한국심리학회지 : 문화 및 사회문제
Korean Journal of Culture and Social Issues
2026, Vol. 32, No. 3, 387~409.
http://dx.doi.org/10.20406/kjcs.2026.8.32.3.387

OPEN ACCESS

멸칭 혐오표현 사용 및 허용 판단에 대한 사회인지적 특성*

양 현 보†

멸칭 혐오표현은 오늘날 사회적 배제와 적대적 상호작용을 강화하는 문제적 언어 현상으로 주목받고 있다. 온라인 공간에서 이러한 표현이 확산될수록 멸칭은 일탈적 발언이 아닌 현실에서 통용될 수 있는 표현이자 사회적으로 용인될 수 있는 발언으로 자리 잡을 수 있다. 이에 본 연구는 멸칭이 어떤 사람들에게 더 쉽게 사용 가능한 표현으로 남고, 어떤 사람들에게 더 사회적으로 허용 가능한 표현으로 판단되는지를 살펴보았다. 이를 위해 대학생 190명이 한국어 멸칭 및 부정단어에 대해 주관적 사용 빈도, 사회적 허용 수준, 정서가, 지각된 공격성 등 14개 심리언어 차원과 도덕기반, 온라인 혐오 수용성, 폭력 수용성, 공감, 감정표현불능 수준 척도에 응답하였다. 선형 혼합효과 모형 분석 결과, 멸칭은 부정단어에 비해 주관적 사용 빈도와 사회적 허용 수준이 낮고, 보다 공격적인 표현으로 지각되었다. 또한 온라인 혐오 수용성은 멸칭의 주관적 사용 빈도와 선택적으로 관련되었으며, 사회적 허용 수준은 폭력 수용성 및 도덕기반-충성성과 정적으로, 공감 및 도덕기반-돌봄·공정성과 부적으로 관련되었다. 이와 함께, 매개분석에서는 공감과 멸칭 허용 수준의 관계에서 정서가와 지각된 공격성을 경유하는 간접 경로가, 공정성·충성성과 멸칭 허용 수준의 관계에서는 정서가를 경유하는 간접 경로가 관찰되었다. 이러한 결과는 멸칭에 대한 사용 및 허용 경향이 단일한 특정 성향으로 환원되지 않으며, 멸칭 혐오표현의 확산 기제를 이해하고 적절히 대응하기 위해서는 개인의 사회인지적 특성이 면밀히 고려될 필요가 있음을 시사한다.

주제어: 멸칭, 혐오표현, 주관적 사용 빈도, 사회적 허용, 도덕기반

* 이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2023S1A5B5A17089068).

† 교신저자 : 양현보, 한국과학기술원 뇌인지과학과, 대전광역시 유성구 대학로 291
Email: hyeonbo@kaist.ac.kr

Copyright ©2026, The Korean Psychological Association of Culture and Social Issues
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non- commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

혐오표현이란 특정 사회집단을 비하하고 배제하며, 그 집단에 대한 차별과 공격을 정당화하는 언어 행위이다(Bilewicz & Soral, 2020; Cervone et al., 2021). 그 가운데 멸칭(derogatory labels)은 특정 집단을 하나의 열등한 범주로 묶고 경멸적 의미를 부여하는 명사형 혐오표현에 해당한다(Carnaghi & Maass, 2007; Fasoli et al., 2015, 2016). 멸칭은 단순히 부정적인 감정을 전달하는 단어가 아니라, 화자가 누구를 사회적 경계 안의 구성원으로 인정하고 누구를 경계 밖으로 밀어내는지를 드러내는 언어적 표현이다. 따라서 멸칭의 사회적 사용 및 확산 기제를 이해하려면 표현 자체의 언어적 특성뿐 아니라, 그러한 표현을 사용하고 사회적으로 허용하는 사람들의 심리적 특성을 함께 살펴볼 필요가 있다.

멸칭은 일반적인 부정단어와 구분된다. 비멸칭 부정단어 역시 비난과 평가를 담을 수 있지만, 멸칭은 표적을 특정 사회집단의 구성원으로 범주화하고 그 집단에 대한 경멸적 평가를 함께 전달한다는 특수성을 지닌다(Carnaghi & Maass, 2007; Croom, 2014). 예컨대 ‘배신자’, ‘위선자’, ‘도둑놈’ 등의 부정단어가 주로 개인의 행위, 성격, 일탈, 범죄를 비난하는 표현이라면, ‘김치녀’, ‘틀딱’, ‘한남충’ 등의 멸칭은 사람을 특정 범주에 포함시킨 뒤 그 범주 자체를 비하하는 표현에 해당한다(멸칭들의 예시는 부록 B 조사단어 목록 참조). 이러한 차이는 멸칭을 단순히 강한 부정어의 한 종류로 환원하기 어렵게 만들며, 멸칭의 사용과 허용을 설명하는 개인차 구조 역시 일반적인 부정단어에 대한 양상과 상이할 가능성을 시사한다.

한국어 맥락에서 양현보(2025)는 이 문제를 단어 범주를 나누어 구체적으로 다루었다. 해당 연구에서 멸칭은 부정단어와 비교해 정서가(valence), 각성가(arousal), 지각된 공격성, 범주 본질주의 등에서 서로 다른 심리언어적 특성을 나타냈다. 즉, 멸칭은 자극 수준에서 단순 부정단어와 상이한 범주로 기능하였으며, 청자의 해석 방식에서도 구분된 양상을 보였다. 그러나 선행연구는 멸칭 혐오표현이 갖는 심리언어적 특징에 초점을 맞추었을 뿐, 어떤 사회인지적 특성이 그러한 표현을 더 자주 사용하고, 더 쉽게 허용하게 만드는지에 대해서는 충분한 설명을 내놓지 못하였다.

이러한 공백은 멸칭 연구의 분석 초점을 자극 특성에서 사용자 개인차로 확장할 필요를 제기한다. 동일한 멸칭이라도 누군가에게는 장난스럽게 사용 가능한 표현으로 인식되는 반면, 다른 사람에게는 사회적으로 허용되기 어려운 표현으로 인식될 수 있다(박세훈 등, 2025). 이때 멸칭에 대한 주관적 사용 빈도와 사회적 허용 수준은 서로 밀접하게 연결되어 있지만, 구분된 차원에서 이해될 필요가 있다. 주관적 사용 빈도는 실제로 본인이 주어진 표현을 얼마나 자주 사용하는지에 대한 보고인 반면, 사회적 허용 수준은 그 표현이 사회적으로 얼마나 용인될 것인지에 대한 판단에 해당하기 때문이다.

특히 사회적 허용 수준은 단순한 사용 가능성의 문제가 아니라, 특정 표현을 규범적으로 받아들일 수 있는가에 대한 판단과 관련된다. 이러한 규범 판단은 개인이 어떤 도덕적 가치를 더 중요하게 여기는가와 연결될 수 있다. 사회 심리학자 Jonathan Haidt가 제안하고 Graham 등(2011)이 체계화한 도덕기반이론(moral foundations theory)에 따르면, 사람들은 돌봄, 공정성, 충성성, 권위, 순수성과 같은 여러 도덕기반에 서로 다른 가치를 부여하며 사

회적 판단을 내린다. 이 가운데 돌봄과 공정성의 경우 타인의 고통과 부당한 대우에 대한 민감성과 연결된다는 점에서 멸칭의 사용 빈도와 사회적 허용 수준을 낮추는 방향으로 작동할 가능성이 있다. 반면 충성성과 같은 결속적 도덕기반은 집단 경계의 유지와 내집단 결속을 강조하므로(Graham et al., 2011; Kim et al., 2012; Kennedy et al., 2023) 멸칭과 같이 외집단을 배제하거나 비하하는 표현에 대해 상대적으로 관대한 태도와 연결될 수 있다. 이러한 관점에서 멸칭의 사용과 허용은 단순한 언어 습관의 문제를 넘어 개인이 중시하는 도덕기반에 기초한 문제로 이해할 수 있다.

여기에 공감과 공격 정당화에 관한 태도를 함께 고려하면, 멸칭의 사용과 허용을 설명하는 보다 직접적인 사회인지적 경로를 가정할 수 있다. 공감은 타인의 고통과 모욕에 더 민감하게 반응하도록 만드는 특성이라는 점에서(Decety & Yoder, 2016), 공감 수준이 높은 개인일수록 멸칭을 더 공격적이고 덜 허용 가능한 표현으로 받아들일 가능성이 있다. 반대로 폭력 수용성과 온라인 혐오 수용성은 서로 밀접하게 관련되므로(Celuch et al., 2022), 이러한 공격 및 혐오 허용 태도가 높은 개인은 멸칭을 덜 문제적인 표현으로 받아들일 가능성이 있다. 따라서 멸칭에 대한 판단은 해당 표현을 공격적이고 부정적인 것으로 해석하는 정도에 의해 매개될 것이라 추측할 수 있다.

이와 함께, 멸칭은 사회적 상호작용 가운데 사용되고 공유되는 표현이므로, 개인의 도덕 직관이나 공감만으로 설명되지 않는 사회적 맥락 역시 함께 고려될 필요가 있다. 정치성향, 인터넷 사용 시간, 온라인 혐오표현 노출 정도, 그리고 자신의 도덕적 가치관이 타인과 얼마나 유사하다고 생각하는지 등은 멸칭이 어떤 사회적 맥락에서 쉽게 사용되는지를 드러내는 배경 지표로서 기능할 수 있다. 예컨대, 반복적인 노출과 규범적 허용은 혐오표현을 일상화하거나 규범적 저항을 약화시키는 방향으로 작동할 수 있으며(Soral et al., 2018; Pluta et al., 2023), 반대로 자신과 타인 사이에 도덕적 가치관이 공유된다는 인식은 내집단 수용성을 높여 특정 집단에 대한 혐오 반응을 억제하는 기제로서 작동할 수 있다(Čehajić-Clancy et al., 2023; Wolf et al., 2021). 이는 멸칭의 사용과 허용이 개인의 내재적 특성뿐 아니라, 온라인 활동 수준 및 혐오 노출 경험, 도덕적 가치관 등 구체적인 사회적 배경 속에서 함께 이해될 필요가 있음을 시사한다.

