Developing and Revising the Student Survey of Motivational Attitudes Toward Statistics: Results from a Pilot Study

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada 2021 Annual Meeting

9 June 2021

DUE-2013392

Why motivational attitudes?

"People forget what they do not use. But attitudes 'stick'"

(Ramirez et al., 2012, p. 57)

- Long history of measuring attitudes toward statistics
- Proliferation of instruments (Nolan et al., 2012; Ramirez et al., 2012)
- Survey of Attitudes Toward Statistics (SATS) instrument (Schau, 1992, 2003b) is widely used
  - Consistent with Expectancy-Value Theory (Schau, 2003a)
    - … but "developed without a theoretical basis" (Xu & Schau, 2019, p. 42)
  - Growing challenges to the use of the SATS (e.g., Whitaker et al., 2019b, in press)

S-SOMAS: Overview

- Student Survey of Motivational Attitudes toward Statistics (S-SOMAS)
  - For more information see Unfried et al. (2018) and Whitaker et al. (2019a)
  - Based on Expectancy-Value Theory (Eccles (Parsons) et al., 1983; Eccles & Wigfield, 2020)

This material is based upon work supported by the National Science Foundation under Grant No. DUE-2013392.

S-SOMAS: Overview

- Student Survey of Motivational Attitudes toward Statistics (S-SOMAS)
  - For more information see Unfried et al. (2018) and Whitaker et al. (2019a)
  - Based on Expectancy-Value Theory (Eccles (Parsons) et al., 1983; Eccles & Wigfield, 2020)

	Student Instrument	**Instructor Instrument**	**Environment Inventory**
Statistics	S-SOMAS	I-SOMAS	E-SOMAS
Data Science	S-SOMADS	I-SOMADS	E-SOMADS

This material is based upon work supported by the National Science Foundation under Grant No. DUE-2013392.

S-SOMAS: Overview

- Student Survey of Motivational Attitudes toward Statistics (S-SOMAS)
  - For more information see Unfried et al. (2018) and Whitaker et al. (2019a)
  - Based on Expectancy-Value Theory (Eccles (Parsons) et al., 1983; Eccles & Wigfield, 2020)

	Student Instrument	**Instructor Instrument**	**Environment Inventory**
Statistics	S-SOMAS *(Pilot 1)*	I-SOMAS *(in development)*	E-SOMAS *(in development)*
Data Science	S-SOMADS *(in development)*	I-SOMADS *(pre-development)*	E-SOMADS *(pre-development)*

This material is based upon work supported by the National Science Foundation under Grant No. DUE-2013392.

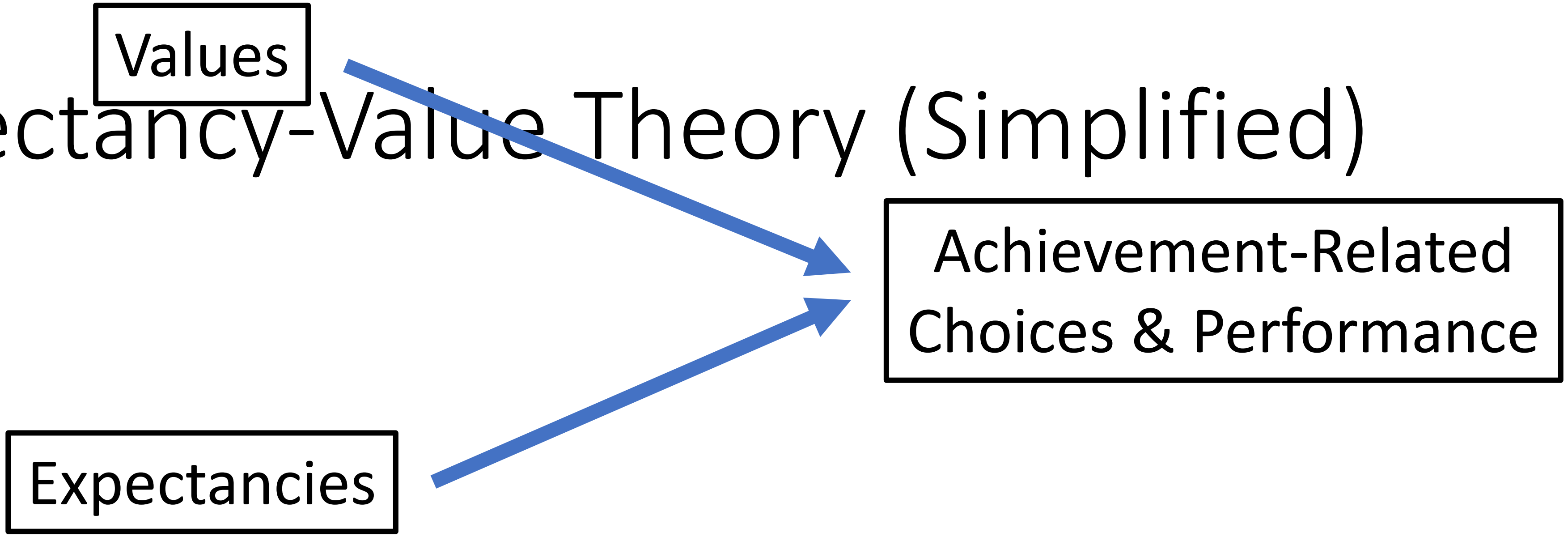
MASDER Team

The *Motivational Attitudes in Statistics and Data Science Education Research* (MASDER) team:

- Leyla Batakci *Elizabethtown College*
- Wendi Bolon *Monmouth College*
- Marjorie Bond *Monmouth College*
- April Kerby *Winona State University*
- Michael Posner *Villanova University*
- Alana Unfried *California State University, Monterey Bay*
- Douglas Whitaker *Mount Saint Vincent University*

Also: numerous undergraduate and graduate student assistants (including Matt Dunham); Research On Statistics Attitudes (ROSA) Working Group; USCOTS 2015 and 2017 Workshop participants; *many more!*

Expectancy-Value Theory (Simplified)



- Originally developed to explain motivation for learning mathematics among students in grades 5-12 (Eccles (Parsons) et al., 1983) and is actively developed (Eccles & Wigfield, 2020)
- Widely used across disciplines and age (Eccles & Wigfield, 2002)
- Has been applied with university students (Eccles & Wigfield, 2020)

