuaian model pengukuran, dilakukan analisis *Confirmatory Factor Analysis* (CFA) tingkat kedua (2nd Order CFA). Berdasarkan hasil analisis diperoleh model pengukuran *Hate Speech* yang terdiri atas tiga dimensi, yaitu D1 (*Targeted Against Identifiable Group)*, D2 (*Stigmatization)*, dan D3 (*Legitimizes Hostility*) seperti yang ditunjukkan pada Gambar 1.

Berdasarkan gambar tersebut, setiap dimensi menunjukkan nilai *loading factor* yang signifikan terhadap konstruk *hate speech*, yakni sebesar 0,94 untuk D1, serta 0,98 untuk D2 dan D3. Di samping itu, seluruh item berhasil melampaui ambang batas 0,50 dengan rentang nilai antara 0,81 hingga 0,93. Hasil analisis ini mengindikasikan bahwa masing-masing item memiliki kemampuan yang sangat baik dalam merepresentasikan dimensi yang diukurnya.

Selain itu, hasil analisis juga menunjukkan nilai RMSEA sebesar 0,073 (< 0,08), yang mengindikasikan bahwa model pengukuran memiliki tingkat kesesuaian yang baik terhadap data empiris.

b. Validitas Alat Ukur (VE) dan Reliabilitas Alat Ukur (CR)

Setelah model fit cocok, langkah selanjutnya adalah menghitung reliabilitas *Construct Reliability* (CR) dan *Variance Extract* (VE) dengan rumus sebagai berikut:

Rumus CR

$$CR = \frac{(\Sigma \text{Standardized Loading})^2}{(\Sigma \text{Standardized Loading})^2 + (\Sigma \text{Measurement Error})^2} \quad (1)$$

Rumus VE

$$VE = \frac{\Sigma \text{Standardized Loading}^2}{\Sigma Standardized\ Loading^2 + \Sigma \text{Measurement Error}^2} \quad (2)$$

Nilai Reliabilitas CR yang diharapkan >0,70 dan >0,50. Berdasarkan pada hasil uji tersebut, Diperoleh nilai CR sebesar 0,96 dan VE sebesar 0,76. Dengan demikian, dapat disimpulkan bahwa alat ukur ini valid dan dianggap mampu mengukur *hate speech* secara konsisten.

c. Model Fit Ukur

Berdasarkan model teoritis yang mengusulkan tiga dimensi *hate speech*, dilakukan *Confirmatory Factor Analysis* (CFA) untuk menguji kesesuaian model pengukuran dengan data empiris. CFA dilakukan untuk mengonfirmasi apakah ketiga dimensi yang diusulkan mampu merepresentasikan konstruk *hate speech* secara memadai. Kesesuaian model dievaluasi menggunakan beberapa indeks fit, yaitu RMSEA, NFI, NNFI, CFI, dan IFI.

Fit Index	*Fit Value*	*Criterion*	*Conclusion*
Root Mean Square Error of Approximation (RMSEA)	0,073	<0,08	*Fit*
Normed Fit Index (NFI)	0,99	>0,90	*Fit*
Non-Normed Fit Index (NNFI)	0,99	>0,90	*Fit*
Comparative Fit Index (CFI)	1,00	>0,90	*Fit*
Incremental Fit Index (IFI)	1,00	>0,90	*Fit*

Tabel 6 Model Fit Indices

Berdasarkan hasil yang ditunjukkan pada Tabel 6, seluruh indeks kecocokan model memenuhi kriteria yang disyaratkan. Nilai RMSEA sebesar 0,073, sedangkan nilai NFI, NNFI, CFI, dan IFI masing-masing sebesar 0,99; 0,99; 1,00; dan 1,00. Oleh karena itu, dapat disimpulkan bahwa model pengukuran Hate Speech yang terdiri atas tiga dimensi memiliki tingkat kecocokan (*goodness of fit*) yang baik atau fit dengan data empiris. Hal tersebut membuktikan bahwa kerangka model yang diajukan mampu merefleksikan hubungan antara konstruk *hate speech*, seluruh dimensi, dan indikator penunjangnya secara representatif.