종합하면, 멸칭 혐오표현의 사용 및 허용 수준은 단순히 표현 자체의 부정성만으로 설명되기보다 도덕기반, 공감, 온라인 혐오에 대한 태도, 폭력 수용성, 표현 해석 방식을 통해 보다 구체적으로 설명될 수 있다. 그러나 선행연구는 멸칭 혐오표현의 심리언어적 특성에 초점을 맞추었으므로, 어떤 사회인지적 특성이 멸칭의 사용과 허용에 관련되는지에 대한 검토는 불충분하다. 이에 본 연구는 멸칭과 비멸칭 부정단어를 비교하여 온라인 혐오 수용성, 공감, 도덕기반과 같은 사회인지적 특성이 멸칭의 주관적 사용 빈도와 사회적 허용 수준에 어떠한 관련성을 갖는지를 폭넓게 검토하였다. 이와 더불어, 멸칭에 대한 정서적 인식과 지각된 공격성이 이러한 관계에서 어떤 역할을 하는지를 함께 살펴보았다.

방 법

참가자

부산 소재 대학교 학부생 196명이 설문에 참여하였다. 자료 선별 과정에서 6명의 참가자가 척도 응답 과정에서 장기간 동일한 값으로 연속 응답하는 패턴을 보여 분석에서 제외되었다. 최종적으로 190명의 자료가 분석에 사용되었으며, 평균연령은 20.75세(SD = 2.05), 성별 분포는 여성 126명, 남성 64명이었다. 참가자들은 설문 전 연구의 목적과 절차에 대한 설명을 제시받고 자발적으로 설문에 참여하였으며, 참가 보상으로 수업 점수를 받았다.

조사 자극 및 평정 도구

평정에 사용된 단어 자극은 한국어 멸칭의 심리언어적 특성을 조사한 선행연구(양현보, 2025)에 수록된 멸칭 20개와 부정단어 20개였다. 이때 자극 범주의 타당성을 단일 선행연구의 분류에만 의존하지 않도록, 국내외 멸칭·혐오표현 문헌에서 제시된 개념적 기준과의 정합성을 함께 검토하였다. 기존 연구에서 멸칭은 특정 사회집단이나 정체성 범주를 지시하면서 해당 범주와 그 구성원에게 경멸적 평가를 부여하는 집단 명칭으로 정의되었으며(Carnaghi & Maass, 2007; Cervone et al., 2021; Croom, 2014), 국내 연구에서도 혐오표현 및 비하 신조어는 특정 집단을 범주화하고 그 집단 전체를 부정적으로 평가하거나 사회적 배제의 대상으로 만드는 표현으로 다루어져 왔다(강희숙, 2018; 장경현, 2021; 홍성수, 2019). 본 연구의 멸칭 자극은 성별, 세대, 인종·국적, 정치적 정체성 또는 기타 사회적 범주를 지시하고 해당 범주 전체에 부정적 평가를 부여하는 단어라는 점에서 이러한 기존 문헌들의 정의와 개념적으로 부합한다.

반면 비멸칭 부정단어는 개인의 행동, 성격, 범죄, 일탈 또는 사회적 상태를 부정적으로 지칭하되, 특정 사회집단이나 정체성 범주 전체에 대한 관습적 비하를 핵심 기능으로 하지 않는 표현으로 구분하였다. 따라서 본 연구의 자극 구분은 양현보(2025)의 분류를 바탕으로, 집단 지시성, 범주 기반 일반화, 경멸·배제 기능이라는 선행 문헌의 공통 기준을 따른다. 전체 단어 목록은 부록에 수록되었다.

참가자는 40개 단어 각각에 대해 의미 이해, 정서가, 각성가, 추상성, 정서 경험, 친숙성, 사용 빈도, 심상성, 지각된 공격성, 사회적 허용 수준, 자기와의 유사성, 성별 지시성, 범주 실체성, 범주 본질주의의 14개 차원을 1-9점 척도로 평정하였다. 이 가운데 의미 이해, 정서가, 각성가, 추상성, 친숙성, 사용 빈도, 심상성은 선행연구(양현보, 2025)와 동일 문항을 포함하였으며, 지각된 공격성과 사회적 허용 수준은 멸칭의 규범적 해악과 허용 판단을 파악하기 위한 차원으로 포함하였다. 반면 정서 경험, 자기와의 유사성, 성별 지시성, 범주 실체성, 범주 본질주의는 멸칭이 개인적 연상과 사회적 범주 표상에 어떻게 연결되는지를 보다 넓게 파악하기 위한 탐색적 차원으로 포함하였다. 전체 설문 문항은 부록에 수록되었다.

단어 평정이 끝난 뒤에는 정치성향, 하루 인터넷 사용 시간, 타인과의 도덕적 유사성 지각, 온라인에서 특정 집단에 대한 부정적 시선 노출을 각각 단일 문항으로 측정하고, 온라인 혐오 수용성(OHA) 5문항, 폭력 수용성(AOV) 15문항, 한국판 토론토 공감 질문지(TEQ) 16문항, 한국판 Perth 감정표현불능증 척도(K-PAQ) 24문항, 도덕기반(K-MFQ) 32문항을 순차적으로 평정하였다.

온라인 혐오 수용성(OHA)과 폭력 수용성(AOV)은 Celuch 등(2022)의 국제 비교연구에서 사용된 문항을 한국어로 직접 번안하여 사용하였다. 번안 과정에서는 원문의 구성 개념과 문항 의미가 한국어 문장에서도 유지되는지, 그리고 응답자가 자연스럽게 이해할 수 있는 표현인지를 4인의 동료 검토를 통해 확인하였다. 다만 이는 엄밀한 한국어 타당화 절차가 아니라는 점을 고려하여, 두 척도와 관련된 결과는 탐색적 증거로 해석하였다. OHA는 온라인 비하·혐오 메시지의 허용 조건을 묻는 5문항으로, AOV는 법 위반, 재산손괴, 보복 폭력, 테러 허용 등을 포함한 15문항으로 각각 구성되었다. 한국판 토론토 공감 질문지(TEQ)는 김환과 한수미(2016)가 타당화한 한국판 토론토 공감 질문지 16문항을 사용하였다. 감정표현불능 수준은 조예진(2024)이 타당화한 한국판 Perth 감정표현불능증 척도(K-PAQ) 24문항을 사용하였다. 도덕기반 문항은 Graham 등(2011)의 Moral Foundations Questionnaire (MFQ)를 번안한 Kim 등(2012)과 석자춘(2022)의 한국어판 문항 구성을 토대로 사용하였다. 단어 평정, 번안 문항 및 응답 방식은 부록에 수록하였다.

분석 방법

혼합모형 분석의 목적은 참가자들이 멸칭과 부정단어를 평정한 자료에서, 참가자의 사회인지적 특성이 멸칭 사용 및 허용 수준에 선택적으로 결합하는지를 살펴보는 것이었다. 이를 위해 Python의 statsmodels.formula.api.mixedlm을 사용하여 선형 혼합효과 모형(linear mixed-effects model, LMEM) 분석을 수행하였다(Baayen et al., 2008; Barr et al., 2013). 모형은 '종속변인 ~ 단어 범주 * (OHA + AOV + TEQ + 돌봄 + 공정성 + 충성성) + 연령 + 성별 + (1|참가자) + (1|단어)'에 해당하는 구조였고, 종속변인은 주관적 사용 빈도와 사회적 허용 수준으로 나누어 각각 추정하였다. 이때 '(1|참가자)'는 참가자별 무선 절편을, '(1|단어)'는 단어별 무선 절편에 해당하는 분산 성분을 뜻한다. 혼합모형에는 핵심 예측변인에 결측이 없는 190명의 7,600개 평정값이 투입되었고, 추정은 최대우도 방식으로 수행하였다. 분석에 투입된 연속형 예측변인은 모두 표준화하였다. 핵심 예측변인과 공변량 사이의 다중공선성은 참가자 수준 설계 행렬의 분산팽창지수(VIF)로 점검하였다. 그 결과, 모든 변인의 VIF는 2.5보다 낮았으며 가장 큰 값은 도덕기반-돌봄의 2.16이었다. 따라서 공감, 돌봄, 공정성과 같이 개념적으로 연관된 변인이 함께 투입되었더라도 혼합모형의 핵심 계수를 주요하게 왜곡할 수준의 다중공선성은 확인되지 않았다.

혼합모형의 핵심 예측변인으로는 온라인 혐오 수용성(OHA), 폭력 수용성(AOV), 공감(TEQ), 도덕기반-돌봄, 도덕기반-공정성, 도덕기반-충성성을 포함하였다. 이 변인들은 각각 멸칭의 규범적 허용, 공격의 정당화, 타인의 고통에 대한 민감성, 그리고 해악·공정성·내집단 충성에 관한 도덕 판단이라는 점에서 연구 질문과 직접 맞닿아 있었다. 반면 연령과 성별은 전반적인 언어 사용 성향 및 규범 판단에 영향을 줄 수 있는 기본 인구통계학적 차이로 보아 공변량으로 통제하였다.

