Applying Elastic-net Regression to Identify the Best Models Predicting Changes in Civic Purpose during the Emerging Adulthood

Hyemin Han Kelsie J. Dawson

University of Alabama

Author Note

Hyemin Han and Kelsie J. Dawson, Educational Psychology Program, University of Alabama.

Correspondence concerning this article should be addressed to Hyemin Han, Educational Psychology Program, University of Alabama, Tuscaloosa, AL 35487. Contact: hyemin.han@ua.edu

The authors report no conflicting interests.

Applying Elastic-net Regression to Identify the Best Models Predicting Changes in Civic Purpose during the Emerging Adulthood

Abstract

Introduction: Changes in civic purpose during the emerging adulthood has been a significant research topic since it is closely associated with active civic engagement later in human lives. While standard regression methods have been used in previous studies to predict civic purpose development, they have limitations that may not always lead to best prediction models. We aimed to address these limitations by utilizing elastic-net multinomial logistic regression, which favors models with the least number of necessary predictors, in exploration of predictors for civic purpose development in a data-driven manner.

Methods: We analyzed data from the longitudinal Civic Purpose Project while focusing on the model that best predicted civic purpose from Wave 1 (before high school graduation) to Wave 2 (two years after Wave 1). The reanalyzed data included responses from 480 participants recruited Californian high schools. The elastic-net regression was performed 5,000 times for predicting three dependent variables, Wave 2 political purpose, community service purpose, and expressive activity purpose, with Wave 1 predictors. We identified which predictors were selected as the constituents of the best regression models during the elastic-net regression process.

Results: Results showed that civic purpose, moral and political identity, and external supports (e.g., parental and peer involvement, school civic opportunities, etc.) in Wave 1 significantly predicted civic purpose in Wave 2. Several predictors were excluded from the regression models during the elastic-net regression process.

Conclusion: We found that the elastic-net regression was able to present the more regularized model for prediction. Implications for promoting civic purpose are discussed as well as utilizing the elastic-net regression method.

Keywords: civic purpose; emerging adulthood; elastic-net regression; positive youth development; data science

Applying Elastic-net Regression to Identify the Best Models Predicting Changes in Civic Purpose during the Emerging Adulthood

Promoting civic engagement during adolescence and emerging adulthood is important in human flourishing because experiences of early civic engagement enable youths to develop and practice skills to continue civic engagement through adulthood (Youniss et al., 2002). Previous research has reported that civic purpose, one's purpose to engage in civic activities, contributes to maintaining civic engagement during adolescence and early adulthood, which are significant transitional periods in a person's life (Malin, Ballard, & Damon, 2015; Malin, Han, & Liauw, 2017). This research has found that civic purpose is constituted by three components, long-term intention to engage in civic activities, actual civic engagement, and motivation to contribute to beyond-the-self beings (Malin, Ballard, et al., 2015). Several factors including identity style (Crocetti, Erentaitė, & Žukauskienė, 2014), moral and political identity (Han, Ballard, & Choi, 2021), parental and peer civic attitude and engagement (Duke, Skay, Pettingell, & Borowsky, 2009; Moran, Bundick, Malin, & Reilly, 2013; Zaff, Malanchuk, & Eccles, 2008), and school atmosphere and support (Andolina, Jenkins, Zukin, & Keeter, 2003; Ballard, Caccavale, & Buchanan, 2015; Keating & Janmaat, 2016) have been found to be conducive to the formation of civic purpose and engagement.

Previous studies have tested which factors would significantly predict civic purpose with standard regression, such as ordinary linear regression. However, due to the methodological limitations of standard regression, they could not examine which regression model is the best model among all possible candidate models consisting of combinations of candidate predictors (Kim, Banerjee, Park, & Pathak, 2016). First, epistemologically, ordinary regression analysis based on null hypothesis significant testing with *p*-values can only reject a null hypothesis (e.g.,

coefficients with a zero effect size) within the one model that is being tested each time (Han, Park, & Thoma, 2018). It is not suitable to examine which predictors (independent variables) are meaningful and should be included in a model (Harrell, 2015; Walter & Tiemeier, 2009). Second, model selection methods based on standard regression, such as the stepwise method, are likely to inflate estimated coefficients associated with smaller *p*-values, while they could not address issues related to collinearity and correlation among predictors (Hammami, Lee, Ouarda, & Lee, 2012; Harrell, 2015; Tibshirani, 1997). Third, when standard regression is performed, only one model, which was set by a researcher, can be tested (Han & Dawson, 2021). So, if the researcher wants to explore the best model among all possible models in terms of combinations of candidate predictors in a hypothesis-free, data-driven manner, such standard regression could not be an ideal method.

Thus, we intend to analyze the civic purpose project (CPP) dataset with elastic-net regression to explore the best model and predictors in a data-driven manner without setting any a priori hypothesis. In the data-driven model exploration, we examined what would be the best regression models predicting changes in civic purpose during emerging adulthood. As candidate predictors to be examined, we employed variables regarding the baseline civic purpose (Malin, Ballard, et al., 2015), civic identity (Han et al., 2021; Malin et al., 2017; Porter, 2013), attitudes and supports regarding civic engagement among peers and parents (Duke et al., 2009; Moran et al., 2013; Zaff et al., 2008), and school environment (Andolina et al., 2003; Ballard, Caccavale, et al., 2015; Keating & Janmaat, 2016) from the CPP dataset based on prior research. Although each individual cited prior study examined several aspects of civic purpose development, the prior studies have not been able to consider diverse person-level and environmental factors in a simultaneous manner, due to the methodological limitations. Hence, we decided to examine how

such diverse factors contributed to prediction of civic purpose development in a data-driven manner for better understanding of the developmental process.

While addressing the question, we were interested in presenting methodological benefits of elastic-net regression in research on adolescence by demonstrating its superiority to standard regression. We decided to employ elastic-net regression because it prefers the simplest model with the least necessary predictors and penalizes the inclusion of unnecessary predictors (see "Analysis" section in supplementary methods for the model estimation mechanism). Although no previous studies using elastic-net have been published in journals for research on adolescence (e.g., Journal of Adolescence, Journal of Youth and Adolescence, Journal of Research on Adolescence), statisticians have proven that elastic-net can produce better outcomes in terms of the accuracy and simplicity of regression models (Grave, Obozinski, & Bach, 2011; Zou & Hastie, 2005). For instance, in epidemiology, when a prediction is made with numerous candidate predictors, elastic-net regression contributes to the estimation of the most stringent model with improved prediction accuracy (Kim et al., 2016).

In addition, we also examined whether the model identified by elastic-net regression can better predict outcomes out of the boundary of data used for regression compared with standard regression through cross-validation (McNeish, 2015). Previous studies reported when the best regression model with parsimony was identified in a data-driven manner, such a model could more accurately predict outcomes even beyond the data used for regression compared with the model from standard regression and less susceptible to over-fitting (Han & Dawson, 2021). Given attrition is a significant issue in research on adolescence (Jeličić, Phelps, & Lerner, 2009), by demonstrating that the result from elastic-net regression can accurately predict developmental

outcomes out of the boundary of data used for regression, we would be able to present another significant practical benefit of this method in research on adolescence.

Methods[1]

Materials

We downloaded and analyzed files containing two-wave survey data from the CPP dataset that is available for public via the Inter-university Consortium for Political and Social Research (https://www.icpsr.umich.edu/icpsrweb/civicleads/studies/36561). The CPP researchers collected Wave 1 data before participants' high school graduation at California high schools, which resulted in 1,578 participants completing the survey. Wave 2 data was collected two years later via Qualtrics and 480 participants completed the survey (see "Materials" section in supplementary methods for further details about the nature of the dataset). We found that the attrition rate was significantly different across different gender and ethnicity groups (see "Materials" section in supplementary methods for additional information).