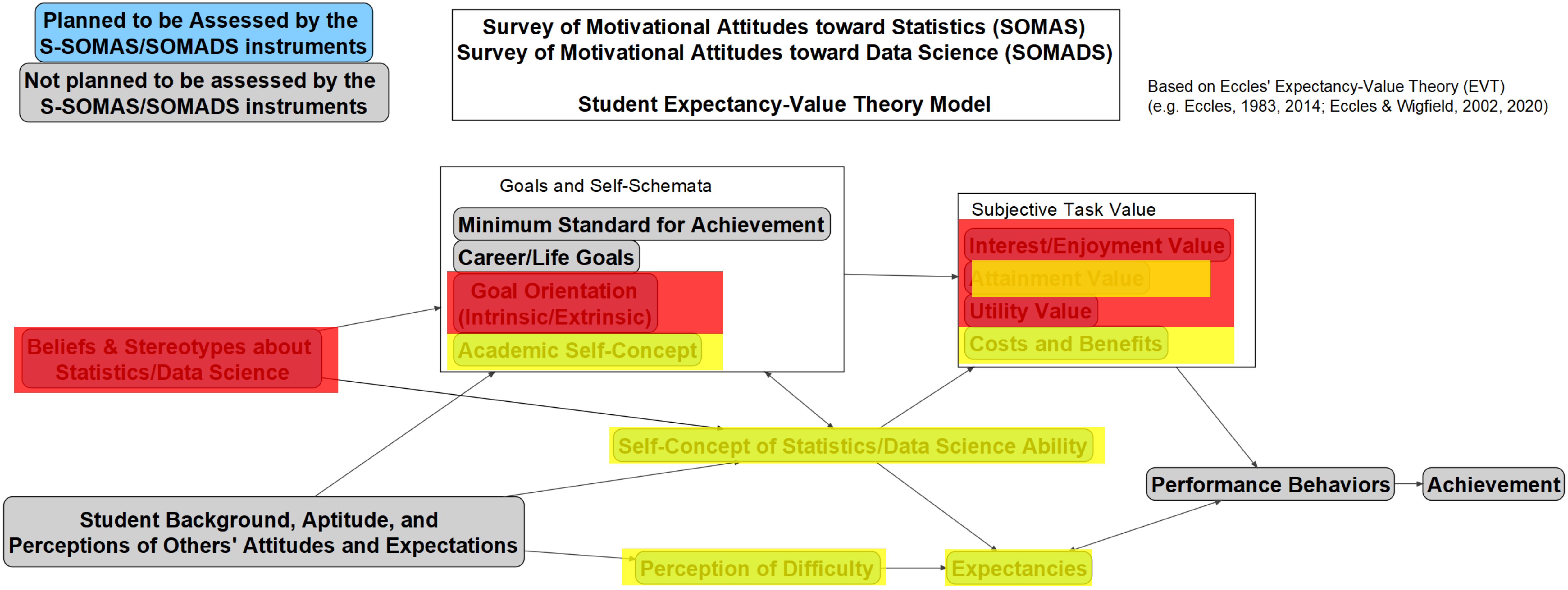
Challenges to using EVT for S-SOMAS

- Want the S-SOMAS to be useful longitudinally… and not require enrolment in a statistics course

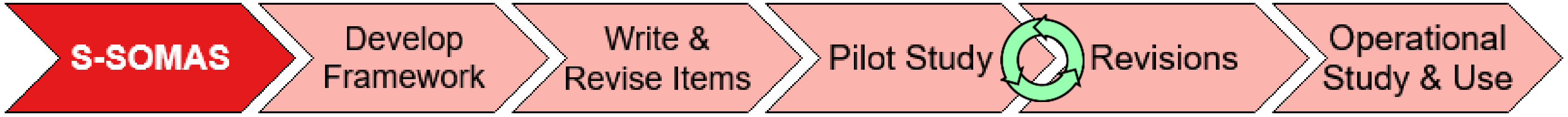
- Some EVT constructs have been researched less than others
    - Especially *Costs & Benefits* (e.g., Flake et al., 2015; Wigfield et al., 2017)

- *How should the EVT constructs be operationalized as scales?*

EVT model for the S-SOMAS instrument



S-SOMAS: Pilot 0



- 92 items measuring 11 constructs
- Split into two forms (one construct on both)
  - Form 1: 49 items and 6 constructs, n = 1155 introductory statistics students
  - Form 2: 50 items and 6 constructs, n = 1159 introductory statistics students
- Main questions:
  - Which constructs are we able to measure?
  - Which scales need to be revised?
  - Which items need to be revised?

- Confirmatory Factor Analysis (CFA)
  - Work presented for Pilot 0 Form 2
- Item Response Theory (IRT)
  - Focus on the *Perception of Difficulty (Difficult)* scale from Pilot 0 Form 2

Results will inform discussions and revisions!

Confirmatory Factor Analysis

- Testing factor structure against hypothesized EVT framework

Confirmatory Factor Analysis

- Testing factor structure against hypothesized EVT framework

	Description	CFI	RMSEA	RMSEA 90% CI Upper	SRMR	Note
Model 2.1	All items to hypothesized constructs	0.943	0.108	0.109	0.090	Covariance matrix not positive definite
Model 2.2	Some items from PCA analysis dropped	0.946	0.103	0.104	0.088	Covariance matrix not positive definite
Model 2.3	Model 2.2, but hierarchical	0.943	0.106	0.107	0.090	Covariance matrix not positive definite
Model 2.4	Model 2.2, but Academic SC and Statistics SC combined	0.935	0.112	0.114	0.095	Covariance matrix not positive definite
Model 2.5	Model 2.2, but Attainment Value and Costs combined	0.945	0.103	0.105	0.088	
Model 2.6	Model 2.2, but Difficulty and Expectancy combined	0.936	0.111	0.113	0.094	Covariance matrix not positive definite
	Recommended value (Hu & Bentler, 1999)	*≥ 0.95*	*≤ 0.05*	*≤ 0.10*	*≤ 0.08*	*None of the above values satisfy the criteria*

Confirmatory Factor Analysis

- Testing factor structure against hypothesized EVT framework
- Evidence of misfit; model modification driven by EVT theory
  - … still substantial model misfit
- Will revisit this after the next pilot!