단어 범주 × 변인 상호작용은 기본 배경 차이를 넘어 멸칭에 선택적으로 결합하는 효과로 해석하였다. 즉, 혼합모형의 상호작용 계수는 핵심 예측변인을 동시에 고려했을 때 비

멸칭 부정단어를 기준으로 멸칭 범주의 기울기가 얼마나 달라지는지를 가리킨다. 정적 상호작용은 해당 예측변인과 종속변인의 관련성이 비멸칭 부정단어보다 멸칭에서 더 강한 방향으로 나타났음을 의미하고, 부적 상호작용은 그 관련성이 멸칭에서 상대적으로 약하거나 더 낮은 방향으로 나타났음을 의미한다. 따라서 상호작용 계수는 특정 변인이 멸칭의 사용 또는 허용 수준을 단독으로 증가·감소시킴을 의미하는 것이 아니라, 두 단어 범주에서 예측변인의 기울기에 차이가 있음을 나타낸다.

감정표현불능 수준(PAQ), 도덕기반-권위·순수성, 정치성향, 인터넷 사용 시간, 지각된 도덕적 일치, 온라인 혐오 노출은 보조 변인으로 다루었다. 이 변인들은 멸칭 사용과 허용의 사회적·정서적 배경을 보여 줄 수 있으나, 참가자 수준 예측변인을 모두 혼합모형에 포함할 경우 고정효과 간 상호 의존성이 커지고, 범주와 예측변인의 상호작용을 해석하기 어려워질 수 있다(Babyak, 2004; Dormann et al., 2013). 이에 혼합모형에는 연구 목적에 우선적으로 부합하는 변인을 중심적으로 포함하였으며, 보조 변인들은 상관분석을 통해 멸칭의 주관적 사용 빈도 및 사회적 허용 수준이 놓인 규범적·사회적 맥락을 해석하는 근거로 활용하였다.

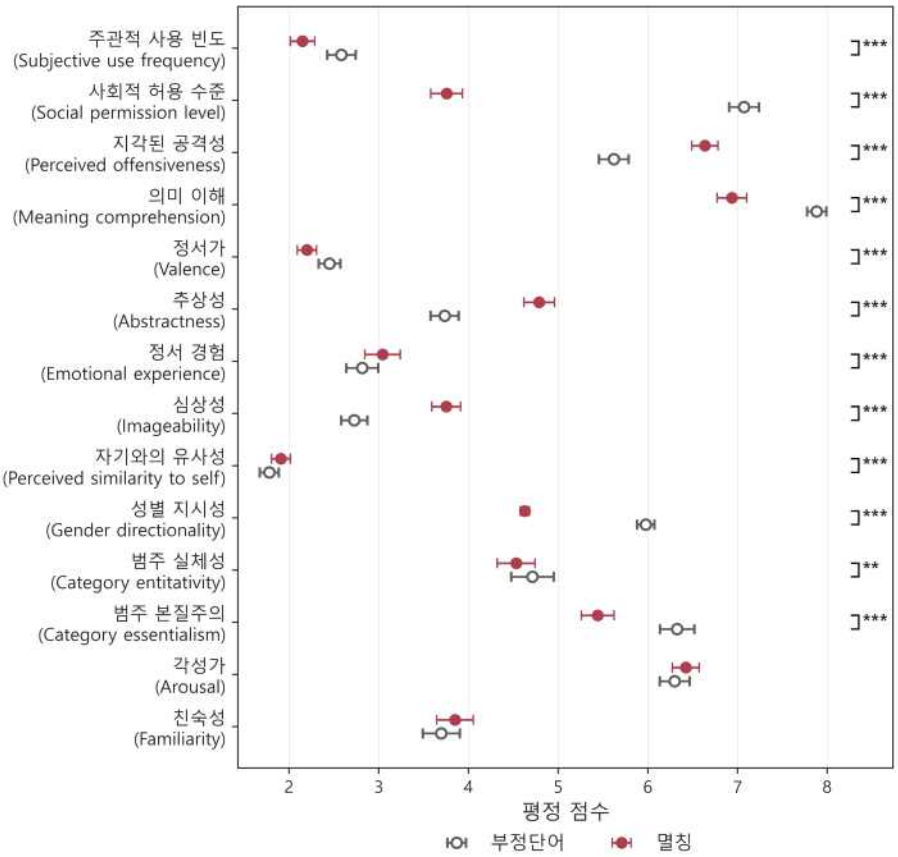
단어 평정에 사용된 14개 심리언어 차원은 참가자의 고정된 성향이 아니라 단어 범주에 따른 인식 차이를 살펴보기 위한 지표였다. 따라서 이 차원들은 멸칭 사용 및 허용에 대한 참가자 수준 예측변인으로 혼합모형에 투입하기보다는 단어 범주별 평균 비교, 상관구조, 사회인지적 변인과의 관계를 매개할 가능성을 확인하는 데 사용하였다.

끝으로, 각 변인이 정서가 또는 지각된 공격성과 같은 심리언어적 변인을 경유하여 멸칭 사용 빈도 및 사회적 허용 수준에 영향을 미치는 간접 경로를 이루는지 살펴보기 위하여 매개분석을 실시하였다. 이 분석은 pyprocessmacro 라이브러리를 사용하여 각 참가자의 평균 자료를 바탕으로 단어 범주별 모형을 추정하였다. 매개모형에는 연령, 성별, 정치성향, 인터넷 사용 시간, 지각된 도덕적 일치, 온라인 혐오 노출을 공변량으로 포함하였다. 해당 변수들은 멸칭의 사용과 허용에 영향을 줄 수 있는 변인이지만, 매개모형에서는 이를 통제함으로써 초점 사회인지 변인과 사회적 허용 수준의 관계가 정서가 및 공격성과 같은 심리언어적 해석 변인을 경유하여 나타나는지를 보다 엄밀하게 확인하고자 하였다. 간접효과의 신뢰구간은 1,000회 부트스트랩을 통해 산출하였다.

결 과

멸칭과 부정단어의 심리언어적 구분

멸칭과 부정단어는 주요 심리언어 차원에서 서로 다른 평정 양상을 보였다. 주관적 사용 빈도는 멸칭(M = 2.15)이 부정단어(M = 2.59)보다 유의하게 낮았다, $t(189) = -6.78$, $p < .001$. 사회적 허용 수준 역시 멸칭(M = 3.76)이 부정단어(M = 7.07)보다 낮았다, $t(189) = -33.62$, $p < .001$. 반면, 지각된 공격성은 멸칭(M = 6.64)이 부정단어(M = 5.62)보다 높았다, $t(189) = 11.54$, $p < .001$. 그 밖의 차원에서 관찰된 단어 유형 간 차이는 그림 1에 제시하였다.



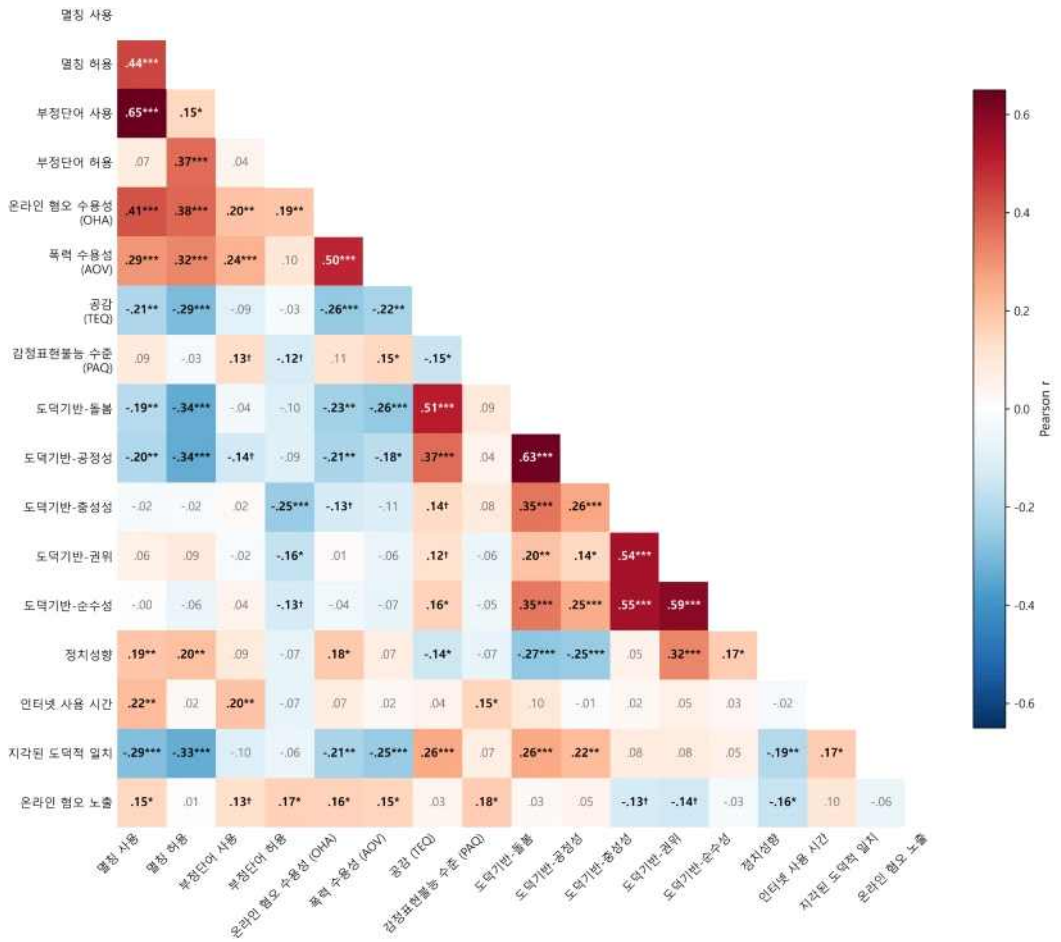
***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; † : $p \leq 0.08$.

