

Online Appendix

Validation Protocols for Algorithm Output

As described in the main narrative, we assessed the validity of our NLP dictionary approach by comparing the algorithm’s output against that of a team of trained human raters. We moreover use a variety of “simulated” algorithm output comparisons to benchmark our actual ICC results against, thus creating plausible upper and lower bounds for possible algorithm performance.

To illustrate this thinking, we can begin by examining the results of this exercise for the phenomenon of top-level leadership in Appendix Figure A1. On the left, we show that the ICC for just the set of human raters was about 0.63. In other words, even among only humans, we should expect a baseline level of disagreement and differences of opinion that make perfect agreement unattainable in concept. When we add the algorithm’s actual output as an additional “rater” and recalculate the ICC, the ICC looks almost identical at 0.68, and the difference is not statistically significant. This indicates that the algorithm disagrees with the human raters, if anything, slightly *less often* than the human raters disagree with one another.

While this is on its face heartening for our results, we should importantly determine the extent to which the algorithm’s agreement with the human raters could be *worse* or *better*. To examine this directly, we simulate a number of possible scenarios. First, we calculate what the ICC would look like if the algorithm simply guessed at random whether a given activity description example indicated top-level leadership using base probabilities informed by the human raters’ responses – we refer to this as “random distributional guessing.” We see that the ICC in this scenario would be significantly lower at 0.32, indicating that our actual algorithm performed far better than such random guessing. Second, we calculate what the ICC would look like if the algorithm instead “cheated” and intentionally provided the response that would

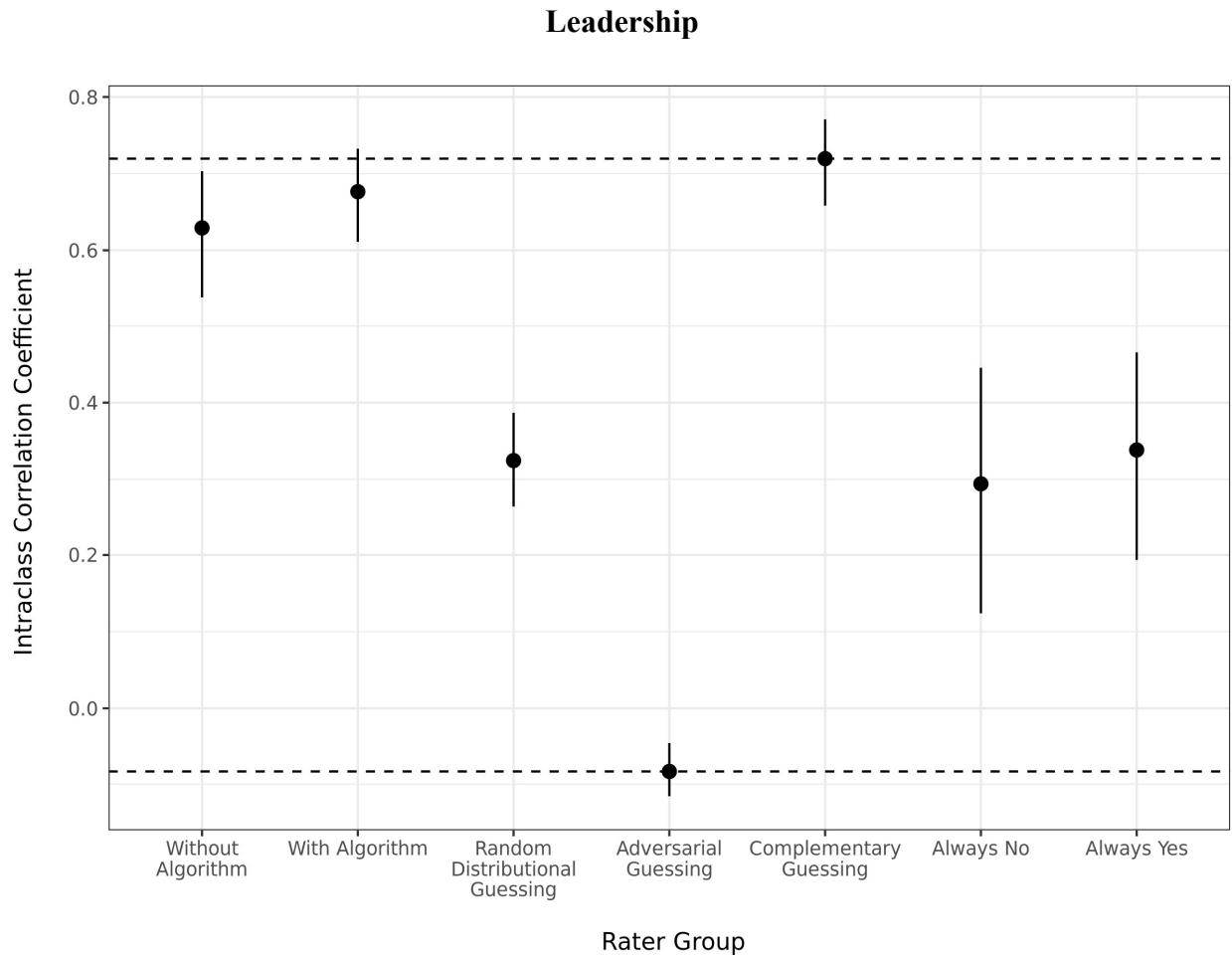
generate the highest possible levels of *disagreement* with the human raters (e.g., say a description does not have top-level leadership in it when all human raters say it does) – we call this the “adversarial guessing” model. The ICC in this case drops precipitously to -0.09, which can be thought of as the theoretical lower bound for algorithm performance. We can also calculate the opposite in what we refer to as the “complementary guessing” model – when the algorithm “cheats” and provides the response that would generate the highest possible levels of *agreement* with the human raters. The ICC in this case is not significantly higher than the actual reality at 0.71, indicating that our actual algorithm’s agreement with the human raters could not theoretically be any higher. We complete this exercise by showing the ICC in cases where the algorithm only guesses “Yes” or “No” for the existence of top-level leadership in any given example. Appendix Figure A2 shows the results of this same process, but for the phenomenon of excellence instead; while the absolute levels of ICC are lower across the board, we again show that our algorithm’s output is extremely high relative to the theoretical upper and lower bounds established here.