Uji Validitas Konvergen dan Diskriminan

Variabel	1	2	3	4	5	AVE	CR
Hate Speech	1					.76	.96
Cyber Bullying	.748**	1				.79	.96
Digital Ethic	-.095*	-.135**	1			.55	.93
Moral Integrity	-.087*	-.082*	.516**	1		.61	.9
Moral Intelligence	-.184**	-.126**	.603**	.532**	1	.64	.95

**. Correlation is significant at the 0.01 level (2-tailed).

Tabel 7 Hasil Uji Konvergen dan Diskriminan

Untuk memperkuat bukti validitas eksternal instrumen *Hate Speech*, dilakukan pengujian validitas konvergen dan validitas diskriminan melalui analisis korelasi Pearson dengan beberapa konstruk yang relevan secara teoritis.

Instrumen *Cyber Bullying, Digital Ethic, Moral Integrity*, dan *Moral Intelligence* dipilih sebagai alat ukur pembanding karena memiliki keterkaitan konseptual dengan konstruk *Hate Speech. Cyber Bullying* digunakan untuk menguji validitas konvergen karena merepresentasikan bentuk perilaku agresif dalam lingkungan digital yang secara teoritis berkaitan dengan hate speech. Sementara itu, *Digital Ethic*, *Moral Integrity*, dan *Moral Intelligence* digunakan untuk menguji validitas diskriminan karena merepresentasikan konstruk yang secara konseptual berbeda dengan hate speech.

Validitas konvergen dievaluasi dengan menguji hubungan antara *Hate Speech* dan *Cyber Bullying*. Hasil analisis menunjukkan adanya korelasi positif yang kuat dan signifikan antara *Hate Speech* dan *Cyber Bullying* ($r = 0{,}748$; $p < 0{,}01$). Temuan ini menunjukkan bahwa individu yang memiliki kecenderungan lebih tinggi dalam perilaku *hate speech* juga cenderung

menunjukkan tingkat *cyber bullying* yang lebih tinggi. Hubungan positif yang kuat tersebut mendukung asumsi teoritis bahwa kedua konstruk berada dalam domain perilaku agresi sosial di lingkungan digital sehingga memberikan bukti yang memadai terhadap validitas konvergen instrumen *Hate Speech*.

Selanjutnya, validitas diskri- minan diuji melalui hubungan antara *Hate Speech* dengan *Digital Ethic*, *Moral Integrity*, dan *Moral Intelligence*. Hasil analisis menunjukkan bahwa *Hate Speech* berkorelasi negatif dan signifikan dengan Digital Ethic ($r = -0,095$; $p < 0,05$), Moral Integrity ($r = -0,087$; $p < 0,05$), dan Moral Intelligence ($r = -0,184$; $p < 0,01$). Arah hubungan yang negatif menunjukkan bahwa semakin tinggi kecenderungan *hate speech*, semakin rendah etika digital, integritas moral, dan kecerdasan moral individu. Meskipun kekuatan hubungan yang ditemukan relatif rendah, arah korelasi yang negatif dan signifikan konsisten dengan landasan teoritis yang menyatakan bahwa perilaku *hate speech* bertentangan dengan etika digital dan nilai-nilai moral. Oleh karena itu, temuan ini menunjukkan bahwa konstruk *Hate Speech* dapat dibedakan dari konstruk-konstruk moral yang secara konseptual berbeda sehingga mendukung validitas diskriminan instrumen.

Selain itu, hasil analisis CFA menunjukkan nilai *Average Variance Extracted* (AVE) sebesar 0,76 dan *Composite Reliability* (CR) sebesar 0,96. Nilai AVE yang melebihi 0,50 menunjukkan bahwa indikator mampu menjelaskan sebagian besar varians konstruk yang diukur, sedangkan nilai CR yang melebihi 0,70 menunjukkan konsistensi internal konstruk yang sangat baik. Dengan demikian, instrumen *hate speech* memiliki validitas konvergen, validitas diskriminan, dan reliabilitas konstruk yang memadai sehingga layak digunakan sebagai alat ukur dalam penelitian.