	Description	CFI	RMSEA	RMSEA 90% CI Upper	SRMR	Note
Model 2.1	All items to hypothesized constructs	0.943	0.108	0.109	0.090	Covariance matrix not positive definite
Model 2.2	Some items from PCA analysis dropped	0.946	0.103	0.104	0.088	Covariance matrix not positive definite
Model 2.3	Model 2.2, but hierarchical	0.943	0.106	0.107	0.090	Covariance matrix not positive definite
Model 2.4	Model 2.2, but Academic SC and Statistics SC combined	0.935	0.112	0.114	0.095	Covariance matrix not positive definite
Model 2.5	Model 2.2, but Attainment Value and Costs combined	0.945	0.103	0.105	0.088	
Model 2.6	Model 2.2, but Difficulty and Expectancy combined	0.936	0.111	0.113	0.094	Covariance matrix not positive definite
	Recommended value (Hu & Bentler, 1999)	*≥ 0.95*	*≤ 0.05*	*≤ 0.10*	*≤ 0.08*	*None of the above values satisfy the criteria*

Item Response Theory

- Goal: identify items that are performing poorly/not fitting well
- We will compare two models:
  - Generalized Partial Credit Model (GPCM; Muraki, 1992)

$$P_{ijc} = \frac{\exp \sum_{k=0}^{c} a_i(\theta_j - \delta_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h} a_i(\theta_j - \delta_{ik})}$$

    - P_{ijc} is the probability of person j scoring c on item i
    - θ_j is the ability of person j
    - δ_{ik} is the (threshold) parameter for item i for responding to category k rather than $k-1$
    - m_i is the number of response categories for item i
    - a_i is the discrimination parameter for item i
  - Partial Credit Model (PCM; Masters, 1982)
    - Special case of the GPCM with $a_i = 1$
  - *Note*: Graded Response Model (GRM; Samejima, 1969) also considered but not presented here – results are *very similar* to GPCM results

Dimensionality

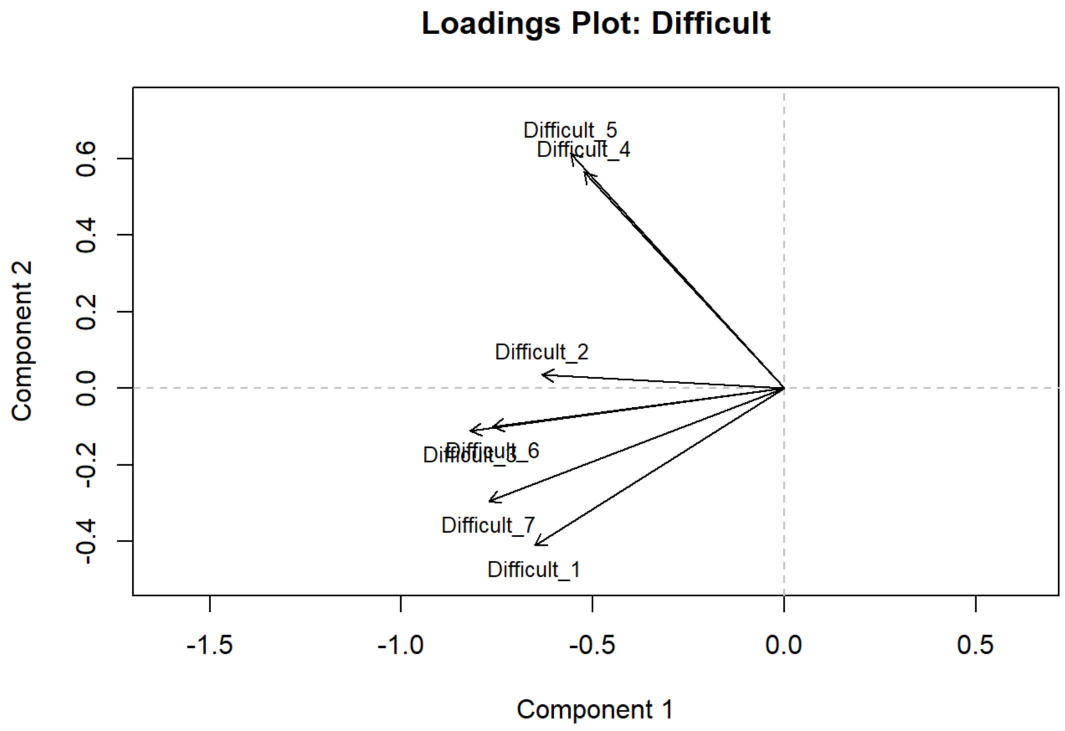
- PCA used to assess unidimensionality assumption for IRT
  - `Gifi` package in R (Mair & De Leeuw, 2019)

- Roughly homogenous loadings on the first two components suggests items are measuring the same construct (Mair, 2018)

Dimensionality

- PCA used to assess unidimensionality assumption for IRT
  - `Gifi` package in R (Mair & De Leeuw, 2019)

- Roughly homogenous loadings on the first two components suggests items are measuring the same construct (Mair, 2018)

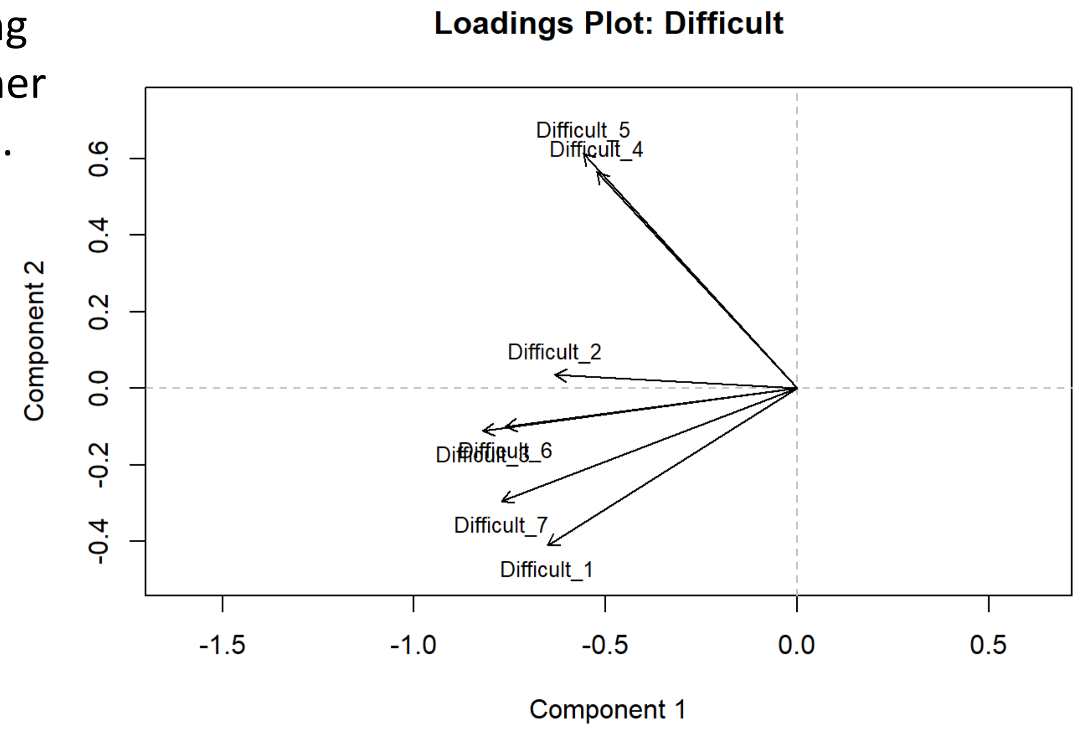


Dimensionality

This construct is an individual's perceived difficulty of statistics. Task difficulty is relative: tasks that require greater information processing power or require higher levels of skill, knowledge, or effort than other tasks are termed *difficult tasks* (Huber, 1985; Mangos & Steele-Johnson, 2001). Statistics is viewed as a "task" to be performed.