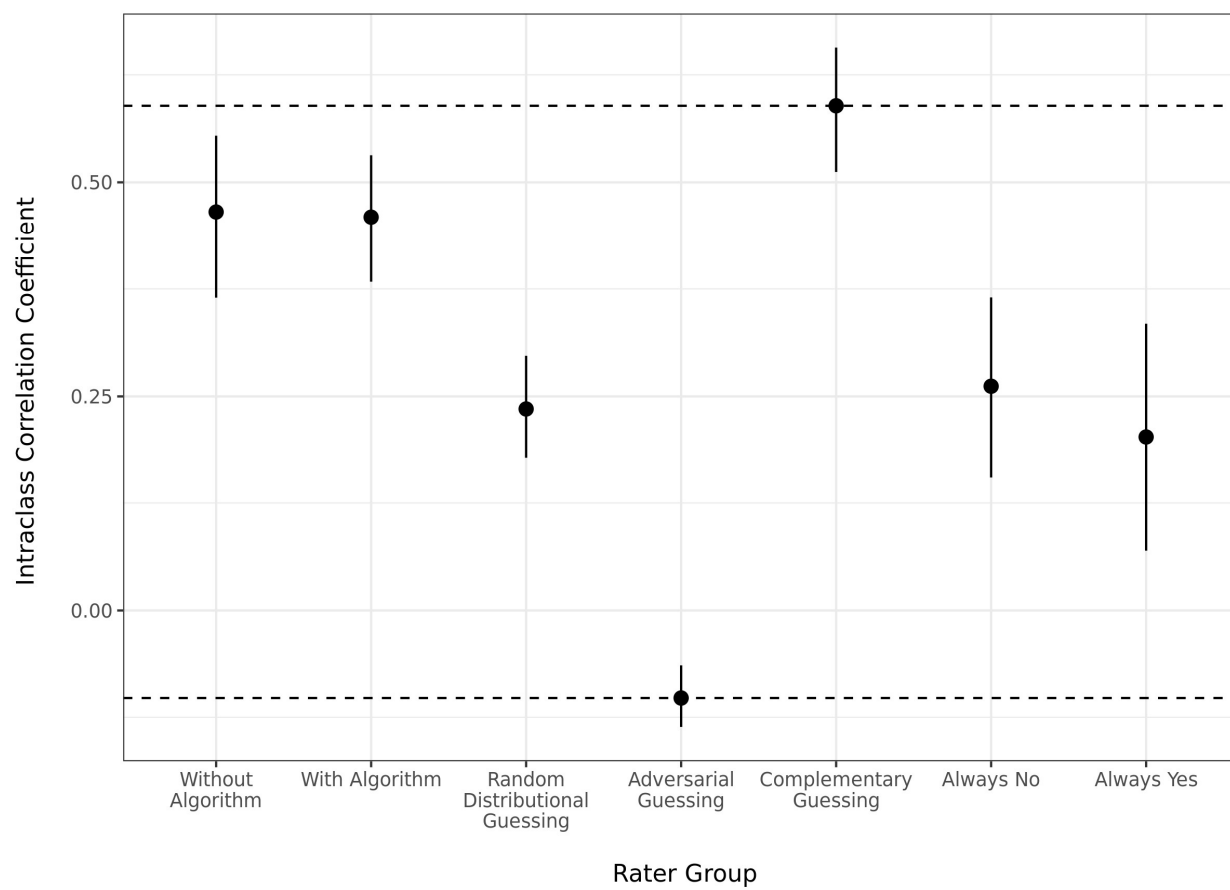
As an additional examination, we also assess the possible existence of algorithmic bias in our model in Appendix Figures A3 and A4. Because we stratified our random sample of activity description examples by applicant sex, URM status, first-gen status, fee waiver receipt, and public/private high school type, we can explicitly examine whether the ICC is significantly different across these groups. If we detect a significant difference in rater ICC across these groups, any disparities we measure across these demographics in our main analyses could be the result of differential accuracy of our algorithm, rather than actual disparities in the data themselves. We show in these figures that this is not the case; that is, we do not observe any detectable difference in ICC across any demographic groups, and thus offer evidence that