Pembahasan

Penelitian ini berhasil mengembangkan alat ukur *hate speech* yang terdiri atas tiga dimensi utama, yaitu *Targeted Against Identifiable Group*, *Stigmatization*, dan *Legitimizes Hostility* yang mengacu pada teori Bhikhu Parekh (2012). Ketiga dimensi tersebut kemudian dioperasionalkan menjadi tujuh item pernyataan yang dirancang untuk mengukur kecenderungan individu dalam melakukan ujaran kebencian.

Hasil pengujian menunjukkan bahwa seluruh item memiliki validitas isi yang baik dengan nilai *Aiken's V* di atas 0,70. Temuan ini menunjukkan bahwa setiap item telah dinilai relevan oleh panel ahli dan mampu merepresentasikan indikator yang diukur. Selain itu, seluruh item memiliki nilai *corrected item-total correlation* di atas 0,30 sehingga dapat membedakan responden yang memiliki tingkat hate speech tinggi dan rendah.

Dari aspek reliabilitas, instrumen memperoleh nilai *Cronbach's Alpha* sebesar 0,945 yang menunjukkan tingkat konsistensi internal yang sangat tinggi. Hasil ini mengindikasikan bahwa setiap item dalam alat ukur bekerja secara konsisten dalam mengukur konstruk hate speech.

Pengujian menggunakan *Confirmatory Factor Analysis* (CFA) juga menunjukkan bahwa struktur tiga dimensi yang diajukan memiliki kesesuaian yang baik dengan data empiris. Seluruh dimensi memiliki *loading factor* yang tinggi dan seluruh indeks *goodness of fit* memenuhi kriteria yang dipersyaratkan. Temuan ini memperkuat bahwa model pengukuran yang dikembangkan telah mampu merepresentasikan konstruk hate speech secara memadai.

Selanjutnya, nilai *Composite Reliability* (CR) sebesar 0,96 dan *Variance Extracted* (VE) sebesar 0,76 menunjukkan bahwa instrumen memiliki reliabilitas konstruk dan validitas konvergen yang baik. Dengan demikian, alat ukur yang dikembangkan tidak hanya konsisten, tetapi juga mampu menjelaskan varians konstruk h*ate speech* secara memadai.

Secara keseluruhan, hasil pengujian validitas dan reliabilitas menunjukkan bahwa alat ukur hate speech yang dikembangkan telah memenuhi standar psikometrik yang baik. Oleh karena itu, instrumen ini layak digunakan sebagai alat ukur untuk mengidentifikasi kecenderungan hate speech pada mahasiswa serta dapat menjadi dasar bagi penelitian lanjutan terkait perilaku komunikasi dan interaksi sosial di lingkungan digital

D. Kesimpulan

Berdasarkan tujuan penelitian yang telah ditetapkan, yaitu mendeskripsikan dan menganalisis fenomena ujaran kebencian di kalangan mahasiswa UIN Sunan Gunung Djati Bandung dengan menggunakan kerangka teoritis Bhikhu Parekh (2012), penelitian ini berhasil mengkonstruksi instrumen pengukuran *hate speech* yang valid, reliabel, dan fit secara empiris. Melalui tiga tahap pengembangan alat ukur (validitas substantif, struktural, dan eksternal) terhadap 721 responden, seluruh item pernyataan dinyatakan valid berdasarkan uji validitas isi

Aiken's (nilai V > 0,7) serta memiliki daya beda yang baik (nilai corrected item-total correlation > 0,3).

Hasil uji reliabilitas menunjukkan koefisien Cronbach's Alpha sebesar 0,945, yang mengindikasikan konsistensi internal yang sangat tinggi. Analisis *Confirmatory Factor Analysis* (CFA) mengonfirmasi bahwa struktur tiga dimensi yang diusulkan Parekh *Targeted Against Identifiable Group*, *Stigmatization*, dan *Legitimizes Hostility* mampu merepresentasikan konstruk *hate speech* dengan baik, ditunjukkan oleh nilai *loading factor* masing-masing dimensi di atas 0,94 serta indeks kecocokan model (RMSEA = 0,073; NFI, NNFI, CFI, IFI ≥ 0,99) yang memenuhi kriteria *fit*. Selain itu, instrumen ini memiliki validitas konvergen yang baik terhadap *cyber bullying* (r = 0,748) dan validitas diskriminan yang memadai terhadap etika digital, integritas moral, serta kecerdasan moral (korelasi negatif signifikan).