Items in the *Difficult* scale on Pilot 0 Form 2:

1. You must work hard to understand statistics.
2. Interpreting statistical results is straightforward.
3. Statistics is easy.
4. Only smart people can do statistics.
5. Anybody can do statistics.
6. It is challenging to solve a problem that requires using statistics.
7. Learning statistics for the first time is hard.

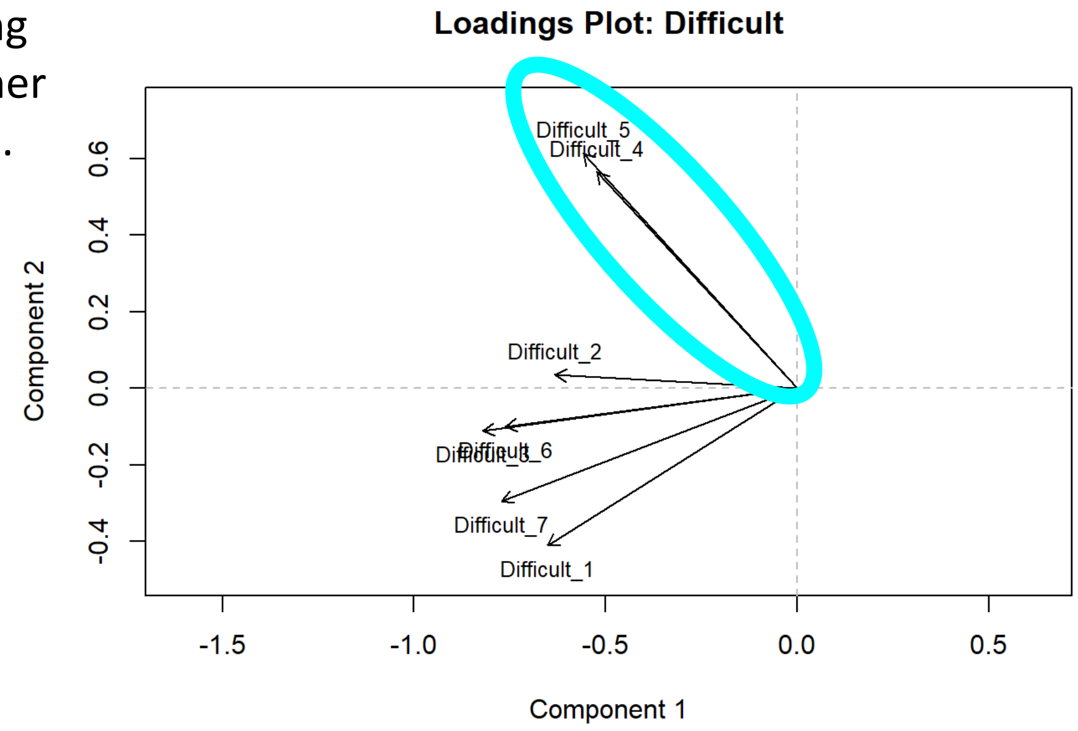


Dimensionality

This construct is an individual's perceived difficulty of statistics. Task difficulty is relative: tasks that require greater information processing power or require higher levels of skill, knowledge, or effort than other tasks are termed *difficult tasks* (Huber, 1985; Mangos & Steele-Johnson, 2001). Statistics is viewed as a "task" to be performed.

Items in the *Difficult* scale on Pilot 0 Form 2:

1. You must work hard to understand statistics.
2. Interpreting statistical results is straightforward.
3. Statistics is easy.
4. Only smart people can do statistics.
5. Anybody can do statistics.
6. It is challenging to solve a problem that requires using statistics.
7. Learning statistics for the first time is hard.

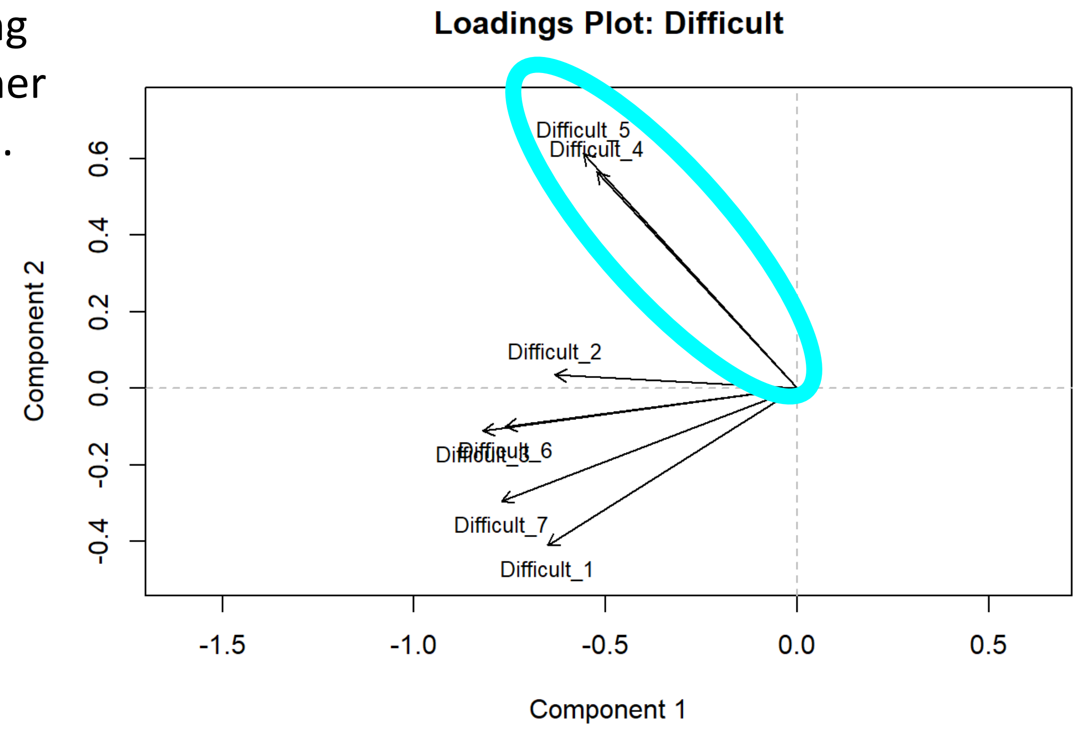


Dimensionality

This construct is an individual's perceived difficulty of statistics. Task difficulty is relative: tasks that require greater information processing power or require higher levels of skill, knowledge, or effort than other tasks are termed *difficult tasks* (Huber, 1985; Mangos & Steele-Johnson, 2001). Statistics is viewed as a "task" to be performed.