Measures

Participants' civic purpose and variables that have been found to be associated with civic purpose (e.g., moral and political identity; supports from parents, peers, and schools) were analyzed. We used three civic purpose statuses as categorical variables, political, community service, and expressive activity purpose statuses, as dependent variables. Further details regarding each measure are available in "Measures" section in supplementary materials.

Civic purpose. For dependent variables, civic purpose was measured in three domains: political, community service, and expressive activities. We classified each participant's political, community service, and expressive activity purpose at Waves 1 and 2 into five categories, i.e., drifting, dreamers, dabblers, self-oriented, and purposeful, with three variables, civic intent, civic

[1] Technical further details regarding the methods used in the present study are elaborated in Supplementary Methods.

engagement, and beyond-the-self motivation (see Table 1 for the classification criteria). Civic intent was intent to engage in civic activities in general. For civic engagement, three different types of engagement, i.e., political, community service, and expressive activity engagement, were measured independently. Beyond-the-self motivation was measured in two domains, political activity and volunteering. The numbers of participants classified into the civic purpose status categories are summarized in Table S1. Some participants did not respond to all civic purpose-related items, so the summed numbers in Table S1 were smaller than 1,578 and 480, the total sample size at Wave 1 and 2, respectively.

Moral and political identity. Moral and political identity were measured at Wave 1 to indicate whether a participant regarded moral and political values as central to his/her self-concept. Participants were presented with items representing different values (either moral, political, or neutral) and asked to score to what extent each value was central to themselves. As one reported more moral and political values as central to him/herself, he/she received the higher moral and political identity score, respectively.

Parental, peer, and school supports. We used four variables that quantify supports from parents, peers, and schools in civic activities at Wave 1 for analysis. These variables were parental civic involvement, peer civic involvement, perceived importance of civic and academic values among peers, and opportunities for involvement in civic activities at school.

Demographics. We used seven demographical variables in this study following prior purpose studies. The variables include gender, college boundedness, ethnicity, participants' place of birth, mother's place of birth, father's place of birth, and socioeconomic status.

Analysis

First, we conducted correlation analysis to examine the relationship between civic purpose-related variables. We focused on Pearson correlation among continuous variables, i.e., civic intent and engagement at both waves, moral and political identity, and variables about parental, peer, and school supports.

Second, we performed multinomial elastic-net multinomial regression to search for the best prediction model for each dependent variable, i.e., Wave 2 political, community service, or expressive activity purpose statuses as categorical variables. For this analysis, we composed our customized R script with *glmnet* package (Friedman et al., 2021). The dependent variables were treated as categorical variables. We entered candidate predictors (i.e., Wave 1 purpose status; moral and political identity; four variables about parental, peer, and school supports; seven demographics variables) to the model.

For each dependent variable, we repeated the aforementioned process for elastic-net regression 5,000 times. Elastic-net regression was repeated multiple times because the dataset was shuffled for cross-validation in a random manner, so we wanted to minimize any random error that could originate from random data shuffling by averaging results across the repetitions (see "Notes on glmnet" subsection in supplementary methods for further details about the cross-validation and random data shuffling). Previous studies that conducted the similar procedure for model exploration with cross-validation repeated model exploration 1,000 to 10,000 times (Han & Dawson, 2021; Han, Lee, & Soylu, 2020). We determined the number of repetitions in our study, 5,000, to compromise required computation time and prediction performance based on the range of the repetition numbers in the previous studies. The repeated cross-validation processes

were performed with multiple processors to save computation time (Friedman et al., 2021; Han, 2021).

We counted how many times each candidate predictor was included in the resultant model out of 5,000 repetitions. We also calculated the mean of each coefficient. According to prior research in computational methods, this procedure was employed and suggested as a way to integrate results from multiple repetitions of elastic-net regression with cross-validation and to improve prediction performance (Boulesteix, De Bin, Jiang, & Fuchs, 2017). Further details regarding elastic-net regression are explained in "Notes on elastic net" subsection in supplementary methods.

Furthermore, we examined whether elastic-net regression better predicted outcomes out of the boundary of data used for regression and, thus, was less susceptible to over-fitting and able to address attrition better compared with standard regression. For this examination, we compared the validation dataset prediction accuracy between the two methods, multinomial elastic-net regression and conventional multinomial regression. Methodological details about this comparison are explained in "Notes on validation dataset prediction accuracy test" subsection in supplementary methods.

The analyzed data file and all source code files are available for public via the Open Science Framework (https://osf.io/zb7um/).

Results

Brief descriptive statistics of analyzed variables were presented in Table S2. The result of correlation analysis is reported in Figure 1. We found significant association among civic purpose-related variables both in Waves 1 and 2. Table 3 presents how many times each predictor showed a non-zero coefficient for each dependent variable among 5,000 trials. Figures

2-4 show the mean value of each coefficient when Wave 2 political, community service, and expressive activity purpose was predicted, respectively. Only the predictors that showed non-zero coefficients at least once out of 5,000 iterations are presented. For better representation of effect sizes, all reported coefficients were standardized. Exact estimated coefficient values are presented in Tables S3-5. Furthermore, when the validation dataset was examined, elastic-net regression significantly better predicted outcomes than standard regression. The comparison results are presented in Table S6.

Discussion

In this study, we used multinomial elastic-net to explore the best prediction model with the most significant predictors for changes in civic purpose statuses in three domains: political, community service, and expressive activities. The results showed that first, as shown in prior research, Wave 1 purpose status significantly predicted Wave 2 purpose status. Second, both moral and political identity were found to be significant predictors from the most of 5,000 iterations in all three domains. Participants who maintained civic purpose were likely to possess strong moral and political identity in Wave 1. Moreover, in general, external supports, including parental and peer involvement, peer civic and academic values, and school civic opportunities were positively associated with Wave 2 civic purpose.

In general, these findings are consistent with what have been reported in previous studies examining civic purpose. The importance of the baseline civic purpose statuses, and moral and political identity in predicting civic purpose development was originally reported in Han et al. (2021) and Malin et al. (2015). The association between demographical and environmental factors (e.g., supports from peers, parents, and schools) and civic purpose was demonstrated by Ballard, Malin, Porter, Colby, and Damon (2015), Malin et al. (2017), and Malin, Tirri, and

Liauw (2015). Despite the aforementioned consistency between findings from our and previous studies, we were able to examine more diverse candidate predictors in predicting civic purpose changes, which could not be examined in the previous studies, with a data-driven analysis method, multinomial elastic-net regression. The relationship between candidate predictors and civic purpose changes could be tested partially in each study as the previous studies employed standard regression based on a priori hypotheses. Unlike the previous studies, we were able to identify the best prediction models among all possible candidate models.

Interestingly, supports from parents, peers, and school differently predicted changes in different domains of civic purpose when standardized coefficients were compared. Compared with other supports, parental supports did not seem to strongly predict civic purpose transitions during emerging adulthood (standardized coefficients << .01). This result is consistent with the existing literature that underscores the role of non-familial mentors in purpose development during adolescence and early adulthood (Damon, 2008). Also, the significant association between peer involvement and values, and civic purpose might support previous research that reported that motivational influences of modeling can be strengthened when peers who are perceived to be relevant are presented as models (Han, Kim, Jeong, & Cohen, 2017; Lockwood & Kunda, 1997). Compared with two other domains, the change in expressive activity purpose was not strongly predicted by external support factors except presence of school opportunities. Expressive activities are somehow more individualistic than traditional political activities (Stanyer, 2005), so influences from others might not significantly predict purpose in this domain.