그림 1. 14개 심리언어 차원에서 멸칭과 부정단어의 평균 비교. 오차막대는 각 범주 평균의 95% 신뢰구간을 나타내며, 우측의 유의성 표기는 참가자 수준 범주 평균에 대한 대응표본 t 검정 결과를 나타낸다.

사용 · 허용 지표와 사회인지적 변인 간 관계

그림 2는 상관분석 결과 멸칭의 주관적 사용 빈도와 사회적 허용 수준이 서로 다른 특성과 관련되는 양상을 나타낸다. 온라인 혐오 수용성(OHA)은 멸칭 사용 빈도와 정적 상관을 보였으며, $r = .413$, $p < .001$, 멸칭 허용 수준과도 정적 상관을 나타냈다, $r = .376$, $p < .001$. 폭력 수용성(AOV)은 멸칭 사용 빈도와 정적 상관을 보였으며, $r = .290$, $p < .001$, 멸칭 허용 수준과도 정적 상관을 나타냈다, $r = .325$, $p < .001$. 이와 대조적으로, 공감(TEQ)은 멸칭 허용 수준과 부적 상관을 보였으며, $r = -.295$, $p < .001$, 도덕기반-공정성 또한 멸칭 허용 수준과 부적 상관을 나타냈다, $r = -.337$, $p < .001$.

보조 변인 가운데 지각된 도덕적 일치는 멸칭 사용 빈도와 부적 상관을 보인 반면, $r = -.288$, $p < .001$, 정치성향, 인터넷 사용 시간, 온라인 혐오 노출은 멸칭 사용 빈도와 정적 상관을 보였다, $r = .189$, $p = .009$; $r = .218$, $p = .002$; $r = .148$, $p = .041$.



***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; † : $p \leq 0.08$.

그림 2. 단어 범주별 사용/허용 응답 및 사회인지적 변인 간 상관 히트맵. 볼드체로 강조된 숫자는 유의한 상관관계를 나타낸다.

주관적 사용 빈도에 대한 혼합모형 결과

사용 빈도에 대한 선형 혼합효과 모형 분석 결과, 멸칭의 범주 주효과는 유의하게 부적인 방향으로 나타났다, b = -0.433, SE = 0.041, z = -10.64, p < .001. 이는 참가자들이 부정단어보다 멸칭을 덜 사용한다고 보고했음을 뜻한다. 단어 범주와 참가자 특성의 상호작용에서는 온라인 혐오 수용성이 유의한 정적 효과를 나타냈다, b = 0.211, SE = 0.048, z = 4.39, p < .001. 즉, 온라인 혐오 수용성이 높을수록 부정단어보다 멸칭의 사용 빈도가 상대적으로 더 높게 평정되었다. 폭력 수용성의 상호작용은 부적 방향으로 유의하였다, b = -0.126, SE = 0.048, z = -2.64, p = .008. 돌봄 역시 부적 방향의 상호작용을 보였다, b = -0.175, SE = 0.059, z = -2.94, p = .003. 공정성의 상호작용은 경향 수준에 머물렀고, b =

0.096, SE = 0.053, z = 1.81, p = .071, 공감과 충성성의 상호작용은 유의하지 않았다, 공감: b = -0.022, SE = 0.048, z = -0.45, p = .650; 충성성: b = 0.016, SE = 0.044, z = 0.36, p = .720.

사회적 허용 수준에 대한 혼합모형 결과

사회적 허용 수준 모형에서도 멸칭의 범주 주효과는 유의한 음의 방향으로 나타났다, b = -3.315, SE = 0.040, z = -83.56, p < .001. 이는 멸칭이 부정단어보다 사회적으로 덜 허용 가능한 표현으로 평정되었음을 뜻한다. 범주와 참가자 특성의 상호작용에서는 폭력 수용성이 정적인 방향으로 유의하였으며, b = 0.167, SE = 0.046, z = 3.60, p < .001, 충성성 또한 정적 상호작용이 관찰되었다, b = 0.437, SE = 0.043, z = 10.27, p < .001.

반면, 공감·돌봄·공정성은 모두 부적인 방향으로의 상호작용을 보였다, b = -0.169, SE = 0.047, z = -3.59, p < .001; b = -0.179, SE = 0.058, z = -3.09, p = .002; b = -0.197, SE = 0.052, z = -3.83, p < .001. 온라인 혐오 수용성의 상호작용은 경향 수준에서 정적인

표 1. 선형 혼합효과 모형 주요 분석 결과 요약

종속변인	효과	b	95% CI	SE	z	p
주관적 사용 빈도	단어 범주 (멸칭 vs. 부정단어)	-0.433	[-0.512, -0.353]	0.041	-10.64	< .001
	멸칭 × OHA	0.211	[0.117, 0.305]	0.048	4.39	< .001
	멸칭 × 폭력 수용성	-0.126	[-0.219, -0.032]	0.048	-2.64	.008
	멸칭 × 공감	-0.022	[-0.116, 0.073]	0.048	-0.45	.650
	멸칭 × 돌봄	-0.175	[-0.291, -0.058]	0.059	-2.94	.003
	멸칭 × 공정성	0.096	[-0.008, 0.199]	0.053	1.81	.071
	멸칭 × 충성성	0.016	[-0.070, 0.101]	0.044	0.36	.720
사회적 허용 수준	단어 범주 (멸칭 vs. 부정단어)	-3.315	[-3.392, -3.237]	0.040	-83.56	< .001
	멸칭 × OHA	0.091	[-0.000, 0.183]	0.047	1.96	.051
	멸칭 × 폭력 수용성	0.167	[0.076, 0.258]	0.046	3.60	< .001
	멸칭 × 공감	-0.169	[-0.261, -0.077]	0.047	-3.59	< .001
	멸칭 × 돌봄	-0.179	[-0.293, -0.065]	0.058	-3.09	.002
	멸칭 × 공정성	-0.197	[-0.298, -0.096]	0.052	-3.83	< .001
	멸칭 × 충성성	0.437	[0.354, 0.521]	0.043	10.27	< .001

주. 전체 선형 혼합효과 모형 분석 결과 중 단어 범주의 주효과 및 주요 사회인지적 변인과의 상호작용을 제시하였다. '멸칭 × 예측변인' 항은 부정단어를 기준으로 멸칭 범주에서 주어진 예측변인의 기울기가 얼마나 추가적으로 변하는지를 나타낸다.

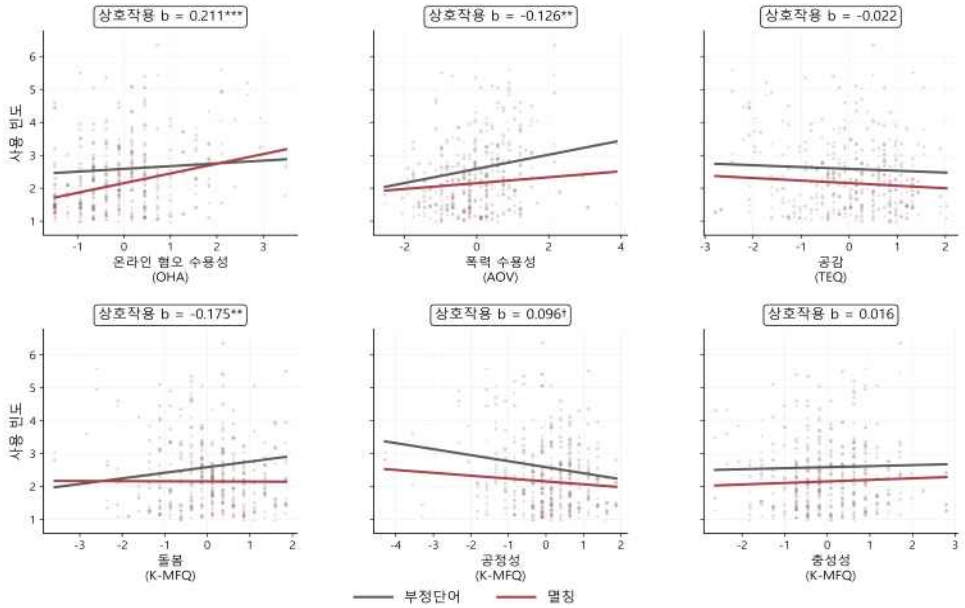


그림 3. 사용 빈도에 대한 범주 × 참가자 특성 상호작용. 각 패널의 점은 참가자별 범주 평균을 나타내며, 선은 혼합모형을 통해 산출된 고정효과 예측선이다. 패널 상단의 값은 부정단어에 비해 멸칭에서의 기울기가 유의하게 달라지는지를 나타내는 상호작용 계수이다.

