The Effect of Frustration on Students’ Assessment Scores in a Digital Learning Platform

Iris Zhong1

1 Stanford University

The Effect of Frustration on Students’ Assessment Scores in a Digital Learning Platform

# Hypothesis

The current dataset comes from a free online tutoring platform – ASSISTments (Feng, Heffernan, & Koedinger, 2009). It is called a “tutoring” system, because if students get a question wrong, they are provided with a small “tutoring” session where they must answer a few questions that break the problem down into steps. The students in this dataset were all middle school students.

ASSISTments also includes sensor-free affect detectors that estimates a student’s likelihood of feeling frustrated, bored, confused, or concentrated (Wang, Heffernan, & Heffernan, 2015). In the previous homework, I explored whether a student’s overall frustration when answering questions in ASSISTments influenced their variance in monthly average assessment score. A side discovery was that students with higher frustration score on average did better in assessments. Therefore, in the current paper, I want to explore how students’ monthly frustration score affected their scores. In particular, the related two research questions are:

1. How does frustration explain between-person variances in scores?

I hypothesize that invidiuals with more frustration scored higher in assessments than those with lower frustration, if the finding from the previous project holds true.

1. How does frustration explain within-person fluctuation in monthly assessment scores?

If frustration stimulates students to perform better in assessments, I hypothesize that in months when frustration score are higher, students’ scores would be higher than their own scores when frustration is lower.

The first research question addresses Objective #5 – determinants of individual differences in intraindividual change. The second research question addresses Objective #4 – determinants of intraindividual change.

# Data analysis

The current paper explores two hypotheses: 1) students with higher overall frustration (trait frustration) tend to score higher in assessments than students with lower frustration; 2) for an individual student, their scores are higher when they are more frustrated (state frustration). Intraindividual covariance model from the multilevel modeling framework is applied, since it can address both between-person variances explained by trait and within-person variances explained by state (Grimm, Ram, & Estabrook, 2017).

First, a simple model with fewer interactions and random effects is attempted. It mainly consists of four parts: tells a person-specific intercept of scores, is a person-specific slope, is the effect of within-person changes in frustration on scores, and is the error term.

, where

,

,

.

describes the estimated score from a prototypical student with average level of frustration overall. indicates the change in assessment score if a student’s trait frustration (i.e. average frustration over time for a person) increases by one point, and it addresses the first research question about the effect of trait frustration on between-person variations in scores. tells a prototypical person’s change of assessment scores over time, and shows the change in the slope if a person’s trait frustration increases by one. The effect of within-person changes in frustration on scores is captured by , which is the target parameter for the second research question. Finally, random effects and are assumed to be multivariate normal.

A full model is then applied. Compared to the first model, it adds – differences in the within-person association between state frustration and scores controlled by trait frustration, and – random effect on person-specific slope.

, where

,

,

.

To make variables at roughly the same scale, I multiply all frustration and score-related fields by 100. The trait frustration predictor is centered by subtracting the mean from the average frustration score for each person. State frustration is derived from subtracting the person-level mean from each frustration score.

The models are run in R (R Core Team, 2022) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). They are fit by the restricted maximum likelihood method, with missing values excluded from analysis. Statistical significance is measured by inspecting whether 0 is within 95% confidence interval.

The entire analysis is supported by the following software and packages: R (Version 4.1.3; R Core Team, 2022) and the R-packages *data.table* (Version 1.14.2; Dowle & Srinivasan, 2021), *dplyr* (Version 1.0.8; Wickham, François, Henry, & Müller, 2022), *ggplot2* (Version 3.3.5; Wickham, 2016), *lme4* (Version 1.1.28; Bates et al., 2015), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *papaja* (Version 0.1.1; Aust & Barth, 2022), *tidyr* (Version 1.2.0; Wickham & Girlich, 2022), and *tidyverse* (Version 1.3.1; Wickham et al., 2019).

# Results

Results from the intraindividual covariance models to estimate the effect of frustration on between-person and within-person variations in assessment scores are shown in the table below.

Table 1: Intraindividual covariance model 1.

| Term |  | 95% CI |  |
| --- | --- | --- | --- |
| Intercept () | 72.44 | [71.96, 72.91] | 296.76 |
| Trait frustration () | 2.50 | [1.96, 3.04] | 9.10 |
| Month ( | -0.23 | [-0.28, -0.18] | -8.63 |
| Trait frustration Month () | 0.08 | [0.02, 0.14] | 2.52 |
| State frustration () | 0.54 | [0.44, 0.64] | 10.40 |