Regarding the treatment of attrition, the results from the validation dataset prediction accuracy test may suggest that elastic-net regression would be one of possible solutions to address the issue. As shown, the regression models generated by elastic-net regression

significantly more accurately predicted outcomes out of the boundary of the training dataset used for regression. The reported superiority of elastic-net regression in validation dataset prediction accuracy might support the point that the method would possibly better predict outcomes in the cases where missing occurred due to attrition or withdrawal. It would be because a regression model based on elastic-net regression is more parsimonious and less likely to be overfitted and biased compared with a regression model from standard regression (Han & Dawson, 2021).

In our study, we explored the best model to predict changes in civic purpose in three different domains with elastic-net multinomial logistic regression. The results will provide information regarding factors that most significantly predict changes in civic purpose during emerging adulthood. Furthermore, we were able to search for the simplest prediction model that could not be feasibly identified with classical regression methods (Walter & Tiemeier, 2009). Previous studies were able to test their own hypotheses focusing on different individual predictors. However, they could not examine which predictors should be used to constitute the best prediction model due to the methodological limitations. Thus, our study will provide researchers in the field with insights about how to use the novel method in data science for their future research, particularly that aim at exploring data without established hypotheses.

However, there are several limitations warranting further investigation. First, the significant gender and ethnicity-related bias in attrition could be further examined. Of course, as discussed, elastic-net regression might potentially be a viable way to address attrition given the result from the validation dataset prediction accuracy test and its tendency to regularize coefficients. However, we could not test other conventional methods to address the issue, such as imputation and inverse probability weighting (Seaman & White, 2013). Second, we focused on only one type of regularized regression, elastic-net regression. Two other more specific types of

regularized regression methods, LASSO and ridge regression (Zou & Hastie, 2005), may need to be examined in future studies. Third, regarding interpretation of resultant standardized coefficients, several predictors, particularly those related to external supports and demographics, were non-zero but their effect sizes were very small (< .01). Follow-up confirmatory studies that utilize the reported regression models as sources for a priori hypotheses shall be conducted to examine the effect sizes of such predictors with additional data.

References

Andolina, M. W., Jenkins, K., Zukin, C., & Keeter, S. (2003). Habits from home, lessons from school: Influences on youth civic engagement. *PS - Political Science and Politics*, *36*(2), 275–280. doi:10.1017/S104909650300221X

Ballard, P. J., Caccavale, L., & Buchanan, C. M. (2015). Civic Orientation in Cultures of Privilege: What Role Do Schools Play? *Youth & Society*, *47*(1), 70–94. doi:10.1177/0044118X14538464

Ballard, P. J., Malin, H., Porter, T. J., Colby, A., & Damon, W. (2015). Motivations for Civic Participation Among Diverse Youth: More Similarities than Differences. *Research in Human Development*, *12*(1–2), 63–83. doi:10.1080/15427609.2015.1010348

Boulesteix, A.-L., De Bin, R., Jiang, X., & Fuchs, M. (2017). IPF-LASSO: Integrative L1-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. *Computational and Mathematical Methods in Medicine*, *2017*, 1–14. doi:10.1155/2017/7691937

Crocetti, E., Erentaitė, R., & Žukauskienė, R. (2014). Identity styles, positive youth development, and civic engagement in adolescence. *Journal of Youth and Adolescence*, *43*(11), 1818–1828. doi:10.1007/s10964-014-0100-4

Damon, W. (2008). *The path to purpose: Helping our children fnd their calling in life*. New York: Free Press.

Duke, N. N., Skay, C. L., Pettingell, S. L., & Borowsky, I. W. (2009). From Adolescent Connections to Social Capital: Predictors of Civic Engagement in Young Adulthood. *Journal of Adolescent Health*, *44*(2), 161–168. doi:10.1016/j.jadohealth.2008.07.007

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2021). glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. Retrieved from https://cran.r-project.org/web/packages/glmnet/index.html

Grave, E., Obozinski, G. R., & Bach, F. R. (2011). Trace Lasso: a trace norm regularization for correlated designs (PDF Download Available). *Advances in Neural Information Processing Systems*, *24*, 2187–2195.

Hammami, D., Lee, T. S., Ouarda, T. B. M. J., & Lee, J. (2012). Predictor selection for downscaling GCM data with LASSO. *Journal of Geophysical Research: Atmospheres*, *117*(D17), n/a-n/a. doi:10.1029/2012JD017864

Han, H. (2021). BayesFactorFMRI: Implementing Bayesian Second-Level fMRI Analysis with Multiple Comparison Correction and Bayesian Meta-Analysis of fMRI Images with Multiprocessing. *Journal of Open Research Software*, *9*. doi:10.5334/jors.328

Han, H., Ballard, P. J., & Choi, Y.-J. (2021). Links between moral identity and political purpose during emerging adulthood. *Journal of Moral Education*, *50*(2), 166–184. doi:10.1080/03057240.2019.1647152

Han, H., & Dawson, K. J. (2021). Improved model exploration for the relationship between moral foundations and moral judgment development using Bayesian Model Averaging. *Journal of Moral Education*. doi:10.1080/03057240.2020.1863774

Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and Relevant Moral Exemplars Are More Effective than Extraordinary Exemplars in Promoting Voluntary Service Engagement. *Frontiers in Psychology*, *8*, 283. doi:10.3389/fpsyg.2017.00283

Han, H., Lee, K., & Soylu, F. (2020). Applying the Deep Learning Method for Simulating Outcomes of Educational Interventions. *SN Computer Science*, *1*(2), 70. doi:10.1007/s42979-020-0075-z

Han, H., Park, J., & Thoma, S. J. (2018). Why do we need to employ Bayesian statistics and how can we employ it in studies of moral education?: With practical guidelines to use JASP for educators and researchers. *Journal of Moral Education*, *47*(4), 519–537. doi:10.1080/03057240.2018.1463204

Harrell, F. E. (2015). *Regression Modeling Strategies*. Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-19425-7

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195–1199. doi:10.1037/a0015665

Keating, A., & Janmaat, J. G. (2016). Education Through Citizenship at School: Do School Activities Have a Lasting Impact on Youth Political Engagement? *Parliamentary Affairs*, *69*(2), 409–429. doi:10.1093/pa/gsv017

Kim, M.-H., Banerjee, S., Park, S. M., & Pathak, J. (2016). Improving risk prediction for depression via Elastic Net regression - Results from Korea National Health Insurance Services Data. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2016*, 1860–1869.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on

the self. *Journal of Personality and Social Psychology*, *73*, 91–103. doi:10.1037/0022-3514.73.1.91

Malin, H., Ballard, P. J., & Damon, W. (2015). Civic purpose: An integrated construct for understanding civic development in adolescence. *Human Development*, *58*(2), 103–130. doi:10.1159/000381655

Malin, H., Han, H., & Liauw, I. (2017). Civic purpose in late adolescence: Factors that prevent decline in civic engagement after high school. *Developmental Psychology*, *53*(7), 1384–1397. doi:10.1037/dev0000322

Malin, H., Tirri, K., & Liauw, I. (2015). Adolescent moral motivations for civic engagement: Clues to the political gender gap? *Journal of Moral Education*, *44*(1), 34–50. doi:10.1080/03057240.2015.1014324

McNeish, D. M. (2015). Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behavioral Research*, *50*(5), 471–484. doi:10.1080/00273171.2015.1036965

Moran, S., Bundick, M. J., Malin, H., & Reilly, T. S. (2013). How Supportive of Their Specific Purposes Do Youth Believe Their Family and Friends Are? *Journal of Adolescent Research*, *28*(3), 348–377. doi:10.1177/0743558412457816