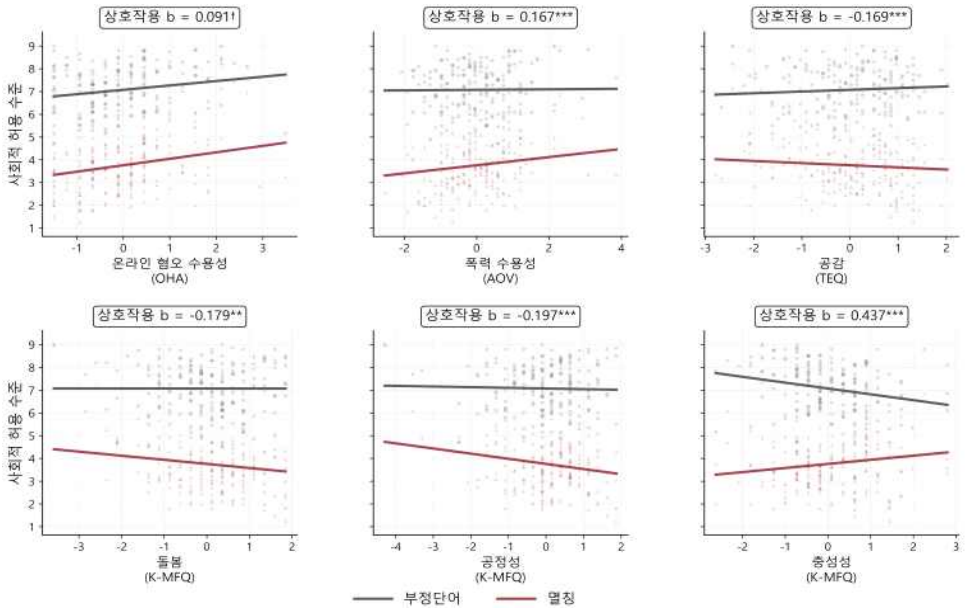


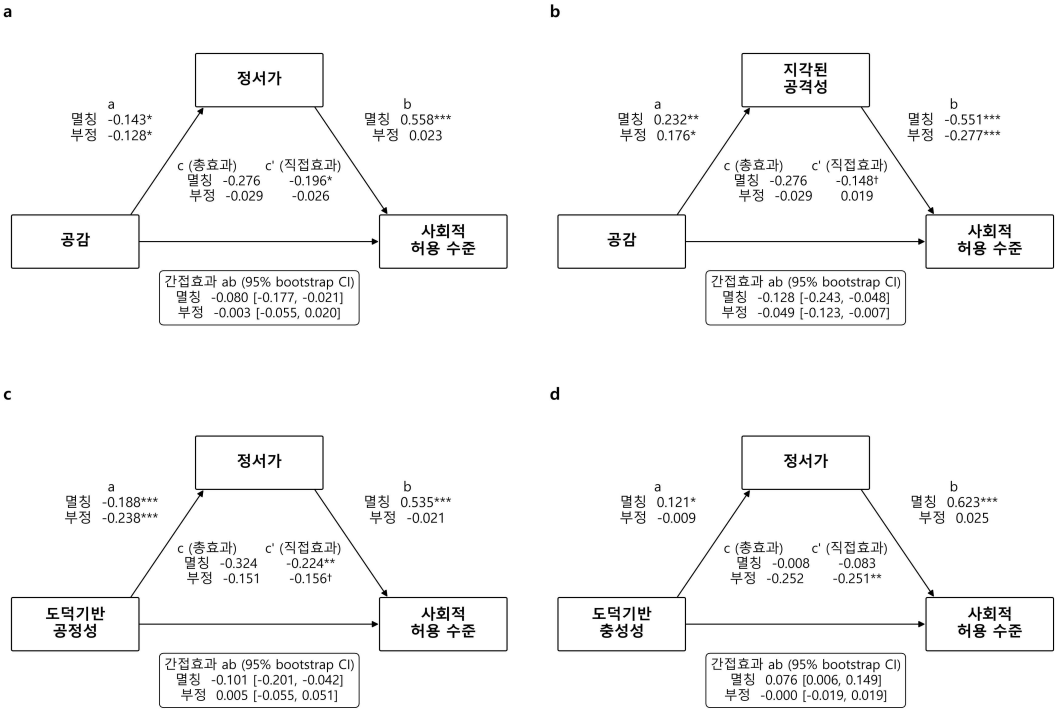
그림 4. 사회적 허용 수준에 대한 범주 × 참가자 특성 상호작용. 각 패널의 점은 참가자별 범주 평균을 나타내며, 선은 핵심 혼합모형을 통해 산출된 고정효과 예측선이다. 패널 상단의 값은 부정단어에 비해 멸칭에서의 기울기가 유의하게 달라지는지를 나타내는 범주 상호작용 계수이다.

방향으로 관찰되었다, b = 0.091, SE = 0.047, z = 1.96, p = .051.

사회적 허용 수준에 대한 매개분석 결과

그림 5는 개인의 사회인지적 성향(공감, 도덕기반-공정성, 도덕기반-충성성)에 따라 멸칭에 관대해지거나 엄격해지는 과정을 단어에 대한 정서적 평가와 지각된 공격성이 매개하는 간접 경로를 나타낸다. 공감과 사회적 허용 수준의 관계에서 정서가를 경유하는 간접효과는 멸칭에 대해 유의하였으나, ab = -0.080, 95% bootstrap CI [-0.177, -0.021], 부정단어에 대해서는 신뢰구간이 0을 포함하여 유의하지 않았다, ab = -0.003, 95% bootstrap CI [-0.055, 0.020](그림 5a). 반면 지각된 공격성을 경유하는 간접효과는 두 범주 모두에서 유의하였다, 부정단어: ab = -0.049, 95% bootstrap CI [-0.123, -0.007]; 멸칭: ab = -0.128, 95% bootstrap CI [-0.243, -0.048](그림 5b).

도덕기반-공정성과 사회적 허용 수준의 관계에서 정서가를 경유하는 간접효과는 멸칭에 대해 유의하였으나, ab = -0.101, 95% bootstrap CI [-0.201, -0.042], 부정단어에 대해서는 유의하지 않았다, ab = 0.005, 95% bootstrap CI [-0.055, 0.051](그림 5c).



***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$.

그림 5. 단어 범주별 사회적 허용 수준 매개분석 모형. 연령, 성별, 정치성향, 인터넷 사용 시간, 지각된 도덕적 일치, 온라인 혐오 노출은 공변량으로 통제되었다. 각 경로 주변의 값들은 멸칭과 비멸칭 부정단어 조건의 비표준화 회귀계수이며, 하단 상자는 1,000회 bootstrap으로 추정한 간접효과(ab)의 95% 신뢰구간을 나타내며, 0이 포함되지 않는 경우 유의한 간접효과로 해석하였다.

도덕기반-충성성과 사회적 허용 수준의 관계에서 정서가를 경유하는 간접효과는 멸칭에 대해 유의하였으나, ab = 0.076, 95% bootstrap CI [0.006, 0.149], 부정단어에 대해서는 유의하지 않았다, ab = -0.000, 95% bootstrap CI [-0.019, 0.019](그림 5d).

주관적 사용 빈도에 대해서도 동일한 매개분석을 실시하였으나, 유의한 간접효과는 일부 비멸칭 부정단어 모형에 제한되었으며, 멸칭 사용 빈도의 경우 정서가 및 지각된 공격성을 경유하는 간접효과가 관찰되지 않았다.

논 의

본 연구는 멸칭 혐오표현이 어떤 사람들에게 더 쉽게 사용되고 사회적으로 허용되는지를 이해하기 위하여 멸칭의 주관적 사용 빈도와 사회적 허용 수준이 어떠한 사회인지적 특성과 관련되는지를 살펴보았다. 먼저, 자극 수준에서 멸칭은 비멸칭 부정단어보다 주관적 사용 빈도와 사회적 허용 수준이 더 낮고, 공격성은 더 높은 표현으로 평정되었다. 이러한 결과는 멸칭이 일반 부정단어와 구분되는 심리언어적 특성을 갖는다는 선행연구의 결과와 부합한다(양현보, 2025). 나아가 본 연구는 이러한 멸칭의 특수성이 자극 수준에 그치지 않고 사용자 수준의 개인차와도 연결됨을 중점적으로 확인하였다. 이때 온라인 혐오 수용성은 멸칭의 주관적 사용 빈도와 선택적으로 관련된 반면, 사회적 허용 수준은 폭력 수용성, 공감, 돌봄, 공정성, 충성성과 더 폭넓게 연결되었다. 이러한 결과는 멸칭 혐오표현의 사용 가능성과 허용 판단이 동일한 개인차 구조로 설명되지 않음을 시사한다.

주관적 사용 빈도에서는 온라인 혐오 수용성이 가장 뚜렷한 예측 요인으로 나타났다. 온라인 혐오표현을 더 허용 가능한 것으로 보는 사람일수록 비멸칭 부정단어에 비해 멸칭을 더 자주 사용한다고 보고하였다. 이는 멸칭의 사용 경향이 일반적인 공격 정당화 성향보다는 온라인 혐오에 대한 허용 태도와 더욱 밀접하게 관련될 수 있음을 보여준다. 반면, 상관분석에서 돌봄 성향은 멸칭 사용 빈도를 낮추는 방향으로 나타났으며, 혼합모형에서도 돌봄과 사용 빈도의 관계는 비멸칭 부정단어보다 멸칭에서 상대적으로 더 부적인 방향으로 나타났다. 이는 타인의 고통과 취약성에 민감한 개인일수록 멸칭 사용에 더 큰 저항을 보일 가능성을 시사한다.