Items in the *Difficult* scale on Pilot 0 Form 2:

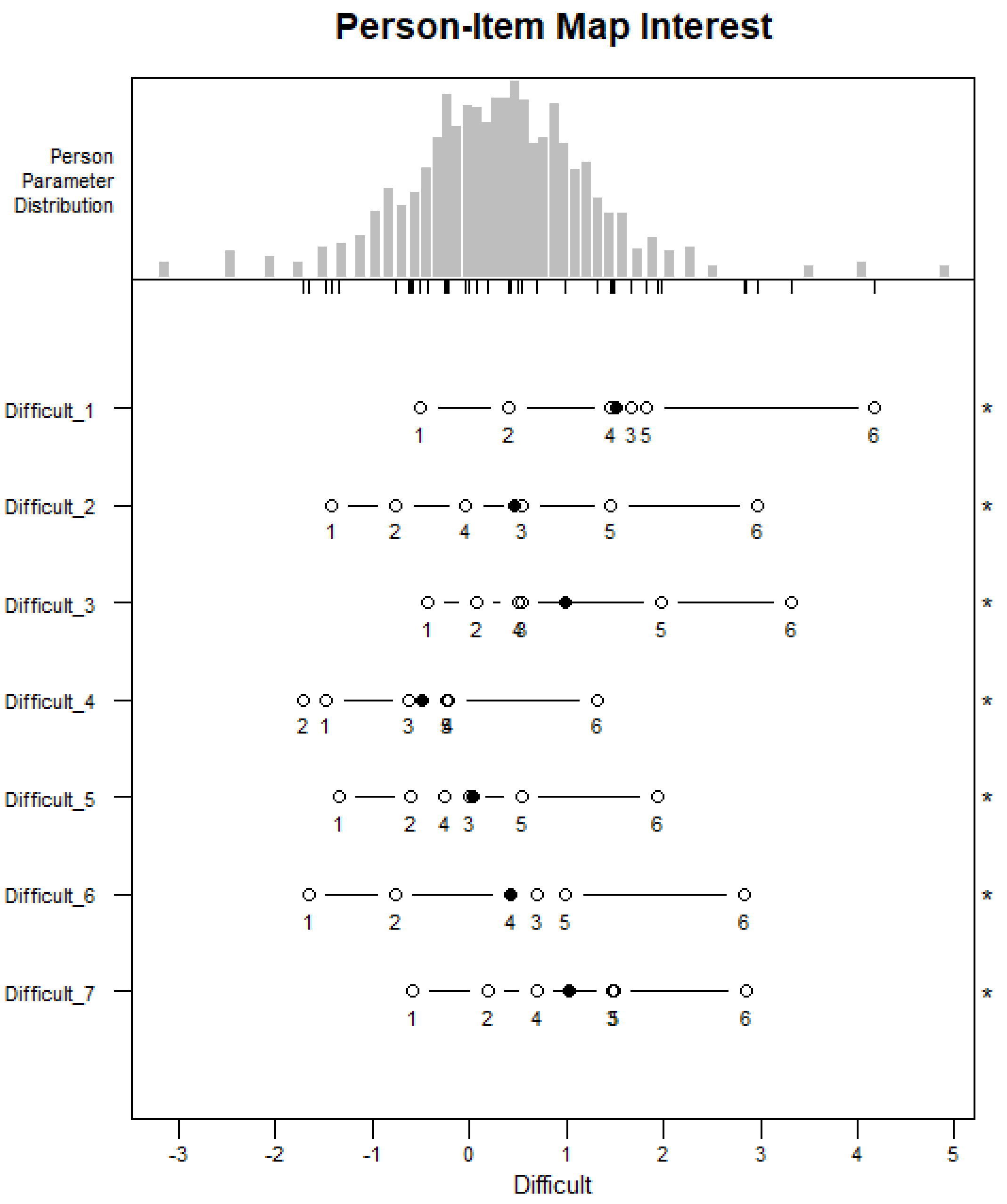
1. You must work hard to understand statistics.
2. Interpreting statistical results is straightforward.
3. Statistics is easy.
4. Only smart people can do statistics.
5. Anybody can do statistics.
6. It is challenging to solve a problem that requires using statistics.
7. Learning statistics for the first time is hard.



Partial Credit Model

- eRm package (Mair et al., 2021)
- Disordered category thresholds
  - Problematic (Andrich, 2013)