Porter, T. J. (2013). Moral and political identity and civic involvement in adolescents. *Journal of Moral Education*, *42*(2), 239–255. doi:10.1080/03057240.2012.761133

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295. doi:10.1177/0962280210395740

Stanyer, J. (2005). The British public and political attitude expression: the emergence of a self-

expressive political culture? *Contemporary Politics*, *11*(1), 19–32. doi:10.1080/13569770500098623

Tibshirani, R. (1997). THE LASSO METHOD FOR VARIABLE SELECTION IN THE COX MODEL. *Statistics in Medicine*, *16*(4), 385–395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

Walter, S., & Tiemeier, H. (2009). Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology*, *24*(12), 733–736. doi:10.1007/s10654-009-9411-2

Youniss, J., Bales, S., Christmas-Best, V., Diversi, M., McLaughlin, M., & Silbereisen, R. (2002). Youth Civic Engagement in the Twenty-First Century. *Journal of Research on Adolescence*, *12*(1), 121–148. doi:10.1111/1532-7795.00027

Zaff, J. F., Malanchuk, O., & Eccles, J. S. (2008). Predicting Positive Citizenship from Adolescence to Young Adulthood: The Effects of a Civic Context. *Applied Developmental Science*, *12*(1), 38–53. doi:10.1080/10888690801910567

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*, 301–320. doi:10.1111/j.1467-9868.2005.00503.x

Tables

Table 1

Classifying civic purpose statuses.

Category	Civic intention	Civic engagement	Possession of beyond-the-self motivation
Drifting	Low	Low	-
Dreamer	High	Low	-
Dabbler	Low	High	-
Self-oriented	High	High	No
Purposeful	High	High	Yes

Figures

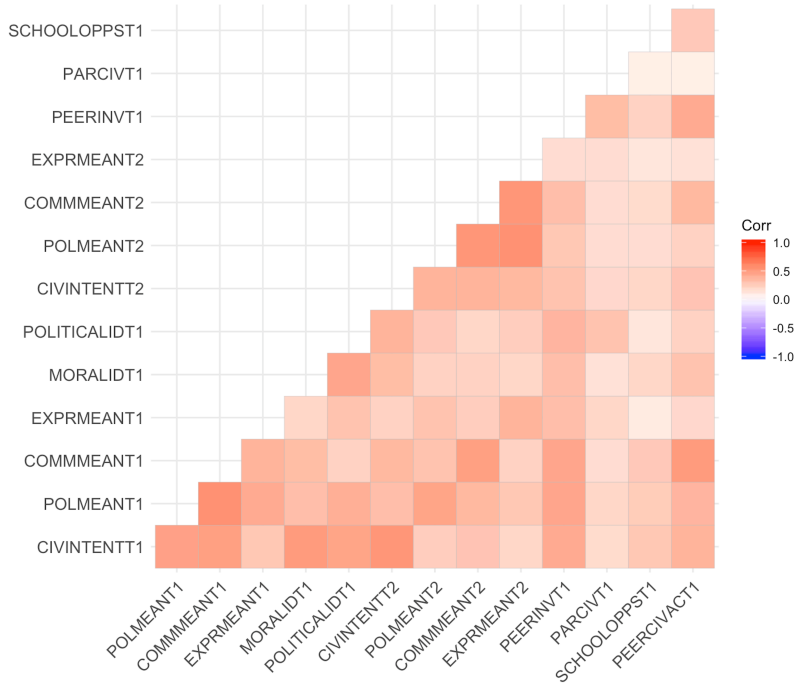


Figure 1. Correlation table. All colored cells represent at least $p < .05$. CIVINTENT: T1 or T2 civic intention. POLMEAN: T1 or T2 political engagement. COMMMEAN: T1 or T2 community service engagement. EXPRMEAN: T1 or T2 expressive activity engagement. MORALIDT1: T1 moral identity. POLITICALIDT1: T1 political identity. PEERINVT1: T1 peer civic involvement. PARCIVT1: T1 parental civic involvement. SCHOOLOPPST1: T1 school opportunity. PEERCIVACT1: T1 peer civic-academic value.

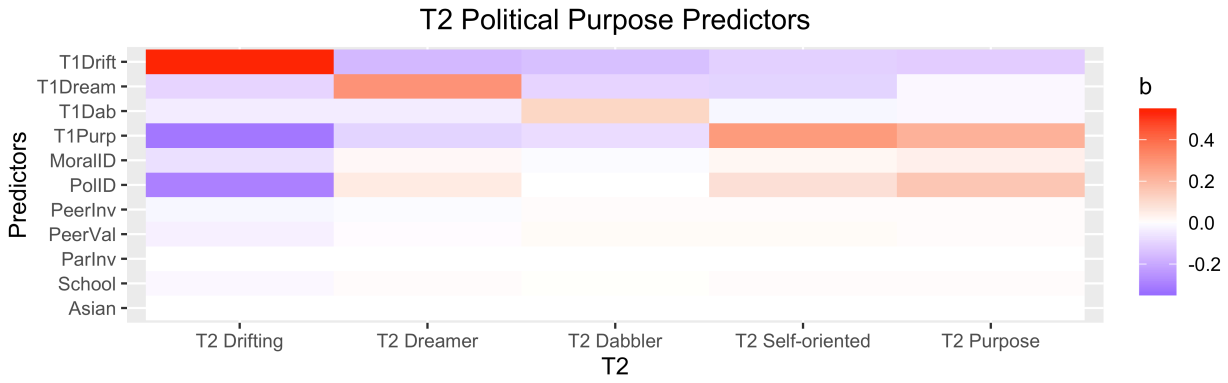


Figure 2. Estimated coefficients in T2 political purpose status prediction averaged over 5,000 iterations. MoralID: T1 moral identity. PolID: T1 political identity. PeerInv: T1 peer civic involvement. PeerVal: T1 peer civic-academic value. School: T1 school opportunity.

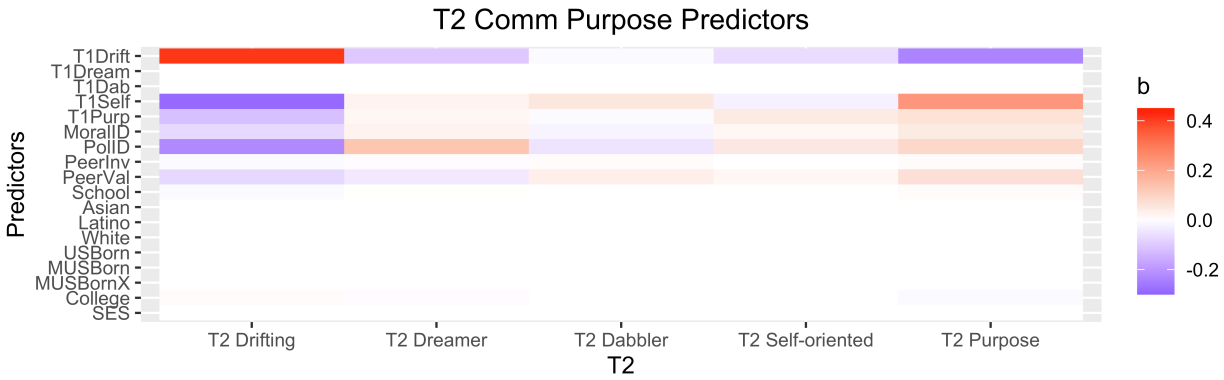


Figure 3. Estimated coefficients in T2 community service purpose status prediction averaged over 5,000 iterations. USBorn: Born in the USA. MUSBorn: Mother born in the USA. MUSBornX: Mother not born in the USA. College: T2 college boundedness.