폭력 수용성은 사용 빈도와 사회적 허용 수준에서 서로 다른 양상을 보였다. 상관분석에서는 폭력 수용성이 멸칭 사용 빈도와 정적으로 관련되었지만, 혼합모형에서는 멸칭 × 폭력 수용성 상호작용이 부적으로 나타났다. 이는 폭력 수용성이 높은 사람이 멸칭을 덜 사용한다고 보고했다는 뜻이 아니며, 다른 사회인지적 변인들을 고려했을 때 폭력 수용성과 사용 빈도의 관계가 비멸칭 부정단어보다 멸칭에서 상대적으로 약했음을 의미한다. 반면 사회적 허용 수준에서는 멸칭 × 폭력 수용성 상호작용이 정적으로 나타나, 폭력 수용성이 높은 사람일수록 비멸칭 부정단어에 비해 멸칭을 더 허용 가능한 표현으로 판단하는 경향을 보였다. 이러한 차이는 주관적 사용 빈도가 개인의 언어 사용 습관을 반영하는 반면, 사회적 허용 수준은 공격적인 표현을 얼마나 용인할 수 있는지에 대한 규범 판단을 반영한다는 차이점에서 비롯되었을 가능성이 있다.

사회적 허용 수준에서는 폭력 수용성뿐 아니라 공감, 돌봄, 공정성과도 뚜렷한 관련이 관찰되었다. 폭력 수용성이 멸칭 허용 수준을 높이는 방향으로 나타난 것과 달리, 공감, 돌봄, 공정성은 이를 낮추는 방향으로 나타났다. 이는 멸칭 허용 판단이 단순히 특정 단어를 자주 사용하는가의 문제가 아니라, 그 표현이 타인에게 해를 가하는지, 부당한 배제를 정당화하는지, 혹은 사회적으로 용인될 수 있는지를 평가하는 규범 판단과 관련됨을 시사한다. 특히 공감은 멸칭을 공격적이고 부당한 표현으로 해석하는 경향과 연관되어 허용 수준을 낮추었을 가능성이 있다(Decety & Yoder, 2016). 돌봄과 공정성 역시 해악과 부당한 대우에 대한 민감성을 반영한다는 점에서 유사한 역할을 추측할 수 있다(Graham et al., 2011).

사회적 허용 수준과 관련된 사회인지적 특성은 멸칭을 정서적·공격적으로 평가하는 방식과도 연결되었다. 매개분석 결과, 공감과 공정성은 멸칭을 더 부정적이거나 공격적인 표현으로 해석하는 경향과 함께 낮은 허용 수준으로 이어지는 간접 경로를 나타내었다. 반대로 충성성은 멸칭을 덜 부정적으로 해석하는 경향과 함께 더 높은 허용 수준과 연결되었다. 특히 정서가를 통한 간접효과는 공감, 공정성, 충성성 모두에서 멸칭에 한하여 선택적으로 관찰되었다. 또한, 지각된 공격성을 경유하는 공감의 간접효과는 멸칭 및 부정단어 모두에서 관찰되었으나 그 효과는 멸칭에서 보다 두드러졌다. 이는 멸칭에 대한 사회적 허용이 단순한 태도의 문제가 아니라, 해당 표현을 얼마나 공격적이고 부정적인 것으로 해석하는지와도 밀접하게 연결될 수 있음을 시사한다.

그러나, 일부 도덕기반 관련 결과는 하위척도의 측정 신뢰도를 고려하여 보수적인 해석을 적용할 필요가 있다. 충성성은 멸칭 허용 수준과 정적으로 관련되었으나, 본 연구에서 충성성 하위척도의 신뢰도는 비교적 낮았다(α = .52). 따라서 이를 충성성 도덕기반의 안정적인 효과로 단정하기보다는, 집단 경계와 관련된 도덕 판단이 멸칭 허용에 관여할 가능성을 보여주는 탐색적 결과로 이해하는 것이 타당하다. 같은 이유로 권위(α = .51)와 순수성(α = .54)이 뚜렷한 관련성을 보이지 않았다는 결과 역시 두 도덕기반이 멸칭 판단과 무관함을 의미한다고 단정하기 어렵다. 향후 연구에서는 신뢰도가 충분히 확보된 측정 도구를 통해 이러한 결과를 반복 검증할 필요가 있다.

한편, 사회적 맥락을 반영하는 배경 변인들도 멸칭의 사용과 허용을 이해하는 데 보조적인 정보를 제공하였다. 인터넷 사용 시간과 온라인 혐오 노출 경험이 많을수록 멸칭에 대한 주관적 사용 빈도가 높았으며, 지각된 도덕적 일치는 멸칭 사용과 허용 모두를 낮추는 방향으로 나타났다. 이는 멸칭 사용이 개인의 사회인지적 특성뿐 아니라 온라인에서 접하는 담론 환경, 자신과 타인이 도덕적 기준을 공유한다고 느끼는 정도와도 관련될 수 있음을 시사한다. 반면, 감정표현불능 수준(PAQ)은 멸칭의 주관적 사용 빈도나 사회적 허용 수준과 뚜렷한 관련을 보이지 않았다. 이는 멸칭 사용과 허용 판단이 정서 인식의 일반적인 어려움보다는 온라인 혐오에 대한 허용 태도, 폭력 수용성, 공감, 도덕기반과 같은 사회인지적·규범적 특성과 더욱 직접적으로 연결되어 있을 가능성을 시사한다. 종합하면, 본 연구의 결과는 멸칭 혐오표현의 사용 경향과 허용 판단이 서로 다른 사회인지적 특성 및 매개 경

로와 관련될 수 있음을 보여준다.

본 연구에는 몇 가지 제한점이 존재한다. 첫째, 본 연구는 횡단 자료에 근거하므로 사회인지 변인과 주관적 사용 빈도, 사회적 허용 수준 사이의 인과 관계를 단정하기 어렵다. 온라인 혐오 수용성이 멸칭 사용 경향을 높일 수도 있지만, 반복적인 멸칭 사용 및 노출 경험이 온라인 혐오에 대한 허용 태도를 강화할 가능성을 배제할 수 없기 때문이다. 후속 연구에서는 종단 설계 또는 실험을 통해 이러한 전후 관계를 보다 명확히 파악할 필요가 있다.

둘째, 본 연구에서 사용한 주관적 사용 빈도는 실제 발화 행동이 아니라 참가자가 스스로 보고한 사용 경향을 반영한다. 따라서 본 연구의 결과를 실제 온라인 게시글, 댓글, 대화 상황에서의 발화 빈도와 곧바로 연결 지어 해석하는 데는 주의가 필요하다. 후속 연구에서는 실제 온라인 발화 자료나 행동 관찰 자료를 활용하여, 주관적 사용 경향과 실제 언어 사용 사이의 관계를 보다 직접적으로 검증할 필요가 있다.

셋째, 본 연구의 자극은 양현보(2025)에서 사용한 멸칭 20개와 비멸칭 부정단어 20개에 한정된다. 따라서 본 연구의 결과는 자극 집합 내에서 관찰된 멸칭과 부정단어의 차이를 보여주지만, 이 결과를 모든 한국어 멸칭 및 부정단어에 일괄적으로 적용하는 데에는 신중할 필요가 있다. 또한, 부정어는 실제 대화나 온라인 게시글 안에서 사용될 때 발화자의 의도, 사용 맥락, 대상 집단과의 관계, 인용 여부 등에 따라 의미와 허용 가능성이 달라질 수 있다(Almagro et al., 2022). 후속 연구에서는 보다 다양한 유형의 멸칭과 문장·담화 수준의 사용 사례를 포함하는 한편, 본 연구에서 관찰된 사회인지적 개인차 패턴이 실제 언어 사용 맥락에서 어떻게 나타나는지를 함께 검토할 필요성이 제기된다.

넷째, 측정 측면에서도 몇 가지 제약이 존재한다. 본 연구는 탐색적 목적에서 다양한 사회인지적 변인을 함께 조사하는 과정에서 온라인 혐오 수용성(OHA)과 폭력 수용성(AOV) 등 일부 척도를 한국어로 직접 번안하여 사용하였다. 따라서 후속 연구에서는 해당 척도들에 대해 역번역, 전문가 평정, 탐색적·확인적 요인분석, 수렴·변별타당도 검토를 포함한 별도의 심리측정 검증을 수행할 필요가 있다. 또한 K-MFQ 하위척도 가운데 충성성, 권위, 순수성의 경우 비교적 신뢰도가 낮았으므로 관찰된 효과와 방향성의 해석에 주의가 요구된다. 향후 보다 정밀한 타당화 절차를 거친 한국판 심리측정 도구를 활용하여 반복 검증을 수행함으로써 본 연구 결과의 일반화 가능성을 제고할 필요가 있다.

본 연구는 한국어 멸칭 혐오표현의 사용과 허용이 단일한 혐오 성향으로 설명되지 않으며, 서로 다른 사회인지적 조건 속에서 형성될 수 있음을 보여주었다. 이는 멸칭이 단순히 공격적인 단어가 아니라, 실제로 사용할 수 있는 표현이라고 생각하는 정도와 사회적으로 용인 가능하다고 판단하는 정도가 온라인 혐오 태도, 공격 정당화, 도덕기반, 공감, 표현 해석 방식과 밀접하게 상호작용하며 조절되는 언어 현상임을 시사한다. 그러므로, 심각한 사회문제로 대두되고 있는 멸칭 혐오표현의 확산 기제를 이해하고 적절히 대응하기 위해서는 해당 표현의 언어적 특성을 넘어 그러한 표현이 사용 가능하고 허용 가능한 것으로 받아들여지는 사회인지적 조건을 함께 고려할 필요성이 제기된다. 본 연구가 멸칭의 사용과 허용의 심리사회적 배경을 이해하고

혐오표현의 확산에 대응하는 방안을 모색하는 데 기여할 수 있기를 희망하는 바이다.

참고문헌

강희숙 (2018). 혐오 표현의 특징 및 변주 양상. 호남학, 64, 33-64.