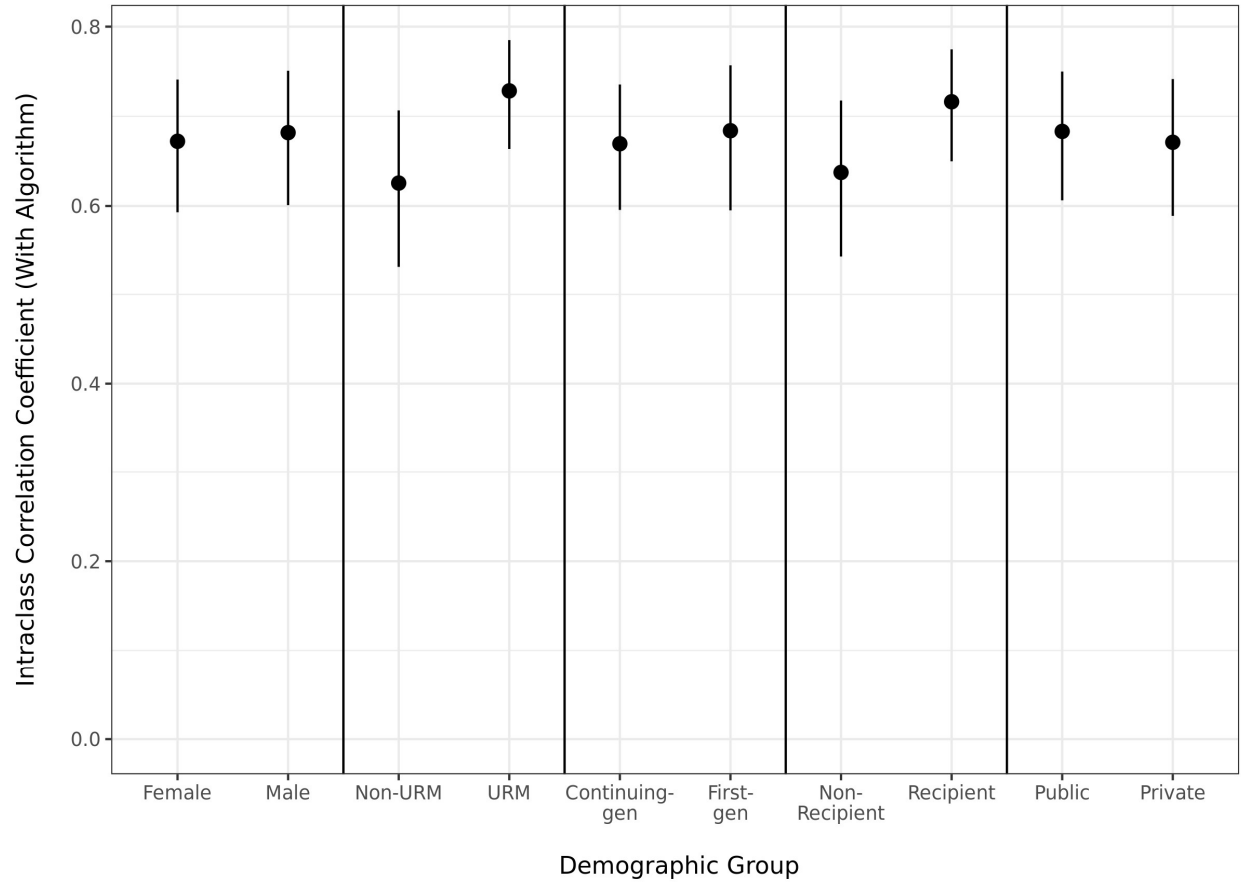
algorithmic bias is not an issue in our context. The only demographic difference approaching significance is the female-male difference for excellence; however, as we do not examine disparities on this axis in this paper, this possible algorithmic bias is unlikely to affect our main findings.