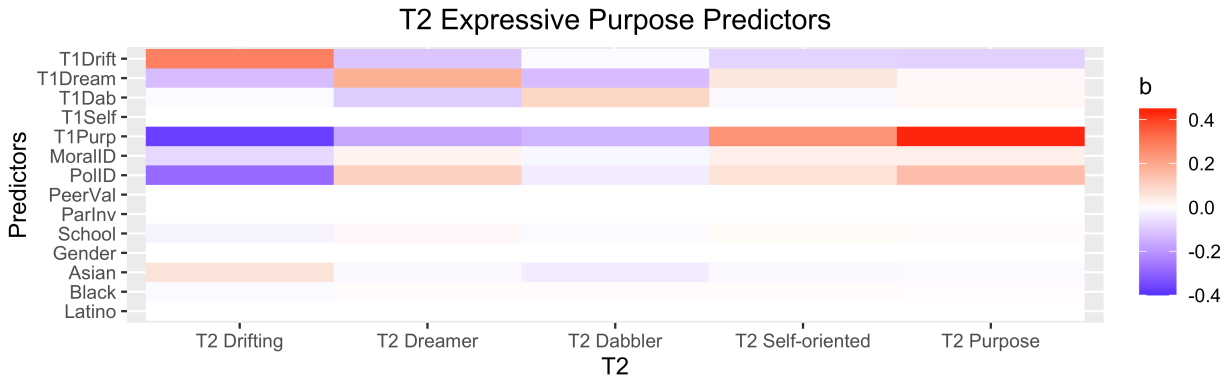


Figure 4. Estimated coefficients in T2 expressive activity purpose status prediction averaged over 5,000 iterations.

Supplementary Materials

Supplementary Methods

Materials

At Wave 1, 1,565 out of 1,578 participants reported their gender. 754 were males and 811 were females. 1,567 participants reported their ethnicity. The following are the responses:

401 Asian or Asian American, including Chinese, Indian, Korean, and others

85 Black or African American

723 Hispanic or Latino, including Mexican American, Colombian, Nicaraguan, and others

98 White, Caucasian, European American; not Hispanic

5 American Indian/Native American

153 Mixed; Parents are from different groups

102 Other

1,565 participants reported their birthplace. 1,319 were born in the United States while 246 were born out of the United States. 1,562 participants reported the place of birth of their mother. 461 were born in the United States, 1,060 were born out of the United States, and 41 were unknown. Also, 1,562 participants reported the place of birth of their father. 427 were born in the United States, 1,060 were born out of the United States, and 74 were unknown.

At Wave 2, out of 480 responses, we excluded 4 responses as they were apparently redundant or incorrectly collected. Thus, responses from 476 participants at Wave 2 were used for our analysis. Among them, 473 participants reported their gender. 186 were males and 287 were females. 473 participants reported their ethnicity. The following are the responses:

163 Asian or Asian American, including Chinese, Indian, Korean, and others

22 Black or African American

191 Hispanic or Latino, including Mexican American, Colombian, Nicaraguan, and others

29 White, Caucasian, European American; not Hispanic

0 American Indian/Native American

43 Mixed; Parents are from different groups

25 Other

471 participants reported their birthplace. 396 were born in the United States while 75 were not. 473 reported the birthplace of their mother. 105 were born in the United States, 360 were born out of the United States, and 8 were unknown. 474 participants reported the birthplace of their father. 101 were born in the United States, 356 were born out of the United States, and 17 were unknown. 476 participants reported their college boundedness. 438 were college-bounded while 38 were not. Also, 476 participants reported their socioeconomic (SES) status (M = 4.16, SD = 1.88). The distribution is presented in a histogram below:

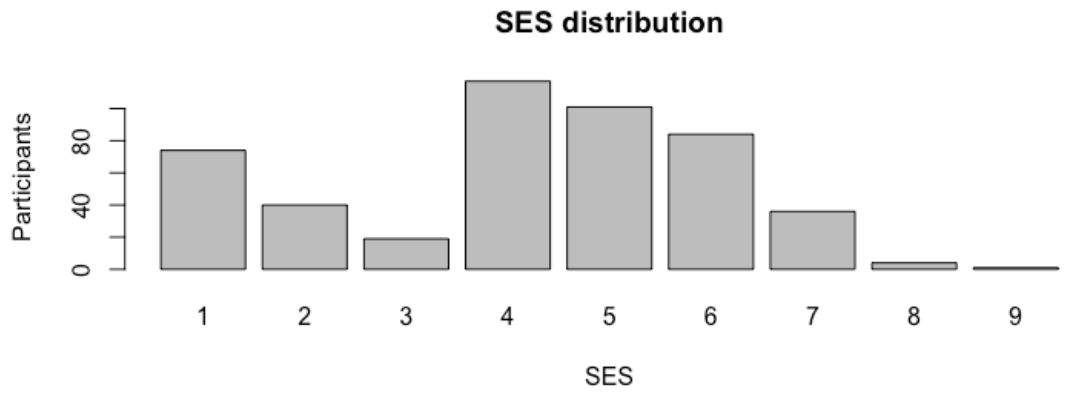


Figure S1. Distribution of participant SES.

In order to examine whether attrition occurred disproportionally across different gender and ethnicity groups, we conducted a series of logistic regression. First, we tested whether

attrition occurred in a gender biased manner. Among 811 female participants at Wave 1, 35.39% completed Wave 2 survey while 64.61% did not. Of 754 male participants at Wave 2, 24.67% participated in our study at Wave 2 while 75.33% did not. The result from logistic regression showed that there was a significant gender effect on attrition, logOR = .51, SE = .11, z = 4.60, p < .01.

Second, we also tested whether the attrition rate was significantly different across different ethnicity groups. The attrition rates were 59.35%, 74.12%, 73.58%, 70.41%, 100.00%, 71.90%, and 75.49% for Asian, Black, Hispanic, White, American Indian/Native American, Mixed ethnicity, and Other groups, respectively. The result from logistic regression suggested that there was a significant ethnicity effect on the attrition rate, χ^2 (6) = 31.05, p < .001.

Participants' civic purpose statuses at Waves 1 and 2 are summarized in Table S1.

Table S1

Participants' civic purpose statuses at Waves 1 and 2

	Drifting (0)	Dreamers (1)	Dabblers (2)	Self-oriented (3)	Purposeful (4)	Total
Wave 1						
Political	806	283	160	145	86	1,480
Community service	756	250	166	118	122	1,412
Expressive activity	712	302	204	129	62	1,409
Wave 2						
Political	247	91	51	35	31	455
Community service	242	88	50	12	47	439
Expressive activity	226	77	70	44	28	445

Measures

In this section, we briefly describe the variables that were analyzed in the present study. All the analyzed variables were imported from the CPP dataset. For the full description of the CPP dataset, including all measures and coding manuals, refer to the dataset available for public (https://www.icpsr.umich.edu/icpsrweb/civicleads/studies/36561).

Civic purpose. To classify participants into five different civic purpose categories (i.e., drifting, dreamers, dabblers, self-oriented, and purposeful), three civic purpose variables, civic intention, civic engagement, and presence of beyond-the-self motivation were measured. For further details regarding how participants were classified into five categories, refer to Han, Ballard, and Choi (2019).

Civic intention. Participants' civic intent was quantified by calculating the composite score of five items. Participants were asked whether civic activities were considered to be important and meaningful to their life goals. Then, they were presented with five civic-intention related items. The five items include "Being involved in politics," "Making a difference through volunteering," "Becoming a leader in my community," "Making positive changes in my community," and "Having an impact on a social cause or issue that is important to me." Participants' answers were anchored on a five-point Likert scale (*not at all meaningful* to *extremely meaningful*).