In the simple model (see Table 1), a prototypical learner with a mean level of trait frustration is estimated to score 72.44% of the points at the start of the school year ( = 72.44, t = 296.76, 95%CI = [71.96, 72.91]). If a student’s trait frustration is one point above the mean, their score is predicted to be 2.5 points higher ( = 2.50, t = 9.10, 95%CI = [1.96, 3.04]), and the result is consistent with the first hypothesis that students with a higher level of overall frustration tend to score higher in assessments. A prototypical student’s scores is estimated to decrease by 0.23 points every month ( = -0.23, t = -8.63, 95%CI = [-0.28, -0.18]). The interaction between month and trait frustration has a coefficient of 0.08, suggesting that for a student with a trait frustration one point higher than average, their slope is predicted to be 0.08 greater than a prototypical student’s slope ( = 0.08, t = 2.52, 95%CI = [0.02, 0.14]). Finally, regarding within-person fluctuations in frustration, if a student’s state frustration at one timepoint is greater than their trait frustration by one point, their score is estimated to be 0.54 higher than their average score ( = 0.54, t = 10.40, 95%CI = [0.44, 0.64]). This finding is in line with the second hypothesis that for each individual, in months when frustration score are higher, their scores are higher. All of the predictors in the model are statistically significant, according to their confidence intervals.

For random effects, the variance of the intercept is 107.12, indicating large between-person differences in assessment scores. The variance of state frustration is 1.429, suggesting small between-person differences in the within-person association between state frustration and scores. The correlation between the two is -0.37. Thus, a person with high scores at the start are more likely to have a more negative association between frustration and score.

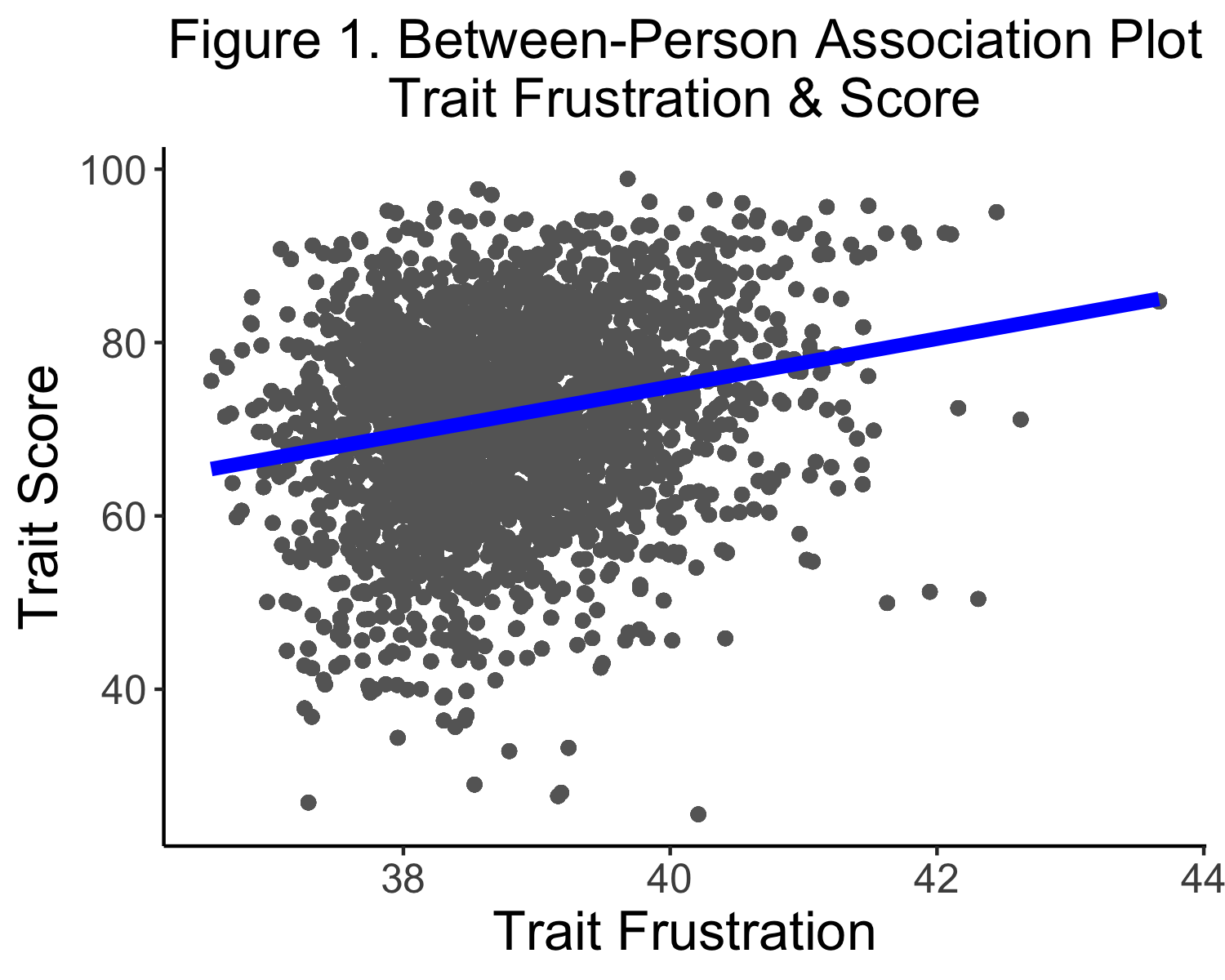


Figure 1 demonstrates the relationship between a person’s trait frustration with their average score. Same as the finding from the model parameters, students with higher trait frustration are more likely to have a higher mean score.