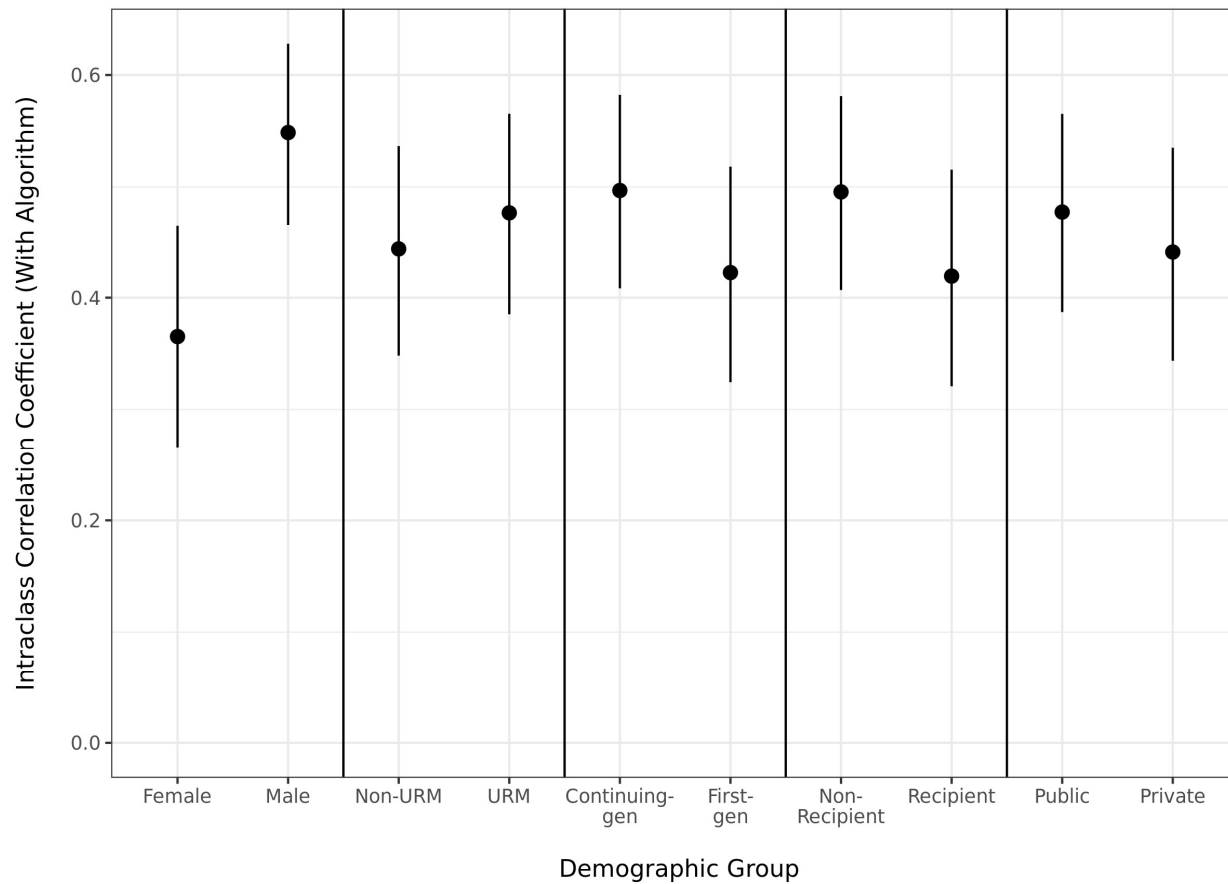
Appendix Tables and Figures

Appendix Figure A1. Intraclass Correlation Coefficient Analysis Results: Top-level



Appendix Figure A2. Intraclass Correlation Coefficient Analysis Results: Excellence

Appendix Figure A3. Algorithmic Bias Analysis: Top-level Leadership

Appendix Figure A4. Algorithmic Bias Analysis: Excellence

Appendix Table A1. Top-level Leadership Phrases and Exclusions

Phrase	Examples of Exclusions (random subset of 5 displayed per phrase for concision)
cadet_chief_master_sergeant	worked_for_the_cadet_chief_master_sergeant, work_for_the_cadet_chief_master_sergeant, help_the_cadet_chief_master_sergeant, coordinate_with_the_cadet_chief_master_sergeant, cadet_chief_master_sergeant_intern
cadet_colonel	shadowed_the_cadet_colonel, asst_cadet_colonel, supported_cadet_colonel, coordinate_with_cadet_colonel, interned_for_the_cadet_colonel
cadet_sergeant_major	intern_for_the_cadet_sergeant_major, coordinate_with_cadet_sergeant_major, worked_for_the_cadet_sergeant_major, shadow_the_cadet_sergeant_major, shadow_cadet_sergeant_major
captain	intern_for_captains, vice_captain, supported_captains, junior_captains, assist_the_captain
captains	coordinate_with_the_captains, work_for_the_captains, captains_council, my_captains, coordinate_with_captains
captian	shadowed_captian, intern_for_captian, associate_captian, interned_for_the_captian, jr_captian
ceo	support_the_ceo, help_ceo, coordinated_with_the_ceo, supported_the_ceo, assc_ceo
chair	coordinated_with_the_chairperson, asst_chairwoman, shadowed_the_chairman, assc_chairman, assisted_the_chairman
chairman	intern_for_the_chairman, assisted_chairman, work_for_the_chairman, shadow_the_chairman, support_the_chairman
chairperson	jr_chairperson, interned_for_the_chairperson, chairperson_intern, assisted_chairperson, junior_chairperson
chairwoman	shadowed_the_chairwoman, supported_chairwoman, shadow_chairwoman, work_for_chairwoman, assisted_the_chairwoman
cocaptain	assisted_cocaptain, coordinated_with_cocaptain, asst_cocaptain, my_cocaptain, assisted_the_cocaptain