Civic engagement. Participants were presented with items representing different types of activities in order to calculate engagement in three activity domains. Then, they were asked how often they engaged in each activity since they entered high school. Their answers were anchored to a four-point Likert scale (never, once or twice, a few times, and regularly). Political engagement was measured with six items (i.e., "Held a leadership position in a school club").

Community service engagement was measured with five items (i.e., “Helped with a fund-raising project”). Expressive activity engagement was measured with four items (i.e., Wrote a letter to a school or community newspaper or publication”). We used the composite score of each activity domain.

Beyond-the-self motivation. Beyond-the-self motivation in the domains of political activity and volunteering was also coded in two binary variables. Beyond-the-self motivation in the political activity domain was used for the classification of political and expressive activity purpose; beyond-the-self motivation in volunteering was used for the classification of community service activity purpose.

Before examining beyond-the-self motivation, participants were presented with one question about whether they participated in political or volunteering activities. In the case of political beyond-the-self motivation, participants were presented with one item, “How involved in political activities are you?” Their answer was anchored to a four-point Likert scale (“not involved in political activities and don't want to get involved in next 6 months” to “very involved in political activities”). If a participant was somewhat or very involved in political activities, the participant was directed to the political beyond-the-self motivation question. The participant was presented with the following list of different motives for political activities:

To do something about an issue I care about. (*)

I wanted to take action on my beliefs. (*)

It is important for my religious/ethnic/cultural group. (*)

I wanted to be the kind of person who helps others. (*)

I’ve been given a lot; I want to give back. (*)

I became upset by something I saw happening. (*)

It is required at school.

It makes me feel good about myself.

To further my education or career goals.

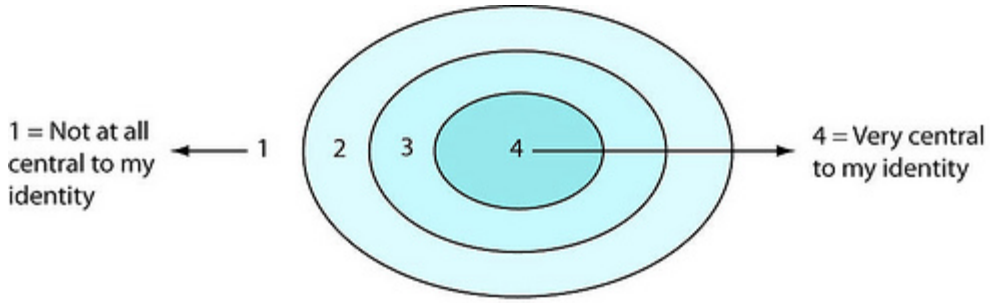
Somebody asked or encouraged me to participate.

To build skills or prepare for the future.

It sounded fun.

The items with a star (*) represent beyond-the-self motives; the others represent self-oriented motives. Once the list was presented to the participant, the participant was asked to select the three most important motives for political engagement from the list. Participants who selected at least one of the six beyond-the-self motive items as the most important motive for political engagement were identified as possessing beyond-the-self motivation in political activities (political beyond-the-self motivation = 1). The other participants were identified as not possessing political beyond-the-self motivation (0). The same procedures were applied to identify beyond-the-self motivation in volunteering as well.

Moral and political identity. To measure moral and political identity, participants were presented with a list of different values and asked to score to what extent each value was central to themselves. To assist their scoring process, the question was presented with a figure:



The following is a list of the moral values:

Being fair

Willing to stand up for what I believe is right

Compassionate, concerned about all kinds of people

Honest

Concerned about justice and human rights

Responsible, someone other can depend on

The following are the political values:

Concerned about international issues

Politically involved

Concerned about government decisions and policies

The moral identity score was calculated by averaging the quantified centralities (1: “not at all central to my identity” to 4: “very central to my identity”) of the six moral value items.

Similarly, the political identity score was calculated by averaging the quantified centralities of the three political value items.

Parental, peer, and school supports. Variables related to parental, peer, and school supports include four variables, i.e., parental civic involvement, peer civic involvement, whether peers highly regard civic and academic values, and opportunities for involvement in activities at school.

Parental civic involvement. Parental civic involvement was quantified by calculating the composite score of these four items: "My parents/guardians are active in the community," "My parents/guardians are active in local politics (e.g., school board, city council)," "My parents/guardians do volunteer work in the community," and "I talk to my parents/guardians about problems in society and political issues." Participants' answer to each item was anchored to a five-point Likert scale ("How much do you agree or disagree with the following statements?"; "strongly disagree" to "strongly agree").

Peer civic involvement. Peer civic involvement was measured by calculating the composite score of three items, i.e., "I talk to my friends about problems in society and political issues," "I have close friends who participate in political activities," and "I have close friends who do volunteer work in the community." Participants' answers were anchored to a five-point Likert scale ("How much do you agree or disagree with the following statements?"; "strongly disagree" to "strongly agree").

Peer civic and academic values. To what extent peers value civic and academic values was quantified by calculating the composite score of three items, i.e., "Participate in school clubs like drama, sports, and music," "Get good grades," and "Do volunteer service work." Participants were asked to what extent each value was regarded to be important among their

friends ("Among your friends, how important is it to do the following things?"). Their answers were anchored to a five-point Likert scale ("not at all important" to "extremely important").

Opportunities for involvement in activities at school. We also used a variable about whether opportunities for involvement in civic activities were available at school. This variable was measured by calculating the composite score of these six items: "Be involved in making decisions about the school," "Take on leadership roles in clubs and organizations," "Work on social issues or causes such as the environment or human rights," "Get involved in political activities," and "Participate in student government." Participants were asked whether each of the aforementioned opportunities was available in their school ("How much do you agree or disagree with the following? At my school, there are opportunities to…"). Their answers were anchored to a five-point Likert scale ("strongly disagree" to "strongly agree").

Demographics. Surveyed demographics included the following seven variables: gender, college boundedness, ethnicity, participants' place of birth, mother's place of birth, father's place of birth, and socioeconomic status (SES). Gender (male/female), college boundedness (yes/no), and participants' place of birth (in the US or not) were treated as binary variables. Ethnicity, mother's and father's place of birth were treated as categorical variables. Three options (i.e., born in the US or not, unknown) were available for mother's and father's place of birth. The following seven options were available for ethnicity:

Asian or Asian American, including Chinese, Indian, Korean, and others

Black or African American

Hispanic or Latino, including Mexican American, Colombian, Nicaraguan, and others

White, Caucasian, European American; not Hispanic

American Indian/Native American

Mixed; Parents are from different groups

Other

Participants' SES was quantified in a continuous variable. The participants were asked to report their relative position based on their family's assets, educational background, and occupations. Their answers were anchored to a ten-point Likert scale (first = lowest SES to tenth decile = highest SES).

Analysis

Notes on elastic net. In the case of regression based on ordinary least squares (OLS), the ultimate goal is to estimate coefficients that minimize the residual sum of squares (RSS). Let us briefly review the process of ordinary linear regression. If we have *n* predictors, $x_1 \ldots x_n$ in our prediction model, then the predicted response, $\hat{y}$ becomes,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n$$

when $\hat{\beta}_0 \ldots \hat{\beta}_n$ are the estimated coefficients. Then, we can define the vector of coefficients, $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_n)$. The RSS can be calculated as follows:

$$RSS = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

when *N* is the sample size, $\hat{y}_i$ is the predicted response and y_i the actual response value at the i_{th} observation. In ordinary linear regression based on OLS, coefficients are estimated to minimize the RSS. However, doing so would not be an ideal method when a researcher is particularly interested in model exploration and thus intends to search for the simplest model with acceptable predictability. Also, the ordinary method that uses the whole dataset to estimate

coefficients is likely to cause the issue of overfitting, so the estimated prediction model based on this method might not be ideal to predict and explain the reality beyond the boundary of the collected data.