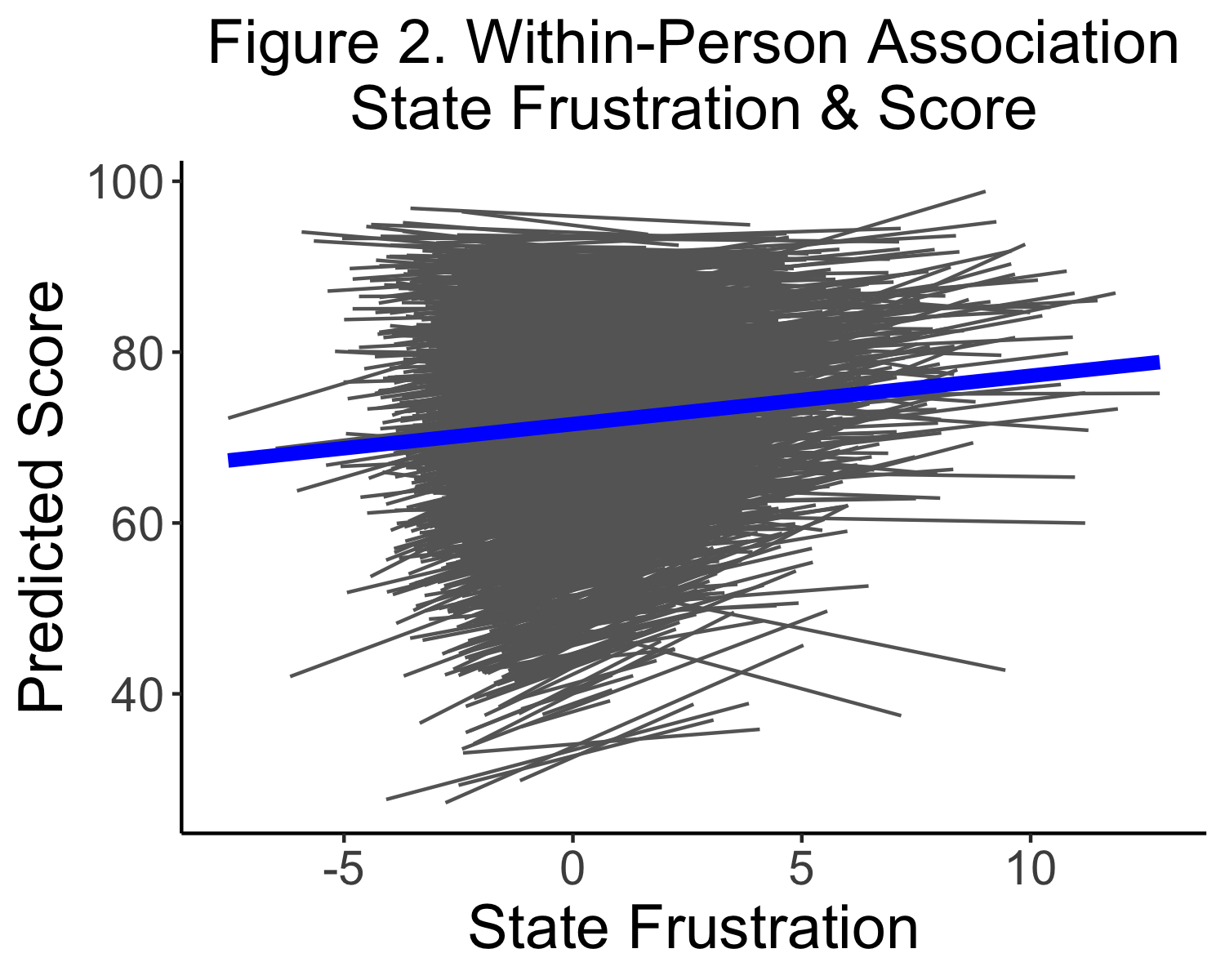


Figure 2 shows the association between state frustration (centered) and predicted assessment score form the model. Each line represents the predicted score for a student. Most lines in the plot have a positive slope, demonstrating the positive association between within-person state frustration and score.

Model 2 fits the full model, with the interaction between trait frustration and state frustration, and a random effect on month added. This allows the within-person association between state frustration and score to depend on trait frustration, and person-specific slopes to vary outside of the effect of frustration. However, this model fails to converge (max|grad| = 0.008), so the current paper does not report the results from Model 2.

# References

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of ‘data.frame‘*. Retrieved from <https://CRAN.R-project.org/package=data.table>

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243–266. <https://doi.org/10.1007/s11257-009-9063-7>

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. New York, NY: Guilford Press.

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Wang, Y., Heffernan, N. T., & Heffernan, C. (2015, March 16). *Towards better affect detectors: Effect of missing skills, class features and common wrong answers*. 3135. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2723576.2723618>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>