cofounder	cofounder_asst, shadow_cofounder, assisted_cofounder, work_for_cofounder, work_for_the_cofounder
commander	supported_commander, associate_commander, helped_commander, assisted_the_commander, assisted_commander
concert_master	assisted_concert_master, shadow_concert_master, assc_concert_master, coordinate_with_concert_master, helped_the_concert_master
concertmaster	assist_the_concertmaster, work_for_concertmaster, support_the_concertmaster, coordinated_with_the_concertmaster, junior_concertmaster
conductor	shadow_the_conductor, associate_conductor, interned_for_the_conductor, dep_conductor, junior_conductor
copresident	interned_for_copresident, assc_copresident, assistant_copresident, my_copresident, assist_the_copresident
director	assist_director, intern_for_the_director, dep_director, asst_director, help_the_director
drum_major	shadow_drum_major, dep_drum_major, assisted_drum_major, work_for_drum_major, worked_for_the_drum_major
drummajor	help_drummajor, interned_for_the_drummajor, assist_drummajor, vice_drummajor, drummajor_asst
editor_in_chief	help_the_editor_in_chief, coordinated_with_the_editor_in_chief, work_for_editor_in_chief, vice_editor_in_chief, assc_editor_in_chief
editorinchief	assisted_editorinchief, shadow_editorinchief, assist_editorinchief, support_editorinchief, asst_editorinchief
entrepreneur	support_entrepreneur, my_entrepreneur, help_the_entrepreneur, entrepreneur_congress, assist_entrepreneur
entrepreneur	intern_for_the_entrepreneur, coordinated_with_entrepreneur, assisted_entrepreneur, helped_the_entrepreneur, support_entrepreneur
executive_officer	coordinate_with_the_executive_officer, worked_for_the_executive_officer, vice_executive_officer, help_executive_officer, deputy_executive_officer

executiveofficer	associate_executiveofficer, assistant_executiveofficer, coordinated_with_the_executiveofficer, asst_executiveofficer, executiveofficer_asst
founder	coordinated_with_the_founder, intern_for_founder, cofounder_asst, assist_founder, intern_for_the_founder
head	assist_the_head, shadow_head, shadow_the_head, helped_the_head, head_asst
key_club_lieutenant_governor	vice_key_club_lieutenant_governor, helped_key_club_lieutenant_governor, coordinated_with_key_club_lieutenant_governor, my_key_club_lieutenant_governor, interned_for_the_key_club_lieutenant_governor
key_club_lt_governor	junior_key_club_lt_governor, helped_the_key_club_lt_governor, key_club_lt_governor_assistant, assist_key_club_lt_governor, shadow_key_club_lt_governor
lieutenant_governor_key_club	lieutenant_governor_key_club_asst, helped_lieutenant_governor_key_club, assisted_the_lieutenant_governor_key_club, lieutenant_governor_key_club_assistant, supported_lieutenant_governor_key_club
lt_governor_key_club	shadow_the_lt_governor_key_club, associate_lt_governor_key_club, worked_for_lt_governor_key_club, asst_lt_governor_key_club, assc_lt_governor_key_club
owner	assist_the_owner, my_owner, asst_owner, worked_for_the_owner, shadowed_the_owner
president	jr_president, shadow_president, intern_for_president, associate_president, copresident_asst
secretary_general	helped_secretary_general, worked_for_secretary_general, assist_the_secretary_general, shadowed_secretary_general, interned_for_secretary_general
secretarygeneral	support_the_secretarygeneral, shadowed_the_secretarygeneral, asst_secretarygeneral, coordinate_with_the_secretarygeneral, helped_the_secretarygeneral
sole_proprietor	supported_sole_proprietor, supported_the_sole_proprietor, vice_sole_proprietor, coordinate_with_the_sole_proprietor, work_for_sole_proprietor

supervisor	shadow_supervisor, assisted_supervisor, shadow_the_supervisor, interned_for_supervisor, intern_for_the_supervisor
------------	---

Appendix Table A2. Excellence Phrases and Exclusions

Phrase	Examples of Exclusions (random subset of 5 displayed per phrase for concision)
1st_chair	
1st_place	
1st_team_all_league	
2nd_place	
3rd_place	
4_h_diamond_star	
4_h_emerald_star	
4h_diamond_star	
4h_emerald_star	
all_american	all_american_steak_house, all_american_steakhouse
all_county	all_county_festival
all_regional	
all_star	all_star_cheer
all_state	all_state_insurance
allamerican	
allcounty	
allregional	
allstar	allstar_cheer
allstate	allstate_insurance
author	shadow_the_author, author_assistant, assisted_author, junior_author, jr_author
award	award_committee, award_show
awarded	
black_belt	