To address these issues, the elastic net-applied regularization has been utilized in model exploration and coefficient estimation (Zou & Hastie, 2005). To enable the exploration of the simplest prediction model, elastic net regularization tends to penalize the inclusion of unnecessary coefficients. While ordinary regression tends to minimize RSS, elastic net regularization tends to minimize RSS plus two penalty terms. When we have two non-negative terms, α and λ,

$$L(\lambda, \beta) = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 + \lambda[(1-\alpha)|\beta|^2 + \alpha|\beta|_1]$$

$|\beta|_1$ and $|\beta|^2$ are two penalized terms, l_1 and l_2 norms, respectively, and defined as

$$|\beta|_1 = \sum_{j=1}^{n} |\beta_n|$$

and

$$|\beta|^2 = \sum_{j=1}^{n} {\beta_n}^2$$

Due to the presence of these two penalization terms, L becomes relatively greater when more predictors are included in a model while the RSS remains the same. Thus, when the predictability of two different models are the same (RSSs are equal) elastic net regularization prefers the simpler model to minimize L with l_1 and l_2 norms.

There are two specific forms of the regularization (Zou & Hastie, 2005). When only the l_2 norm is included in L, in other words, when $\alpha = 0$, it becomes ridge regression. Although ridge

regression also penalizes unnecessary predictors and coefficients it cannot completely exclude such predictors and coefficients from a model because the l_2 norm is the quadratic term of coefficients, and when Lagrange multiplier is applied to find the minimal L value, β_n cannot reach zero. Thus, ridge regression might not be ideal for model exploration although it tends to shrink coefficients with the l_2 norm.

On the other hand, when $\alpha = 1$, only the l_1 norm is included in L. It is least absolute shrinkage and selection operator (LASSO). Unlike ridge regression with the l_2 norm, it is possible to completely exclude certain predictors and coefficients with LASSO. In general, for model exploration, LASSO would be more suitable compared with ridge regression. However, when predictors are significantly correlated with each other, LASSO does not perform desirably. With highly correlated predictors, LASSO could not produce stable outcomes when LASSO is performed multiple times; in other words, estimated coefficients can change significantly across trials (Grave, Obozinski, & Bach, 2011). Moreover, LASSO tends to select only one predictor while neglecting other predictors that might also be significantly associated with the dependent variable (Zou & Hastie, 2005). Given that civic purpose-related variables that were used in the present study as candidate predictors are significantly correlated with each other, LASSO would not be the ideal exploration method like the case of ridge regression.

When $0 < \alpha < 1$, both l_1 and l_2 norms are included in L. It becomes the general form of elastic net. This method can overcome the limitations of the aforementioned two methods by employing both l_1 and l_2 norms. According to previous statistical studies, elastic net is likely to produce the best outcome particularly when predictors show significant correlation with each other. Hence, we used elastic net with $\alpha = .5$, which is the midpoint between ridge regression and LASSO, as used in Zou and Hastie (2005).

Notes on glmnet. We used an R package, *glmnet*, to implement elastic net in our study. We performed elastic net regression for each of three dependent variables, Wave 2 political purpose, community service purpose, and expressive activity purpose statuses. All input variables, independent variables or candidate predictors, were converted into a 2D matrix. Because we tested three different models for three dependent variables, three sets of 2D matrices containing predictors were created.

Elastic net regression was performed 5,000 times with cross-validation to minimize the possibility of overfitting. The whole dataset was randomly separated into ten folds. For each trial, *glmnet* randomly shuffled the whole dataset and used 90% of them for the estimation of coefficients (learning dataset) and 10% of them for accuracy testing (validation dataset for cross-validation). This method was used to examine whether the regression model can explain the rest of the data that was not used for estimation and was robust against overfitting. Given that both the learning and validation datasets were randomly selected for each trial, we decided to repeat the overall process 5,000 times to better examine the overall trend. If we rely on the result from only one trial, it would be possible that the resultant estimated regression model was estimated by chance.

For each trial, one block of the separated datasets (90% of the whole dataset) was used to estimate coefficients and then the prediction accuracy of the prediction model with the estimated coefficients was evaluated with another data block (10% of the whole dataset). Computationally, for each trial, different sets of coefficient values were estimated while changing λ. According to *glmnet* manual, there are two options to select λ to identify the best prediction model. First, when the minimum λ value among all tested λ, λ_{min}, is employed, the prediction model produces the smallest *L* and highest prediction accuracy. Second, when λ_{1se}, is employed, the prediction model

becomes simpler with less predictors and more regularized coefficients while not significantly compromising the prediction accuracy; λ_{1se} gives the most regularized model with smaller and less coefficients while maintaining the regression error within one standard error of the possible minimum error (Hastie & Qian, 2014). We used λ_{1se} to estimate the simplest model with the reasonable prediction accuracy and regularized coefficients. As a result, we estimated coefficients with λ_{1se} used in *L*.

We used the multinomial logistic regression option that was implemented in *glmnet*. With this option, *glmnet* estimated coefficients for each Wave 2 civic purpose status. For instance, if the estimated coefficient of Wave 1 political drifting in the prediction of Wave 2 political drifting is greater than zero, being a drifting participant at Wave 1 positively contributes to being a drifting participant at Wave 2. At the end of each trial, estimated coefficients were stored in a list of matrices. After completing 5,000 trials, first, we examined which predictors were estimated to have non-zero coefficients in at least one trial. We filtered out predictors that were not selected by *glmnet* at least once out of 5,000 trials. Second, we calculated the mean of the coefficient of each survived predictor.

Notes on validation dataset prediction accuracy test. We examined the validation dataset accuracy between multinomial elastic-net regression and standard multinomial regression. During this test process, we tested to what extent the regression model generated by each method could accurately predict outcomes out of the boundary of the data used for regression. For this test, before performing regression, we randomly separated the whole dataset into a training dataset (50%) and validation dataset (50%). Then, only the half of the dataset, the training dataset, was used to generate a regression model with each regression method; the validation dataset was not used during this phase.

Once a regression model was generated, we entered the validation dataset to the model to predict categorical outcomes beyond the boundary of data used for regression. Then, we compared the predicted outcomes and the actual outcomes, which existed in the validation dataset, to calculate a prediction accuracy. For instance, if a validation dataset size was 100, and predicted categorical outcomes were identical to actual outcomes in 50 cases, then a calculated validation dataset accuracy became 50%. Similar to the aforementioned elastic-net regression process to examine coefficients, we repeated this accuracy test procedure 5,000 times. Because the dataset was randomly shuffled to create the training and validation dataset, we decided to repeat the same test procedure multiple times to address any random error emerging from randomization.

Results

Descriptive statistics

First, brief descriptive statistics of analyzed variables at both Waves are presented in Table S2.