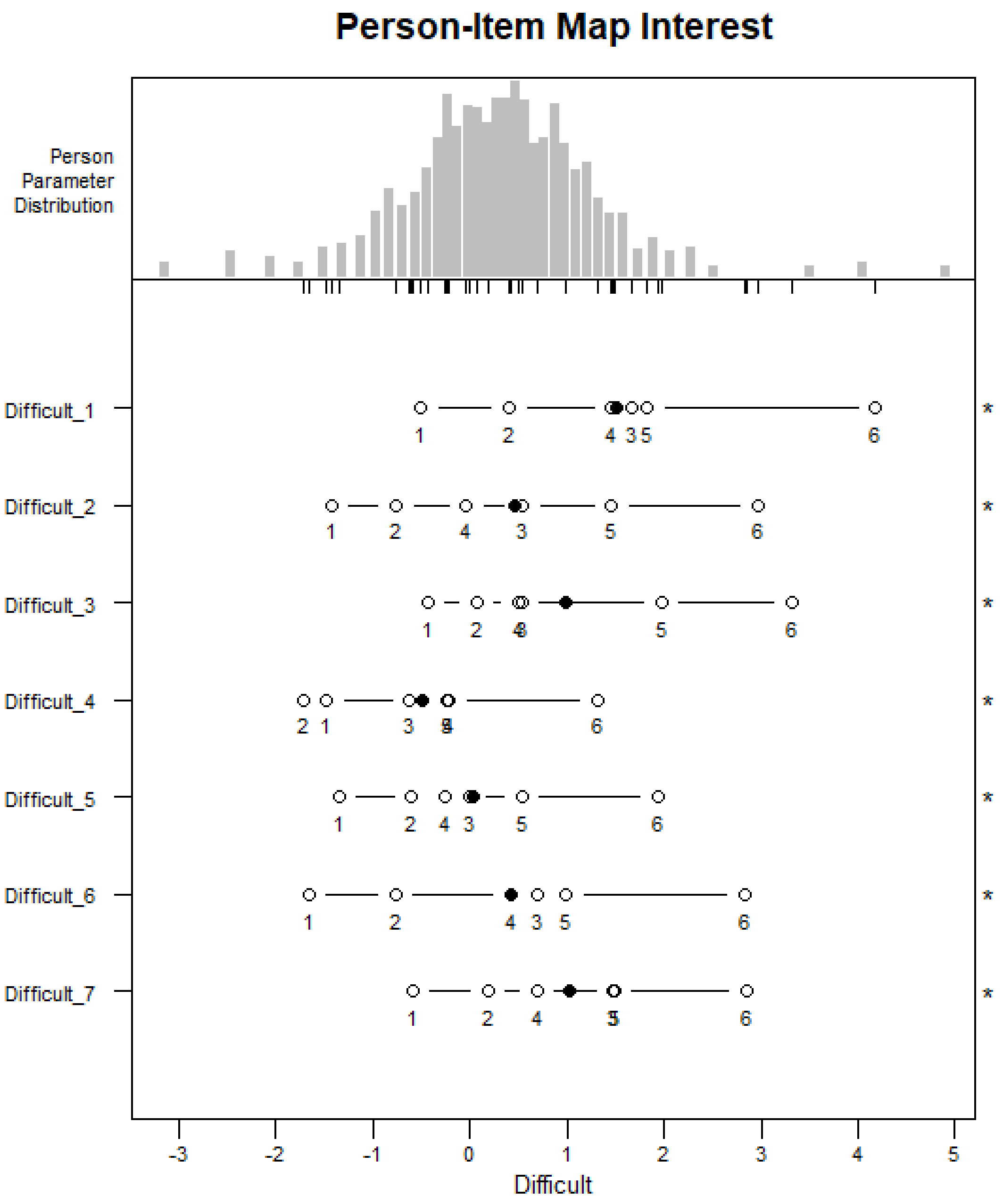
	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
Difficult_1	1044.168	1153	0.990	0.905	0.912	-2.204	-2.056	0.554
Difficult_2	1142.479	1153	0.582	0.990	0.978	-0.238	-0.557	0.532
Difficult_3	725.729	1153	1.000	0.629	0.623	-10.376	-11.339	0.778
Difficult_4	1278.371	1153	0.006	1.108	1.136	2.330	3.069	0.428
Difficult_5	1324.238	1153	0.000	1.148	1.086	3.318	2.098	0.472
Difficult_6	858.580	1153	1.000	0.744	0.745	-7.035	-7.224	0.689
Difficult_7	850.043	1153	1.000	0.737	0.712	-6.596	-7.668	0.705



Partial Credit Model

- eRm package (Mair et al., 2021)
- Disordered category thresholds
  - Problematic (Andrich, 2013)

	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
Difficult_1	1044.168	1153	0.990	0.905	0.912	-2.204	-2.056	0.554
Difficult_2	1142.479	1153	0.582	0.990	0.978	-0.238	-0.557	0.532
Difficult_3	725.729	1153	1.000	0.629	0.623	-10.376	-11.339	0.778
Difficult_4	1278.371	1153	0.006	1.108	1.136	2.330	3.069	0.428
Difficult_5	1324.238	1153	0.000	1.148	1.086	3.318	2.098	0.472
Difficult_6	858.580	1153	1.000	0.744	0.745	-7.035	-7.224	0.689
Difficult_7	850.043	1153	1.000	0.737	0.712	-6.596	-7.668	0.705

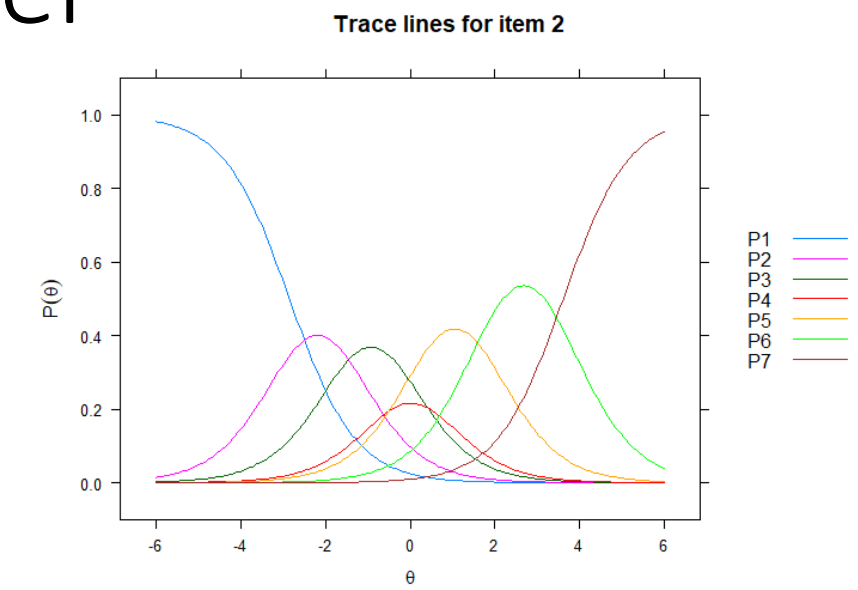


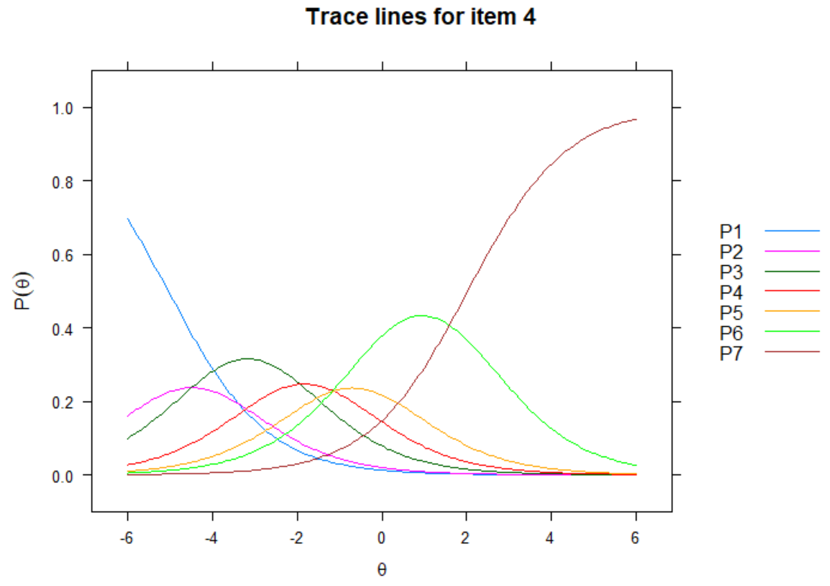
Generalized Partial Credit Model

- `mirt` package (Chalmers, 2012)

