Leveraging Trace Data in Game-Based Assessments

Elena M. Auer, Sebastian Marin, and Richard N. Landers

University of Minnesota, Twin Cities

Andrew Collmus

Flex

Michael B. Armstrong

Google

Salih Mujcic and Jason A. Blaik

Revelian, Inc.

University of Minnesota

Driven to Discover®

Agenda

1. Trace data in Game-based Assessments
2. Predicting *g* with GBA game scores (prior work)
3. Predicting *g* with GBA trace data (present study)

Trace Data & Game-based Assessments

- GBAs produce two types of data reflecting in-game behaviors as indicators of KSAOs
  1. **Game Score:** resulting from planned measurement approach
  2. **Trace Data:** micro-behavioral data points typically used for system monitoring and debugging

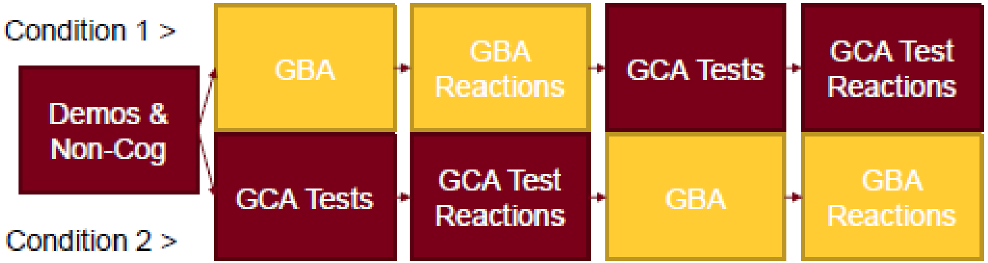
Predicting *g* with Game Scores

Revelian's Cognify

- GBA with a series of mini games designed to target specific cognitive abilities
- Game score composite for general intelligence

Predicting *g* with Game Scores

- Previous validation study supports this link (Landers, Armstrong, Collmus, Mujcic, & Blaik, 2017)
  - 530 undergraduate students in semi-controlled environment



- Improved motivational and attitudinal reactions for GBA (*d*s ranging from .106 - .814)
- Construct validity evidence: *latent g* $\longrightarrow$ *latent game performance* (*r* = .968)

Predicting *g* with Trace Data

- Is trace data untapped potential for improving construct measurement and predicting criteria?
  - Greater range of behaviors
  - But, log files can be big, messy, and tricky to derive meaning from
  - Requires data science techniques: feature engineering & modern prediction methods

Feature Engineering

- Raw Trace data (from previous study):
  - 3 GB
  - ~ 13 million rows (~26 k/participant)
  - Structured by event & time stamp

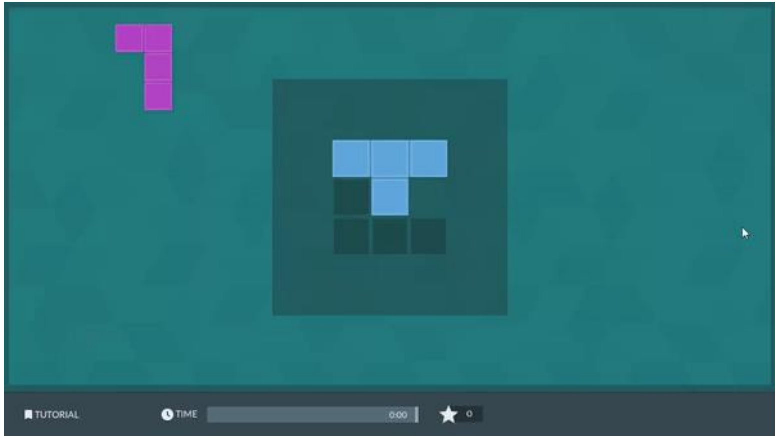
sequence_number	event_time	event_type	event_content
1195	2017-02-13 15:58:57.8+00	DestroyObject	{""id"":""56dad974-d5ef-4940-84b6-f3b1dca8
1196	2017-02-13 15:58:57.957+00	MousePosition	{""y"":127,""context"":null,""x"":722}
1197	2017-02-13 15:58:58.304+00	StartRound	{""roundNumber"":35,""round"":""6b8012c9-24
1198	2017-02-13 15:58:58.372+00	MousePosition	{""y"":127,""context"":null,""x"":722}
1199	2017-02-13 15:58:58.781+00	MousePosition	{""y"":127,""context"":null,""x"":722}
1200	2017-02-13 15:58:59.182+00	MousePosition	{""y"":127,""context"":null,""x"":722}
1201	2017-02-13 15:58:59.38+00	DestroyObject	{""id"":""5e66cef2-550a-4098-b217-3bb6aa8e
1202	2017-02-13 15:58:59.38+00	CreateObject	{""id"":""d8a6bcaa-1fcf-41fc-96ac-6b9b54fd8