김환, 한수미 (2016). 한국판 토론토 공감 질문지의 타당화 연구. 한국심리학회지: 임상, 35(4), 809-821.

박세훈, 한승희, 유금란 (2025). 농담이 되어버린 혐오: 폄하 유머와 유머 지각 그리고 정당한 세상에 대한 개인적인 믿음의 관계. 한국심리학회지: 문화 및 사회문제, 31(4), 769-793.

석자춘 (2022). 한국형 간문화 도덕성 검사 도구의 설계와 활용. 도덕윤리과교육, 77, 227-255.

양현보 (2025). 멸칭 혐오표현의 심리언어적 특성 및 정서 처리에 미치는 영향. 한국심리학회지: 문화 및 사회문제, 31(4), 645-666.

장경현 (2021). 신조어 비하 표현의 특성 연구. 우리말글, 90, 1-24.

조예진 (2024). 한국판 Perth 감정표현불능증 척도(Korean-Perth Alexithymia Questionnaire: K-PAQ)의 타당화 연구. 한국심리학회지: 건강, 29(5), 837-860.

홍성수 (2019). 혐오표현의 해악과 개입의 정당성: 금지와 방치를 넘어서. 법철학연구, 22(3), 27-64.

Almagro, M., Hannikainen, I. R., & Villanueva, N. (2022). Whose words hurt? Contextual determinants of offensive speech. *Personality and Social Psychology Bulletin, 48*(6), 937-953.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic: The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology, 41*(S1), 3-33.

Carnaghi, A., & Maass, A. (2007). In-group and out-group perspectives in the use of derogatory group labels: Gay versus fag. *Journal of Language and Social Psychology, 26*(2), 142-156.

Čehajić-Clancy, S., Janković, A., Opačin, N., & Bilewicz, M. (2023). The process of becoming 'we' in an intergroup conflict context: How enhancing intergroup moral similarities leads to common-ingroup identity. *British Journal of Social Psychology, 62*(3), 1251-1270.

Celuch, M., Oksanen, A., Räsänen, P., Costello, M., Blaya, C., Zych, I., Llorent, V. J., Reichelmann, A., & Hawdon, J. (2022). Factors associated with online hate acceptance: A cross-national six-country study among young adults. *International Journal of*

Environmental Research and Public Health, 19(1), 534.

Cervone, C., Augoustinos, M., & Maass, A. (2021). The language of derogation and hate: Functions, consequences, and reappropriation. *Journal of Language and Social Psychology, 40*(1), 80-101.

Croom, A. M. (2014). The semantics of slurs: A refutation of pure expressivism. *Language Sciences, 41*, 227-242.

Decety, J., & Yoder, K. J. (2016). Empathy and motivation for justice: Cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Social Neuroscience, 11*(1), 1-14.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27-46.

Fasoli, F., Maass, A., & Carnaghi, A. (2015). Labelling and discrimination: Do homophobic epithets undermine fair distribution of resources? *British Journal of Social Psychology, 54*(2), 383-393.

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology, 46*(2), 237-248.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366-385.

Kennedy, B., Golazizian, P., Trager, J., Atari, M., Hoover, J., Mostafazadeh Davani, A., & Dehghani, M. (2023). The (moral) language of hate. *PNAS Nexus, 2*(7), pgad210.

Kim, K. R., Kang, J. S., & Yun, S. H. (2012). Moral intuitions and political orientation: Similarities and differences between South Korea and the United States. *Psychological Reports, 111*(1), 173-185.

Pluta, A., Mazurek, J., Wojciechowski, J., Wolak, T., Soral, W., & Bilewicz, M. (2023). Exposure to hate speech deteriorates neurocognitive mechanisms of the ability to understand others' pain. *Scientific Reports, 13*(1), 4127.

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior, 44*(2), 136-146.

Wolf, L. J., Hanel, P. H. P., & Maio, G. R. (2021). Measured and manipulated effects of value similarity on prejudice and well-being. *European Review of Social Psychology, 32*(1), 123-160.

논문 투고일 : 2026. 05. 28
1 차 심사일 : 2026. 07. 09
게재 확정일 : 2026. 08. 08

한국심리학회지 : 문화 및 사회문제
Korean Journal of Culture and Social Issues
2026, Vol. 32, No. 3, 387～409.

Social-Cognitive Characteristics of Subjective Use Frequency and Social Permission Level of Derogatory Hate Labels

Hyeonbo Yang†

Department of Brain and Cognitive Sciences, Korea Advanced Institute of Science and Technology (KAIST)

Derogatory hate labels have increasingly been recognized as a problematic form of language that reinforces social exclusion and hostile interaction. As such expressions become more widespread in online environments, they may come to be perceived not as deviant utterances, but as expressions that can realistically be used and socially tolerated. Accordingly, the present study examined which social-cognitive characteristics were associated with reporting more frequent use of Korean derogatory labels and judging such labels as more socially acceptable. A total of 190 university students rated Korean derogatory labels and non-derogatory negative words on 14 psycholinguistic dimensions, including subjective use frequency, social permission level, valence, and perceived offensiveness, and completed measures of moral foundations, online hate acceptance, acceptance of violence, empathy, and alexithymia. Linear mixed-effects model analyses showed that derogatory labels were rated as less frequently used and less socially permissible, but as more offensive, than non-derogatory negative words. In addition, online hate acceptance was selectively associated with the subjective use frequency of derogatory labels, whereas social permission level was positively associated with acceptance of violence and moral foundation-loyalty, and negatively associated with empathy and moral foundation-care and fairness. Mediation analyses further revealed indirect associations between empathy and social permission level of derogatory labels through valence and perceived offensiveness. Besides, indirect associations through only valence were seen in the connection of fairness and loyalty with social permission level of derogatory labels. These findings suggest that tendencies to use and socially permit derogatory labels cannot be reduced to a single dispositional characteristic. Rather, understanding the mechanisms underlying the spread of derogatory hate labels and reacting to them appropriately require careful consideration of individual social-cognitive characteristics.

Key words : Derogatory labels, Hate speech, Subjective use frequency, Social permission, Moral foundations

† Corresponding author: Hyeonbo Yang, Department of Brain and Cognitive Sciences, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea. Email: hyeonbo@kaist.ac.kr

부 록

부록 A. 척도별 문항 수, 점수 범위, 기술통계 및 신뢰도

척도	문항 수	점수 범위	M	SD	신뢰도(α)
온라인 혐오 수용성 (OHA)	5	1-5	2.08	0.73	.79
폭력 수용성 (AOV)	15	1-5	2.08	0.43	.75
한국판 토론토 공감 질문지 (TEQ)	16	1-5	3.80	0.53	.84
한국판 Perth 감정표현불능증 척도 (PAQ)	24	1-5	2.09	0.60	.92
부정 감정 묘사의 어려움 (PAQ-부정 DDF)	4	1-5	2.29	0.84	.76
부정 감정 인식의 어려움 (PAQ-부정 DIF)	4	1-5	2.19	0.78	.72
긍정 감정 묘사의 어려움 (PAQ-긍정 DDF)	4	1-5	2.04	0.74	.78
긍정 감정 인식의 어려움 (PAQ-긍정 DIF)	4	1-5	1.82	0.62	.74
외부지향적 사고 (PAQ-EOT)	8	1-5	2.11	0.80	.90
도덕기반-돌봄 (K-MFQ-Care)	6	0-5	3.42	0.68	.60
도덕기반-공정성 (K-MFQ-Fairness)	6	0-5	3.40	0.68	.66
도덕기반-충성성 (K-MFQ-Loyalty)	6	0-5	2.45	0.62	.52
도덕기반-권위 (K-MFQ-Authority)	6	0-5	2.44	0.67	.51
도덕기반-순수성 (K-MFQ-Purity)	6	0-5	2.45	0.66	.54

부록 B. 조사 단어 목록

멸칭	비멸칭 부정단어
개저씨	강도
걸레	깡패
급식충	노름꾼
김여사	노숙자
김치녀	놈팽이
깜둥이	도둑놈
꼰대	방관자
꼴페미	배신자
대깨	변태
도태남	불량배
뚱녀	불효자
맘충	사기꾼
병신	살인자
빨갱이	아첨꾼
설명충	악당
오크녀	양아치
외노	위선자
찐따	절도범
틀딱	죄수
한남충	폭주족

주. 아래 표는 본문 분석에 사용한 멸칭 20개와 비멸칭 부정단어 20개를 가나다순으로 정리한 것이다.

