

Appendix C: Variable contributions from the item-level predictive models

In Figure 6, the variable contributions are depicted for the models fitted using the RIASEC item scores as predictors. The plots show a similar, but more nuanced view, compared to the variable contributions in the models fitted using the subscale scores (Figure 3, main paper). Again, the Social preferences scale contains the strongest predictors of university major completed, followed by Realistic subscale, followed by Enterprising, Investigative and Conventional preferences scales. The item-level variable contributions do provide a more nuanced view. For example, in the analyses using the item scores as predictors, the Investigative subscale does seem to contribute more strongly compared to the analyses based on the subscales.

All methods found items S5 (“I would like to help people with family-related problems”) and S3 (“I would like to help people who have problems with drugs or alcohol”) and to contribute most. For the penalized GLM, the GAM and the boosted ensemble, this was followed by I2 (“I would like to study animal behavior”). For the single tree, the third most important predictor was item R6 (“I would like to fix a broken faucet”). For the prediction rule ensemble and the random forest, the third and fourth most important predictors were items R3 (“I would like to work on an offshore oil-drilling rig”) and E5 (“I would like to manage a department within a large company”).

Figure 6

Variable contributions for each of the models fitted using RIASEC item scores as predictors