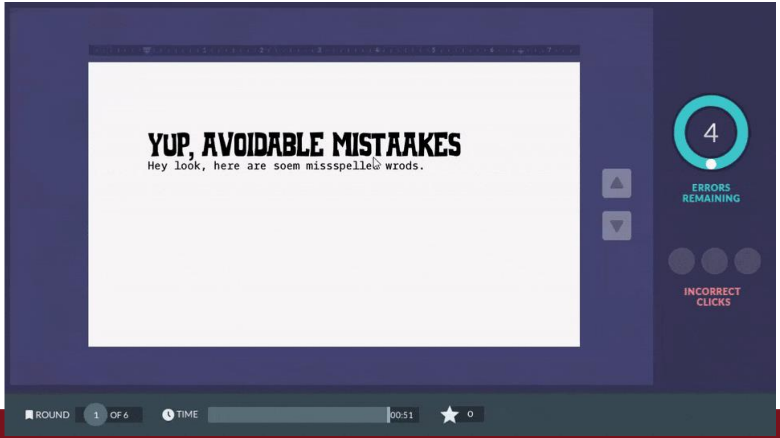
Feature Engineering & Selection

- Top-down (theory-driven) and bottom-up (data mining and machine learning) approach
  – SMEs examined log dataset
  – Machine learning models with feature selection

Feature Engineering

- Engineered Features: 65
- A few examples:
  - Time spent/game or /round
  - Average # of clicks/game
  - # of rotations (game specific)
  - # of words selected (game specific)





Predicting *g*: Modeling Approach

- Use-cases for modern prediction methods (e.g., machine learning) in social science (Putka, Beatty & Reeder, 2017)
  - lack of comprehensive theory
  - a need to balance model complexity with parsimony
  - predictors occur on a variety of measurement scales
  - high degree of uncertainty in the model selection process
- Working with trace data involves all of these

Predicting *g*: Modeling Approach

- R's Caret Machine Learning Package (Kuhn, 2014)
  - k-fold cross-validation (for tuning model parameters)
  - Training & holdout split (to calculate validity estimates)
- Models: *Putka, Beatty & Reeder (2018) provides an overview of each of these models

Elastic Net	Random Forest	Gradient Boosted Trees	Support Vector Machines
Regularized regression (ridge & lasso)	Regression trees with bootstrapped sampling & random subset of predictors	Regression trees with a forward step-wise-like regression component	Robust regression variant, focused on predicting difficult to predict cases

Findings

Predicting g (from traditional measures) using GBA **trace data**

Model	Validity
Elastic Net	0.605
Random Forest	0.693
Gradient Boosted Trees	0.683
Support Vector Machines	0.630

Note: Validity refers to the correlation between the predicted g score from each machine learning model and the composite g score from the holdout sample.

Findings

Predicting g (from traditional measures) using GBA **composite game score**

Model	Validity
OLS Regression	0.633

Note: Validity refers to the correlation between the predicted g score from OLS Regression model and the composite g score from the holdout sample.