boys_nation	boys_nation_finalist, boys_nation_nomination, nominated_for_boys_nation
boys_state	
captain	my_captain, coordinated_with_the_captain, help_captain, worked_for_the_captain, deputy_captain
captian	asst_captian, supported_the_captian, vice_captian, junior_captian, associate_captian
champion	food_champion
champions	
championship	
champs	champs_sports
coauthor	assc_coauthor, assisted_coauthor, help_coauthor, work_for_coauthor, interned_for_the_coauthor
cocaptain	support_the_cocaptain, junior_cocaptain, dep_cocaptain, intern_for_the_cocaptain, vice_cocaptain
coeditor_in_chief	
competitive	competitive_market_place, competitive_marketplace, non_competitive, not_competitive
copresident	supported_the_copresident, associate_copresident, dep_copresident, work_for_the_copresident, work_for_copresident
delivered_tedx	
eagle_scout	becoming_an_eagle_scout, toward_eagle_scout
eaglescout	
editor_in_chief	
editorinchief	editorinchief_assistant, coordinate_with_the_editorinchief, helped_editorinchief, jr_editorinchief, work_for_editorinchief
elite	elite_learning_group
finalist	boys_nation_finalist, girls_nation_finalist
first_chair	
first_place	

first_team_all_league	
girls_nation	girls_nation_finalist, girls_nation_nomination, nominated_for_girls_nation
girls_state	
governors_school	
inaugural_tedx	
intel	shadowed_at_intel, shadowed_intel
isef	
junior_olympic	junior_olympic_volunteer
junior_olympics	junior_olympics_volunteer
lead	support_lead, assist_lead, coordinated_with_lead, my_lead, coordinated_with_the_lead
life_scout	
lifescout	support_the_lifescout, coordinated_with_lifescout, lifescout_asst, asst_lifescout, supported_lifescout
medal	
medalist	
merit_scholar	
most_valuable_player	
mostvaluableplayer	
mvp	
olympic_development	
olympics_development	
premier	adobe_premier, premier_service, premier_software, service_premier, wrote_premier
president	coordinate_with_the_president, vice_president, intern_for_president, worked_for_the_president, supported_the_president
prize	gave_prize, prize_ceremony, prize_desk

qualifier	
recipient	
scholar_athlete	
scholarship	focused_on_scholarship, scholarship_organization, scholarship_organizer
second_place	
selective	
semifinalist	
siemens	
solo	deputy_soloist, shadowed_soloist, supported_soloist, shadow_soloist, helped_soloist
soloist	work_for_soloist, junior_soloist, supported_the_soloist, dep_soloist, shadowed_soloist
special_citation	
starter	starter_culture, starter_cultures
state_level	
statelevel	
talent_search	
talentsearch	
tedx_speaker	
tedx_talk	
third_place	
top	on_top_of, top_colleges
upward_bound_scholar	
varsity	junior_varsity
westinghouse	
winner	helped_the_winner, associate_winner, dep_winner, shadowed_the_winner, assisted_winner

Appendix Table A3. Full List of Original and Condensed Activity Type Options

Original Activity Type Option	Condensed Activity Type
Academic	Academic
Computer/Technology	Academic
Debate/Speech	Academic
Journalism/Publication	Academic
Research	Academic
Robotics	Academic
Science/Math	Academic
Art	Arts
Dance	Arts
Music: Instrumental	Arts
Music: Vocal	Arts
Theater/Drama	Arts
Athletics: Club	Athletics
Athletics: JV/Varsity	Athletics
Career Oriented	Career
Internship	Career
Junior R.O.T.C.	Career
Work (Paid)	Career
Cultural	Culture and Identity
Foreign Exchange	Culture and Identity
Foreign Language	Culture and Identity
LGBT	Culture and Identity
Religious	Culture and Identity
Social Justice	Culture and Identity
Environmental	Other
Family Responsibilities	Other
Other Club/Activity	Other
School Spirit	School Governance/Spirit
Student Government/Politics	School Governance/Spirit
Community Service (Volunteer)	Service