Using AI to Quantify Personal Qualities from College Applications

Lila DiMasi

University of Pennsylvania, COL'2025

Mentor: Dr. Angela L. Duckworth, College of Arts & Sciences – Department of Psychology; Wharton – Operations, Information & Decisions Department

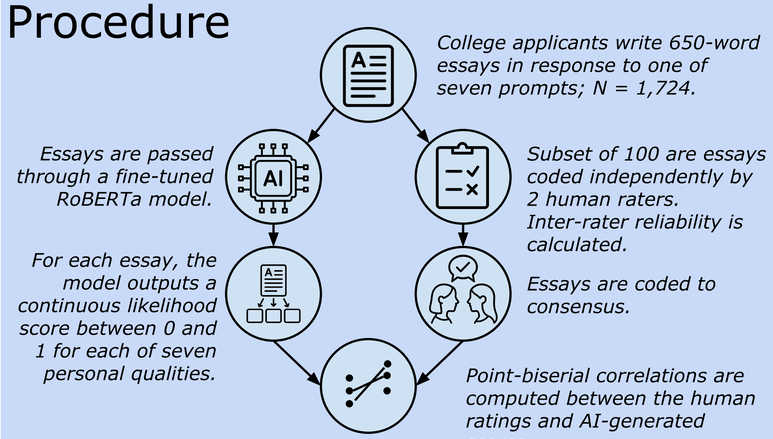
Introduction

- **In college admissions, grades and test scores don't tell the whole story.** Holistic admissions processes aim to evaluate the "whole candidate," including personal qualities like perseverance, leadership, and purpose.
- **Personal qualities** are conceptually distinct from cognitive ability, predictive of long-term success, and potentially responsive to intervention.[1]
- **Existing measures aren't scalable.** Self-report surveys and teacher ratings are vulnerable to faking, while performance tasks and biodata are time-intensive and inconsistent across applicants.[1,2]
- **AI presents an opportunity** to identify personal qualities from application text.
- **Lira et al. (2023) demonstrated proof of concept.** RoBERTa models[4] trained on short 150-word essays scored seven personal qualities and modestly predicted six-year college graduation (AUC = 0.557–0.568), independent of demographics and test scores.

Aims

1. Test generalizability: Evaluate whether a RoBERTa model trained on 150-word essays about extracurriculars[3] can validly assess personal qualities in full-length college essays.
2. Examine extent to which AI scores relate to demographic variables and academic metrics (SAT, GPA).

Procedure



Results

Face validity: The model's top words make intuitive sense: *president* predicts leadership, *collaborating* drives teamwork, and *learned* signals learning.

Figure 1. *Complete or partial words on which RoBERTa models relied most for generating personal quality scores. Font size is proportional to word importance. Dark blue words are more common.*

AI-predicted personal qualities aligned with human ratings and showed minimal association with demographics or academic metrics.

	Human ratings						
Personal quality	**PP**	**LD**	**TW**	**LR**	**PS**	**IM**	**GP**
AI-generated likelihoods							
1. Prosocial Purpose (PP)	**0.62**	0.12	-0.11	-0.11	-0.04	-0.01	0.07
2. Leadership (LD)	0.24	**0.51**	**0.32**	-0.06	0.09	-0.17	0.12
3. Teamwork (TW)	0.01	**0.27**	**0.73**	0.13	-0.06	-0.02	0.08
4. Learning (LR)	0.08	-0.03	0.13	**0.47**	0.23	0.05	0.09
5. Perseverance (PS)	-0.04	0.05	-0.05	0.05	**0.59**	-0.18	0.14
6. Intrinsic Motivation (IM)	-0.07	0.01	0.13	-0.02	-0.17	**0.61**	0.06
7. Goal Pursuit (GP)	**0.3**	0.1	0.12	-0.09	0.23	0.07	**0.61**
Descriptive statistics							
Human inter-rater reliability (α)	0.34	0.64	0.35	0.33	0.66	0.57	0.52
Frequency of human rating	40%	11%	11%	56%	58%	49%	53%
Mean AI-generated likelihood (SD)	0.33 (0.42)	0.08 (0.23)	0.14 (0.3)	0.69 (0.39)	0.71 (0.34)	0.54 (0.43)	0.72 (0.36)

bold = $p < 0.05$, BH-corrected

Table 1. *Correlations between human ratings and AI-generated likelihoods of personal qualities. Inter-rater reliability measured with Krippendorf's α.*

AI scores align closely with human judgments when predicting the same personal qualities. Moderate to strong correlations between human ratings and computer-generated likelihoods (mean r = 0.60).

Both the AI and human raters distinguish between conceptually distinct personal qualities. Correlations between human ratings and AI scores for *different* traits are near zero (mean $r = 0.005$).

AI-generated scores were largely uncorrelated with demographics and academic metrics.

After BH correction, only two significant relationships remained: male students received slightly lower scores on prosocial purpose compared to female students ($r = -0.11$, *adjusted* $p < .05$), and first-generation students received slightly lower scores on leadership (r = –0.074, *adjusted* p < .05). All other correlations were small ($|r| < 0.08$) and non-significant.

There were no significant correlations between AI-generated personal quality scores and either SAT scores or GPA (all |r| < 0.09)

Discussion

Implications

Scalable, interpretable, and privacy-preserving. Unlike generative models like ChatGPT, RoBERTa models can be run offline on local machines and provide transparent, word-level explanations – practical for sensitive contexts like college admissions.

Limitations & Future Directions

- No predictive outcomes. Future work could examine whether these scores predict enrollment, graduation, or performance in college.
- Only 27% of students reported SAT or ACT scores, limiting analyses of AI scores and academic metrics.
- All essays came from students in National Education Equity Lab courses—a highly motivated, possibly non-representative group.
- Models were not trained on full-length essays. Future work could fine-tune new models on target text sources.

Acknowledgements

This work was supported in part by a Millstein Family Undergraduate Research Grant. I extend my deepest gratitude to Angela L. Duckworth, Benjamin Lira, and Maya Brown-Hunt for their contributions to this project, as well as our partners at Common App, National Education Equity Lab, and Mathematica.

References

1. Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. Educational Researcher (Washington, D.C. : 1972), 44(4), 237–251. https://doi.org/10.3102/0013189X15584327
2. Kuncel, N., Tran, K., & Zhang, S. H. (Charlene). (2020). Measuring Student Character: Modernizing Predictors of Academic Success. In C. Wendler & M. E. Oliveri (Eds.), Higher Education Admissions Practices: An International Perspective (pp. 276–302). Cambridge University Press. https://doi.org/10.1017/9781108559607.016
3. Lira, B., Gardner, M., Quirk, A., Stone, C., Rao, A., Ungar, L., Hutt, S., Hickman, L., D'Mello, S. K., & Duckworth, A. L. (2023). Using artificial intelligence to assess personal qualities in college admissions. Science Advances, 9(41), eadg9405. https://doi.org/10.1126/sciadv.adg9405
4. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach (No. arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692