Findings

Predicting g using **GBA trace data**

Model	Top "Important" Variables
Elastic Net	• number of Timeout Rounds in Short Circuit game • **time spent on make a splash and short circuit games** • missed choices in quick comparison game
Random Forest	• **time spent on short circuit game** • total time spent per round on balloon blast game • **time spent on make a splash game**
Gradient Boosted Trees	• **total time spent per round on balloon blast game** • **time spent on make a splash and short circuit games** • Time spent on quick comparison game tutorial
Support Vector Machines	• **time spent on short circuit game** • time spent on short circuit tutorial • **total time spent per round on balloon blast game**

Findings

Predicting GPA: incremental validity of trace data (predicted g) over game score composite

Elastic Net

Regressions of GPA on Comparisons of Hierarchically Nested Regression Models

					Comparison to Model 3			
Model	R^2	F	df	p	ΔR^2	F	df	p
Model 1: Game Score Composite	.027	13.98	1, 509	< .001	0.00	0.02	1,509	.891
Model 2: Trace Data Score	.018	9.27	1,509	.002	0.01	4.52	1,509	.032
Model 3: Model 1 + Model 2	.027	6.986	2,508	.001				

Findings

Predicting GPA: incremental validity of trace data (predicted g) over game score composite

Random Forests

Regressions of GPA on Comparisons of Hierarchically Nested Regression Models

					Comparison to Model 3			
Model	R^2	*F*	*df*	*p*	ΔR^2	*F*	*df*	*p*
Model 1: Game Score Composite	.027	13.98	1, 509	< .001	0.01	6.63	1,509	.010
Model 2: Trace Data Score	.039	20.59	1,509	< .001	0.00	0.20	1,509	.653
Model 3: Model 1 + Model 2	.039	10.38	2,508	< .001				

Findings

Predicting GPA: incremental validity of trace data (predicted g) over game score composite

Gradient Boosted Trees

Regressions of GPA on Comparisons of Hierarchically Nested Regression Models

					Comparison to Model 3			
Model	R^2	*F*	*df*	*p*	ΔR^2	*F*	*df*	*p*
Model 1: Game Score Composite	.027	13.98	1, 509	< .001	0.00	0.22	1,509	.636
Model 2: Trace Data Score	.022	11.48	1,509	< .001	0.005	2.67	1,509	.103
Model 3: Model 1 + Model 2	.027	7.09	2,508	< .001				

Findings

Predicting GPA: incremental validity of trace data (predicted g) over game score composite

Support Vector Machines

Regressions of GPA on Comparisons of Hierarchically Nested Regression Models

					Comparison to Model 3			
Model	R^2	F	df	p	ΔR^2	F	df	p
Model 1: Game Score Composite	.027	13.98	1, 509	< .001	0.013	6.84	1,509	.009
Model 2: Trace Data Score	.040	21.00	1,509	< .001	0.00	0.02	1,509	.881
Model 3: Model 1 + Model 2	.040	10.49	2,508	.009				

Summary of findings

- Trace data (from this GBA) can be used to predict composite *g* (avg hold-out sample *r* = .65)

- Across models, some features were consistently weighted more heavily

- Evidence of *some* incremental validity when predicting GPA, depending on modeling approach
  - Machine learning model changes construct measurement

Conclusions & Future Directions

- Trace data vs. Game Score Composite
- **Infinite** number of features and modeling approaches
- Trace data-related findings are generally **context dependent and data-driven**, limiting generalizability
- Research opportunities are abundant
  - Why are some models better than others?
  - Inductive & data driven benefits?

Thank you!

Contact information:

Elena Auer

auer0027@umn.edu

UNIVERSITY OF MINNESOTA
Driven to Discover®

References

Kuhn, M. (2014). Package ‘caret’. The R Project for Statistical Computing.

Landers, R. N., Armstrong, M. B. & Collmus, A. B., Mujcic, S. & Blaik, J. A. (2017, April). *Empirical validation of a general cognitive ability assessment game.* In R. N. Landers (Chair) and B. Hawkes (Discussant), Assessment Games and Gamified Assessment. Symposium presented at the 32nd Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods, 21*(3), 689-732.

Correlation Matrix

Variable	1	2	3	4	5	6
1. Game Score (composite)	1					
2 . g (composite)	.63	1				
3. GPA	.16	.20	1			
4. Predicted g from trace data (GLMNET)	.84	.62	.13	1		
5. Predicted g from trace data (RF)	.77	.91	.20	.77	1	
6. Predicted g from trace data (GBM)	.84	.69	.15	.86	.85	1
7. Predicted g from trace data (SVM Radial)	.84	.77	.20	.82	.90	.87