Dengan demikian, kesimpulan dari penelitian ini adalah bahwa instrumen pengukuran *hate speech* yang dikembangkan terbukti secara ilmiah layak dan andal untuk digunakan dalam mengukur kecenderungan ujaran kebencian di kalangan mahasiswa, khususnya di lingkungan perguruan tinggi Islam Indonesia. Temuan ini sekaligus menjawab kesenjangan literatur terkait belum adanya pemetaan empiris berbasis indikator terstandar pada populasi mahasiswa UIN Sunan Gunung Djati Bandung.

E. Referensi

Abubakar, I., Muchtadlirin, Simun, J., & Nurhidayat, M. (2016). Instrumen Monitoring Hate Speech Tingkat Kabupaten/Kota di Indonesia. Jakarta: Center for the Study of Religion and Culture (CSRC) UIN Syarif Hidayatullah Jakarta & The Asia Foundation.

Alkomah, F., & Ma, X. (2022). A literature review of textual hate speech detection methods and datasets. *Information*, *13*(6), 273. https://doi.org/10.3390/info13060273

Gennaro, G., Bronner, L., Derksen, L., Kubli, M., Kotarcic, A., Kurer, S., Grech, P., Donnay, K., Gilardi, F., & Hangartner, D. (2025). The distribution of hate speech and its implications for content moderation. *Political Science Research and Methods*, 1–9. https://doi.org/10.1017/psrm.2025.10063

Loevinger, J. (1957). Objective tests as instruments of psychological theory. Psychological reports, 3(3), 635-694.

Papcunová, J., Martončik, M., Fedáková, D., Kentoš, M., Bozogáňová, M., Srba, I., Moro, R., Pikuliak, M., Šimko, M., & Adamkovič, M. (2023). Hate speech operationalization: A preliminary examination of hate speech indicators and their structure. *Complex & Intelligent Systems*, *9*, 2827–2842. https://doi.org/10.1007/s40747-021-00561-0

Parekh, B. (2012). Is there a case for banning hate speech? In M. Herz & P. Molnar (Eds.), *The content and context of hate speech: Rethinking regulation and responses* (pp. 37–56). Cambridge University Press. https://doi.org/10.1017/CBO9781139042871.006

Pratama, C. H., & Findawati, Y. (2024). Klasifikasi hate speech dan emosi dalam teks berbahasa Indonesia pada pengguna Twitter menggunakan metode Naïve Bayes classifier. *Indonesian Journal of Applied Technology, 1*(3), 1–10. https://doi.org/10.47134/ijat.v1i3.3105

Sachdeva, P., Barreto, R., Bacon, G., Sahn, A., von Vacano, C., & Kennedy, C. (2022). The measuring hate speech corpus: Leveraging Rasch measurement theory for data perspectivism. In Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022 (pp. 83–94). European Language Resources Association. https://aclanthology.org/2022.nlperspectives-1.11/

Safitri, S. N., & Santoso, I. (2026). Self-control limitations in mitigating hate speech: Role toxic online disinhibition among Generation Z on social media. *Jurnal RAP (Riset Aktual Psikologi), 17*(1), 1–18. https://doi.org/10.24036/rap.v17i1.85

Tontodimamma, A., Nissi, E., Sarra, A., & Fontanella, L. (2021). Thirty years of research into hate speech: Topics of interest and their evolution. *Scientometrics*, *126*, 157–179. https://doi.org/10.1007/s11192-020-03737-6

Toussaint, L. L., Barry, M., Enomoto, M., Anians, W., Rodamaker, K., Keil, A., & Meier, M. (2020). Hateful Emotional Responses Scale (HatERS): Development and initial evaluation. *Journal of Hate Studies, 16*(1), 48–54. https://doi.org/10.33972/jhs.155

Vergani, M., Diallo, T., & O'Brien, K. (2023). Measuring the potential for hateful behaviours: Development and validation of the Hate Behaviours Scale (HBS). Terrorism and Political Violence, 37(2), 1–18. https://doi.org/10.1080/09546553.2023.2283565