item	outfit	z.outfit	infit	z.infit	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
Difficult_1	0.900	-1.777	0.911	-1.737	129.928	91	0.019	0.005
Difficult_2	0.939	-1.514	0.937	-1.616	120.814	105	0.011	0.139
Difficult_3	0.694	-5.389	0.723	-6.649	106.358	77	0.018	0.015
Difficult_4	0.914	-1.809	0.945	-1.208	117.100	108	0.009	0.259
Difficult_5	0.938	-1.555	0.945	-1.458	139.181	125	0.010	0.182
Difficult_6	0.833	-3.916	0.850	-3.699	111.294	91	0.014	0.073
Difficult_7	0.790	-3.566	0.812	-4.022	93.465	82	0.011	0.182

	a1	d1	d2	d3	d4	d5	d6
Difficult_1	1.518	1.880	-0.188	-2.054	-3.152	-4.413	-6.799
Difficult_2	1.285	3.661	1.960	0.411	-0.468	-2.251	-4.644
Difficult_3	2.728	3.265	1.107	-0.600	-2.064	-4.611	-7.641
Difficult_4	0.864	4.352	3.384	2.076	1.067	0.103	-1.753
Difficult_5	0.947	3.468	2.103	1.103	0.346	-0.854	-2.763
Difficult_6	1.903	4.556	2.273	0.166	-0.969	-2.470	-5.137
Difficult_7	2.379	2.772	0.346	-1.748	-2.713	-4.240	-6.590



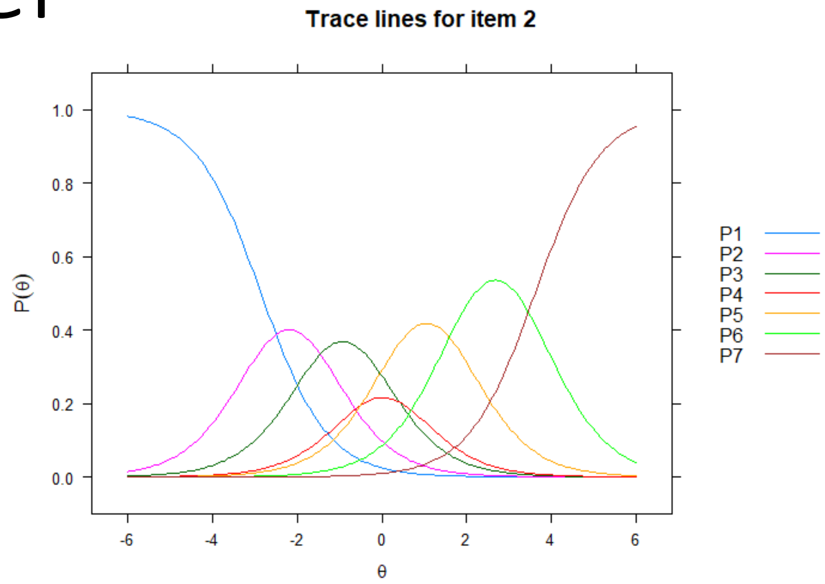


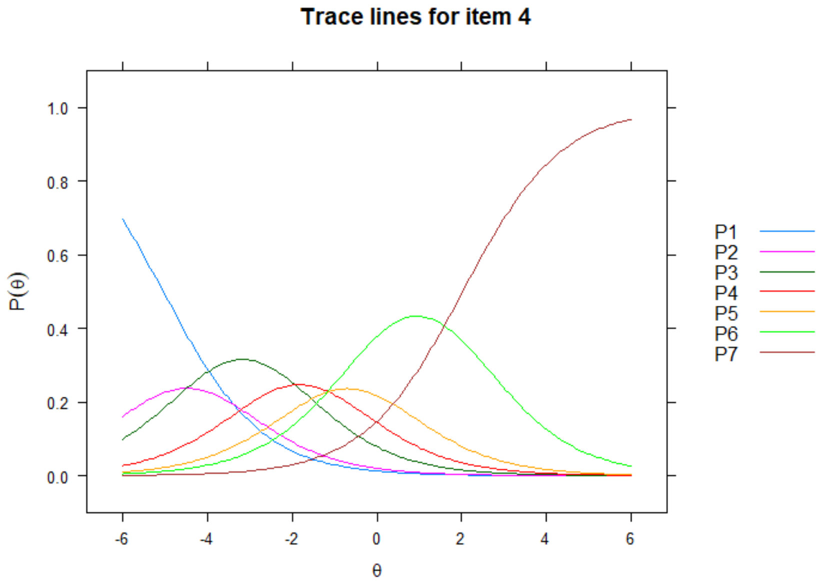
Generalized Partial Credit Model

- `mirt` package (Chalmers, 2012)

item	outfit	z.outfit	infit	z.infit	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
Difficult_1	0.900	-1.777	0.911	-1.737	129.928	91	0.019	0.005
Difficult_2	0.939	-1.514	0.937	-1.616	120.814	105	0.011	0.139
Difficult_3	0.694	-5.389	0.723	-6.649	106.358	77	0.018	0.015
Difficult_4	0.914	-1.809	0.945	-1.208	117.100	108	0.009	0.259
Difficult_5	0.938	-1.555	0.945	-1.458	139.181	125	0.010	0.182
Difficult_6	0.833	-3.916	0.850	-3.699	111.294	91	0.014	0.073
Difficult_7	0.790	-3.566	0.812	-4.022	93.465	82	0.011	0.182

	a1	d1	d2	d3	d4	d5	d6
Difficult_1	1.518	1.880	-0.188	-2.054	-3.152	-4.413	-6.799
Difficult_2	1.285	3.661	1.960	0.411	-0.468	-2.251	-4.644
Difficult_3	2.728	3.265	1.107	-0.600	-2.064	-4.611	-7.641
Difficult_4	0.864	4.352	3.384	2.076	1.067	0.103	-1.753
Difficult_5	0.947	3.468	2.103	1.103	0.346	-0.854	-2.763
Difficult_6	1.903	4.556	2.273	0.166	-0.969	-2.470	-5.137
Difficult_7	2.379	2.772	0.346	-1.748	-2.713	-4.240	-6.590





IRT Summary

- PCM fits poorly, GPCM fits better
- With GPCM there are still many items that exhibit misfit (either infit or outfit)
- Response scale – too many points?
- Most scales have at least a few items that seem fine
    - (Build out new scales using these items?)