Table S2

Descriptive statistics of variables

	n	*M*	*SD*	Median	Skewness	Kurtosis
Wave 1						
Civic intent	1,546	3.56	.75	3.60	-.50	.35
Political engagement	1,503	1.61	.60	1.50	1.10	.85
Community service engagement	1,432	2.65	.78	2.60	-.14	-.73
Expressive activity engagement	1,431	1.79	.71	1.75	.80	.06
Moral identity	1,510	3.20	.58	3.20	-.81	.65

Political identity	1,507	2.16	.73	2.00	.47	-.24
Parental involvement	1,464	2.62	.88	2.75	-.05	-.37
Peer involvement	1,463	3.00	.93	3.00	-.31	-.23
Peer civic academic value	1,451	3.75	.82	3.67	-.55	.12
School opportunities	1,562	3.59	.73	3.67	-.72	1.36
Wave 2						
Civic intent	462	3.61	.73	3.60	-.59	.94
Political engagement	468	1.61	.61	1.42	1.26	1.52
Community service engagement	453	2.30	.75	2.20	.27	-.76
Expressive activity engagement	457	1.64	.66	1.50	1.02	.49

Elastic net regression

Tables S3, 4, and 5 show mean estimated non-zero coefficients in the prediction models for Wave 2 political, community service, and expressive activity purpose, respectively. The non-zero coefficients were coefficients of predictors that survived *glmnet* process at least once out of 5,000 trials. We calculated the mean coefficient of each survived predictor across 5,000 trials. The numbers were not rounded to provide readers with additional information.

Table S3

Mean estimated non-zero coefficients in the prediction model of Wave 2 political purpose

	Wave 1 drifting	Wave 1 dreamers	Wave 1 dabblers	Wave 1 purposeful	Moral identity	Political identity	Parental involvement	Peer involvement	Peer values	School opportunities	Asian
Wave 2 Political											
Drifting	0.5361199	-0.09320287	-0.04429729	-0.32210065	-0.070472127	-0.294960758	-4.26E-04	-0.019749579	-0.028828963	-0.011558982	4.08E-06
Dreamers	-0.168418	0.29902541	-0.0388182	-0.09771547	0.015945394	0.052330837	-2.49E-04	-0.005134788	-0.001972543	0.003650236	-1.71E-06
Dabblers	-0.1477021	-0.09408749	0.11414063	-0.07937907	-0.005764673	0.000714779	9.41E-05	0.009359266	0.010648965	0.001093667	2.33E-07
Self-oriented	-0.1071924	-0.09761362	-0.01917327	0.27860787	0.019958272	0.087621636	2.89E-04	0.009395275	0.012086239	0.004578278	-1.20E-06
Purposeful	-0.1128074	-0.01412143	-0.01185187	0.22058733	0.040333134	0.154293507	2.92E-04	0.006129827	0.008066302	0.0022368	-1.40E-06

Table S4

Mean estimated non-zero coefficients in the prediction model of Wave 2 community service purpose

	Wave 1 drifting	Wave 1 dreamers	Wave 1 dabblers	Wave 1 self-oriented	Wave 1 purposeful	Moral identity	Political identity	Peer involvement	Peer values
Wave 2 community service									
Drifting	0.412435892	-1.88E-04	7.69E-07	-0.2864454	-0.121910598	-0.07076972	-0.22982666	-0.00492147	-0.07465079
Dreamers	-0.102965943	4.80E-04	-1.07E-06	0.02641362	0.015970055	0.02448244	0.13312716	-0.001171881	-0.03936328
Dabblers	-0.008459596	-1.52E-04	2.19E-07	0.04962175	-0.006178787	-0.01794138	-0.04525057	0.002022419	0.03008039
Self-oriented	-0.064864862	5.70E-05	-1.73E-07	-0.02403156	0.046183979	0.0195827	0.04898148	0.000298754	0.01585104
Purposeful	-0.236145491	-1.97E-04	2.53E-07	0.23444159	0.065935351	0.04464597	0.09296859	0.003772179	0.06808264
	School opportunities	Asian	Latino	White	US born	Mother US born	Mother out of US born	College	SES
Wave 2 community service									
Drifting	-0.007283517	0.000210157	-2.93E-07	1.94E-04	3.34E-05	-1.74E-04	3.75E-06	0.004576003	-1.85E-07
Dreamers	0.001128071	-0.000594513	3.03E-07	-3.13E-04	-3.33E-06	-1.32E-04	2.16E-06	-0.001257366	1.39E-06
Dabblers	0.000485694	0.000570056	-6.28E-08	-1.10E-04	-5.96E-06	1.44E-04	-3.17E-06	-0.000919414	7.76E-07
Self-oriented	-0.000264379	-0.000294998	-4.00E-08	1.84E-04	-4.90E-06	1.17E-04	-2.31E-06	0.000242924	-7.18E-07
Purposeful	0.005934131	0.000109298	9.27E-08	4.48E-05	-1.92E-05	4.49E-05	-4.30E-07	-0.002642146	-1.26E-06

Table S5

Mean estimated non-zero coefficients in the prediction model of Wave 2 expressive activity purpose

	Wave 1 drifting	Wave 1 dreamers	Wave 1 dabblers	Wave 1 self-oriented	Wave 1 purposeful	Moral identity	Political identity	Parental involvement	Peer values	School opportunities	Gender	Asian	Black	Latino
Wave 2 expressive activity														
Drifting	0.283657739	-0.12229773	-0.004655493	-8.76E-06	-0.3684759	-0.07218855	-0.28745577	-5.07E-04	-1.68E-04	-0.020306904	-9.91E-06	0.062315299	-0.005209862	-3.89E-04
Dreamers	-0.11001671	0.18295592	-0.098050535	1.29E-05	-0.1643814	0.02451741	0.10667628	-5.13E-05	8.42E-05	0.01286859	6.11E-06	-0.014062271	0.001360277	1.28E-04
Dabblers	-0.008093983	-0.12510251	0.095128397	6.32E-06	-0.14414	-0.01589725	-0.02994681	1.40E-04	5.20E-06	-0.003068088	1.35E-06	-0.030453593	0.003157812	2.25E-04
Self-oriented	-0.080660089	0.05250083	-0.011943748	-2.41E-06	0.2447343	0.02987162	0.06507638	1.21E-04	3.78E-05	0.009297961	-2.10E-06	-0.012127291	0.001886268	-2.26E-05
Purposeful	-0.084886956	0.01194349	0.019521378	-8.01E-06	0.4322629	0.03369677	0.14564992	2.97E-04	4.09E-05	0.001208442	4.55E-06	-0.005672144	-0.001194494	5.87E-05

Validation Dataset Prediction Accuracy Test

Table S6 reports the results from validation dataset prediction based on elastic-net regression and standard regression across 5,000 repetitions. In all civic purpose domains, regression models generated by elastic-net regression better predicted outcomes in the realm of the validation dataset compared with regression models generated by standard regression. Both frequentist and Bayesian *t*-tests indicated that the differences in prediction accuracy indicators were significant. The resultant Cohen's *D*s, indicators for effect sizes, also suggested that the differences were practically significant.

Table S6

Validation dataset prediction accuracy comparison between elastic-net regression and standard regression

	Elastic-net regression		Standard regression		*t*	*df*	*p*	Bayes Factor	Cohen's *D*
	M	*SD*	*M*	*SD*					
Political	52.77%	.03%	44.89%	.10%	54.39	4,947	.000	2.68e+501	-.77
Community Service	54.56%	.03%	49.56%	.06%	55.54	4,999	.000	3.92e+519	-.79
Expressive Activity	49.96%	.03%	39.01%	.12%	61.79	4,999	.000	7.56e+613	-.87

Supplementary References

Grave, E., Obozinski, G. R., & Bach, F. R. (2011). Trace Lasso: a trace norm regularization for correlated designs (PDF Download Available). *Advances in Neural Information Processing Systems*, *24*, 2187–2195.

Han, H., Ballard, P. J., & Choi, Y.-J. (2019). Links between moral identity and political purpose during emerging adulthood. *Journal of Moral Education*.

Hastie, T., & Qian, J. (2014). Glmnet Vignette. Retrieved from https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*, 301–320. doi:10.1111/j.1467-9868.2005.00503.x