부록 C. 14개 심리언어 평정 차원 및 문항 구성

차원	질문	응답
의미 이해 (Meaning comprehension)	이 단어의 의미를 얼마나 잘 알고 있습니까?	전혀 모른다(1점) - 매우 잘 안다(9점)
정서가 (Valence)	이 단어가 얼마나 부정적, 혹은 긍정적으로 느껴집니까?	매우 부정적(1점) - 매우 긍정적(9점)
각성가 (Arousal)	이 단어가 얼마나 편하게, 혹은 긴장되게 느껴집니까?	매우 편함(1점) - 매우 긴장됨(9점)
추상성 (Abstractness)	이 단어의 의미가 얼마나 구체적 혹은 추상적입니까?	매우 구체적(1점) - 매우 추상적(9점)
정서 경험 (Emotional experience)	이 단어로 인해 얼마나 자주 감정을 경험합니까?	매우 드물게(1점) - 매우 자주(9점)
친숙성 (Familiarity)	이 단어를 얼마나 자주 보거나 듣습니까?	매우 드물게(1점) - 매우 자주(9점)
주관적 사용 빈도 (Subjective use frequency)	이 단어를 얼마나 자주 사용합니까?	매우 드물게(1점) - 매우 자주(9점)
심상성 (Imageability)	이 단어의 의미를 얼마나 쉽게 상상할 수 있습니까?	매우 어려움(1점) - 매우 쉬움(9점)
지각된 공격성 (Perceived offensiveness)	이 단어가 얼마나 공격적이라고 느껴집니까?	전혀 공격적이지 않음(1점) - 매우 공격적임(9점)
사회적 허용 수준 (Social permission level)	이 단어가 사회적으로 얼마나 사용되어도 괜찮다고 생각합니까?	절대 사용돼서는 안 됨(1점) - 얼마든지 사용될 수 있음(9점)
자기와의 유사성 (Perceived similarity to self)	이 단어가 가리키는 집단이 자신과 얼마나 닮았다고 생각합니까?	전혀 닮지 않음(1점) - 매우 닮음(9점)
성별 지시성 (Gender directionality)	이 단어가 가리키는 집단이 얼마나 여성적 또는 남성적이라고 생각합니까?	매우 여성적(1점) - 매우 남성적(9점)
범주 실체성 (Category entitativity)	이 단어가 가리키는 집단이 얼마나 의미 있고, 통합되고, 응집력 있다고 생각합니까?	전혀 그렇지 않음(1점) - 매우 그러함(9점)
범주 본질주의 (Category essentialism)	이 단어가 가리키는 집단이 얼마나 시간과 상황에 걸쳐 일관된 특성을 갖고 있다고 생각합니까?	전혀 그렇지 않음(1점) - 매우 그러함(9점)

주. 모든 단어 차원 평정은 1점부터 9점까지의 9점 척도로 제시하였다.

부록 D. 직접 번안하여 사용한 문항

다음 문항은 Celuch 등(2022)의 문항을 본 연구에서 한국어로 직접 번안해 사용한 것이다. 응답 범주는 두 척도 모두 1점(매우 동의한다)부터 5점(매우 동의하지 않는다)까지였으며, 분석에서는 역채점 후 점수가 높을수록 수용 수준이 높도록 처리하였다.

D1. 온라인 혐오 수용성 (Online Hate Acceptance, OHA)

1. 온라인에서 누군가가 먼저 나를 공격하기 시작했다면 혐오 또는 비하 메시지를 보내도 괜찮다.
2. 온라인에서 누군가가 내 친구나 가족을 모욕한다면 혐오 또는 비하 메시지를 보내도 괜찮다.
3. 혐오표현은 재미있기 때문에 온라인에서 누군가에게 혐오 또는 비하 메시지를 보내도 괜찮다.
4. 혐오 또는 비하 메시지는 사람들이 당신을 존중하도록 만드는 방식이므로 보내도 괜찮다.
5. 누군가가 당신의 종교, 출신, 피부색을 이유로 당신을 조롱하거나 모욕한다면 혐오 또는 비하 메시지를 보내도 괜찮다.

D2. 폭력 수용성 (Acceptance of Violence, AOV)

1. 법은 어기도록 만들어져 있다.
2. 다른 사람을 해치지 않는 한 원하는 것은 무엇이든 해도 괜찮다.
3. 친구끼리 또는 가족끼리 싸우는 것은 남일이 아니다.
4. 친구를 곤경에서 벗어나게 하기 위해 거짓말을 해도 괜찮다.
5. 다른 사람들이 사람을 때리고 있는 상황을 고려했을 때, 일부 물건을 부수는 것은 큰 일이 아니다.
6. 사람들이 저지르는 불법적인 일들에 비하면 가게에서 물건을 훔치는 것은 별로 심각하지 않다.
7. 친구들의 압박에 의한 것이었다면, 아이들의 못된 짓은 비난받을 수 없다.
8. 누군가가 당신에게 먼저 싸움을 걸었다면 그 사람에게 폭력을 사용해도 괜찮다.
9. 누군가가 내 친구나 가족을 모욕했다면 그 사람에게 폭력을 사용해도 괜찮다.
10. 폭력은 재미있기 때문에 사용해도 괜찮다.
11. 폭력은 사람들이 당신을 존중하도록 만드는 방식이므로 사용해도 괜찮다.
12. 누군가가 당신의 종교, 출신, 피부색을 이유로 당신을 조롱하거나 모욕한다면 폭력을 사용해도 괜찮다.
13. 세상의 문제를 해결하기 위해 폭력을 사용하는 것은 괜찮다.
14. 테러 행위를 저질러도 괜찮다.
15. 불의에 맞서 싸우기 위해 폭탄을 사용하는 것은 괜찮다.

부록 E. 한국어판 도덕기반 질문지(K-MFQ)

아래 응답 범주는 설문 제시 형식을 따른다. 앞서 부록 A에 수록된 K-MFQ 점수 범위는 분석을 위해 각 문항을 0-5점으로 환산한 뒤 하위척도 평균으로 계산한 값이다. K-MFQ는 30개의 채점 문항과 2개의 검사 문항으로 구성되어 있으며, 검사 문항인 6번과 22번은 하위척도 점수 산출에서 제외하고, 나머지 30문항을 바탕으로 돌봄・공정성・충성성・권위・순수성 점수를 각각 6문항의 평균으로 산출한다.

E1. 관련성 판단 문항 (문항 1-16)

응답 범주: 1점 = 전혀 관련이 없음, 2점 = 별로 관련이 없음, 3점 = 약간 관련 있음, 4점 = 다소 관련 있음, 5점 = 매우 관련 있음, 6점 = 극히 관련 있음.

1. 어떤 사람이 정서적으로 고통받았는지 여부
2. 일부 사람들이 나머지 사람들과 다르게 대우받았는지(차별받았는지) 여부
3. 어떤 사람의 행동이 나라에 대한 사랑을 나타냈는지 여부
4. 어떤 사람이 권위에 대한 존경심을 보였는지 여부
5. 어떤 사람이 순수함과 품위에 관한 정도를 넘어섰는지(침해했는지) 여부
6. 어떤 사람이 수학을 잘했는지 여부
7. 어떤 사람이 약하거나 취약한 사람을 배려했는지 여부
8. 어떤 사람이 불공정하게 행동했는지 여부
9. 어떤 사람이 자신이 속한 집단을 배신하는 어떤 일을 했는지 여부
10. 어떤 사람이 사회의 전통을 따랐는지 여부
11. 어떤 사람이 혐오스러운 행동을 했는지 여부
12. 어떤 사람이 잔인했는지 여부
13. 어떤 사람이 자신의 권리를 거부당했는지 여부
14. 어떤 사람이 충성심 부족을 드러냈는지 여부
15. 어떤 행동이 혼란이나 무질서를 야기했는지 여부
16. 어떤 사람이 신이 허락할 만한 방식으로 행동했는지 여부

E2. 동의 판단 문항 (문항 17-32)

응답 범주: 1점 = 강하게 동의하지 않는다, 2점 = 중간 정도로 동의하지 않는다, 3점 = 약간 동의하지 않는다, 4점 = 약간 동의한다, 5점 = 중간 정도로 동의한다, 6점 = 강하게 동의한다.

17. 고통받는 사람들에 대한 연민은 가장 중요한 미덕이다.
18. 정부가 법을 만들 때, 가장 중요한 원칙은 모든 사람이 공정하게 대우받는 것을 보장하는 것이어야 한다.
19. 나는 우리나라의 역사가 자랑스럽다.
20. 모든 아이는 권위를 존중하는 태도를 배워야 한다.
21. 아무도 피해를 보지 않더라도, 사람들은 혐오감을 유발하는 일을 해서는 안 된다.
22. 나쁜 일을 하는 것보다 좋은 일을 해야 한다.
23. 어떤 개인이 할 수 있는 최악의 일 중 하나는 스스로 방어하지 못하는 동물을 해치는 것이다.
24. 정의는 사회의 가장 중요한 필수조건이다.
25. 가족이 잘못된 일을 했다 하더라도 사람들은 자신의 가족에게 충실해야 한다.
26. 사회에서 남자와 여자는 각각 다른 역할을 지니고 있다.
27. 나는 어떤 행동들을 부자연스럽기 때문에 잘못되었다고 말할 것이다.
28. 사람을 죽이는 것은 결코 옳지 않다.
29. 가난한 아이들이 아무것도 물려받지 못하고 부자 아이들은 많은 재산을 물려받는 것은 도덕적으로 잘못된 일이라고 생각한다.
30. 자기 자신을 드러내는 것보다 팀원으로서의 역할을 하는 것이 더 중요하다.

31. 내가 군인이라면, 지휘관의 명령에 동의하지 않더라도 명령에 따르는 것이 나의 의무이므로 명령에 복종할 것이다.
32. 순결은 중요하고 소중한 미덕이다.

부록 F. 본 연구에서 구성한 단일 문항

1. 정치성향

질문: 자신의 정치적 성향을 표시해 주세요.

응답 범주: 1점 = 매우 진보적, 9점 = 매우 보수적

2. 인터넷 사용 시간

질문: 하루에 인터넷을 몇 시간 정도 사용하십니까?

응답 범주: 1점 = 1시간(또는 그 미만), 9점 = 9시간(또는 그 이상)

3. 지각된 도덕적 일치

질문: 자신의 도덕적 가치관이 다른 사람들과 얼마나 비슷하다고 느낍니까?

응답 범주: 1점 = 전혀 비슷하지 않음, 9점 = 매우 비슷함

4. 온라인 혐오 노출

질문: 온라인에서 어떤 집단에 대한 부정적 시선을 표현하는 내용을 얼마나 자주 목격하십니까?

응답 범주: 1점 = 매우 드물게, 9점 = 매우 자주