	# Items	# PCA	# Misfit	Total
Beliefs & Stereotypes	10	4	4	8
Intrinsic GO	7	0	7	7
Extrinsic GO	8	2	3	5
Utility Value	8	0	4	4
Interest value	9	0	3	3
Attainment Value (1)	7	4	3	4
Attainment Value (2)	7	2	4	5
Academic SC	9	0	1	1
Statistics SC	9	0	6	6
Difficulty	7	2	3	5
Expectancy	11	3	1	4
Costs & Benefits	7	2	2	4

Note: Misfit (infit or outfit) is from a GRM IRT analysis. Some items may have been identified as problematic in both the PCA and IRT analysis, so Total is not the sum of the two columns.

Difficulty	AIC	BIC	log.Lik	LRT	df	p.value
PCM	24660.64	24872.82	-12288.3		42	
GPCM	24599.38	24846.92	-12250.7	75.26	49	<0.001

Conclusions, Limitations, and Next Steps

- Lots of information for the MASDER team to review when revising the S-SOMAS instrument
  - CFA, PCA, IRT
  - Improved definitions
  - EFA results from colleagues (e.g., Unfried et al., 2018)
- Decision to split constructs into two forms limits interpretations
  - Pilot 1 includes all constructs on one form
- Next steps:
  - Revise items, remove items, write new items
  - Change number of response points (e.g., go from 7 to 5)
  - (Change response options? Rewrite items? [Drop Agree/Disagree?])
  - Use lessons when developing I-SOMAS, S-SOMADS

Questions?

Thank you!

Developing and Revising the Student Survey of Motivational Attitudes Toward Statistics: Results from a Pilot Study

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada Annual Meeting
9 June 2021

DUE-2013392

References

- Andrich, D. (2013). An Expanded Derivation of the Threshold Structure of the Polytomous Rasch Model That Dispels Any “Threshold Disorder Controversy.” *Educational and Psychological Measurement, 73*(1), 78–124. https://doi.org/10.1177/0013164412450877
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the *R* Environment. *Journal of Statistical Software*, *48*(6). https://doi.org/10.18637/jss.v048.i06
- Eccles, J. S. (2014). Expectancy-Value Theory. In R. Eklund & G. Tenenbaum (Eds.), *Encyclopedia of Sport and Exercise Psychology*. SAGE Publications, Inc. https://doi.org/10.4135/9781483332222.n110
- Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*, 109–132.
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859
- Eccles (Parsons), J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches*. W.H. Freeman. http://web.archive.org/web/20170701031033/http://www.rcgd.isr.umich.edu/garp/articles/ecclesparsons83b.pdf
- Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
- Huber, V. L. (1985). Effects of Task Difficulty, Goal Setting, and Strategy On Performance of a Heuristic Task. *Journal of Applied Psychology, 70*(3), 492–504.
- Mair, P. (2018). *Modern Psychometrics with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-93177-7
- Mair, P., & De Leeuw, J. (2019). *Gifi: Multivariate Analysis with Optimal Scaling* (R package version 0.3-9) [Computer software]. https://CRAN.R-project.org/package=Gifi
- Mair, P., Hatzinger, R., & Maier, M. J. (2021). *eRm: Extended Rasch Modeling* (1.0-2) [R package]. https://cran.r-project.org/package=eRm
- Mangos, P. M., & Steele-Johnson, D. (2001). The Role of Subjective Task Complexity in Goal Orientation, Self-Efficacy, and Performance Relations. *Human Performance, 14*(2), 169–185. https://doi.org/10.1207/S15327043HUP1402_03
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272
- Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. https://doi.org/10.1177/014662169201600206
- Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students’ attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103–123.
- Ramirez, C., Schau, C., & Emmioğlu, E. (2012). The Importance of Attitudes in Statistics Education. *Statistics Education Research Journal*, *11*(2), 57–71.
- Samejima, F. (1969). *Estimation of Latent Ability Using a Response Pattern of Graded Scores*. Psychometric Society. https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf
- Schau, C. (1992). *Survey of Attitudes Toward Statistics (SATS-28)*. http://evaluationandstatistics.com/
- Schau, C. (2003a). *Survey of Attitudes Toward Statistics (SATS-36)*. http://evaluationandstatistics.com/
- Schau, C. (2003b, August). *Students’ attitudes: The “other” important outcome in statistics education* [Paper]. Joint Statistical Meetings, San Francisco, CA. https://irp-cdn.multiscreensite.com/281322c3/files/uploaded/JSM2003.pdf
- Unfried, A., Kerby, A., & Coffin, S. (2018). Developing a Student Survey of Motivational Attitudes Toward Statistics. *2018 JSM Proceedings*. Joint Statistical Meetings 2018, Vancouver, Canada.
- Whitaker, D., Unfried, A., & Bond, M. (2019a). Design and validation arguments for the Student Survey of Motivational Attitudes toward Statistics (S-SOMAS) instrument. In J. D. Bostic, E. E. Krupa, & J. C. Shih (Eds.), *Assessment in Mathematics Education Contexts: Theoretical Frameworks and New Directions* (1st ed., pp. 120–146). Routledge. http://ec.msvu.ca/xmlui/handle/10587/2125
- Whitaker, D., Unfried, A., & Bond, M. (2019b, May). *Challenges to Using and Interpreting the SATS-36 Instrument: Do you like statistics? Do your students like statistics? How do you know?* [Poster]. United States Conference on Teaching Statistics (USCOTS), State College, PA. http://ec.msvu.ca:8080/xmlui/bitstream/handle/10587/2120/USCOTS%202019%20-%20SATS%20Poster%20-%20Print.pdf?sequence=1&isAllowed=y
- Whitaker, D., Unfried, A., & Bond, M. (in press). Challenges associated with measuring attitudes using the SATS family of instruments. *Statistics Education Research Journal*.
- Wigfield, A., Rosenzweig, E. Q., & Eccles, J. S. (2017). Achievement Values: Interactions, Interventions, and Future Directions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 116–134). Guilford Press.
- Xu, C., & Schau, C. (2019). Exploring Method Effects in the Six-Factor Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal*, *18*(2), 39–53. https://iase-web.org/documents/SERJ/SERJ18(2)_Schau.pdf