Analysis of Massachusetts' Standardized Testing Multi-Group Structural Equation Modelling

Landon Hurley [1] Herb Susmann [2]

[1]Psychology, SUNY Geneseo

[2]Mathematics, SUNY Geneseo

April 12, 2014

Outline

Introduction

- Massachusetts standardized test data

Introduction

- Massachusetts standardized test data
- Is the test psychometrically valid?

Introduction

- Massachusetts standardized test data
- Is the test psychometrically valid?
- **No! It doesn't measure the same thing in the same way across groups.**

Data

- 10,515 10th grade student test results

Data

- 10,515 10th grade student test results
- English, Math, and Science components

Data

- 10,515 10th grade student test results
- English, Math, and Science components
- Answers from every question for each section

Data

- 10,515 10th grade student test results
- English, Math, and Science components
- Answers from every question for each section
- Basic demographic information from questionnaire

Data

- 10,515 10th grade student test results
- English, Math, and Science components
- Answers from every question for each section
- Basic demographic information from questionnaire

Measurement Invariance

- Testing: set of standardized questions designed to elicit responses for use in measuring the traits, capacities, or achievements of an individual.
- Tests imply a global solution: the individual should never affect how items on the test perform.
- In psychometrics, validity is tested in one of two way: predictive invariance & measurement (factorial) invariance.

Measurement Invariance Process

1. Configural invariance: Tests whether the same factor model is found within each subgroup.
2. Weak invariance: Tests whether in addition to the same factor structure, the same items load equivalently onto the same structure.
3. Strong invariance: Tests that the intercepts, item loadings, and factor structure are equivalent across groups.
4. Strict invariance: Tests the assumption that in addition to the preceding steps, the residual variances are equivalent across groups.

Asessing model fit

- χ^2

$$\chi^2_{d.f.} = 2(N-1) \cdot F_0 \tag{1}$$

  Scales with N, making it difficult to fit

- CFI:

$$1 - \frac{\chi^2_t - df_t}{\chi^2_n - df_n} \tag{2}$$

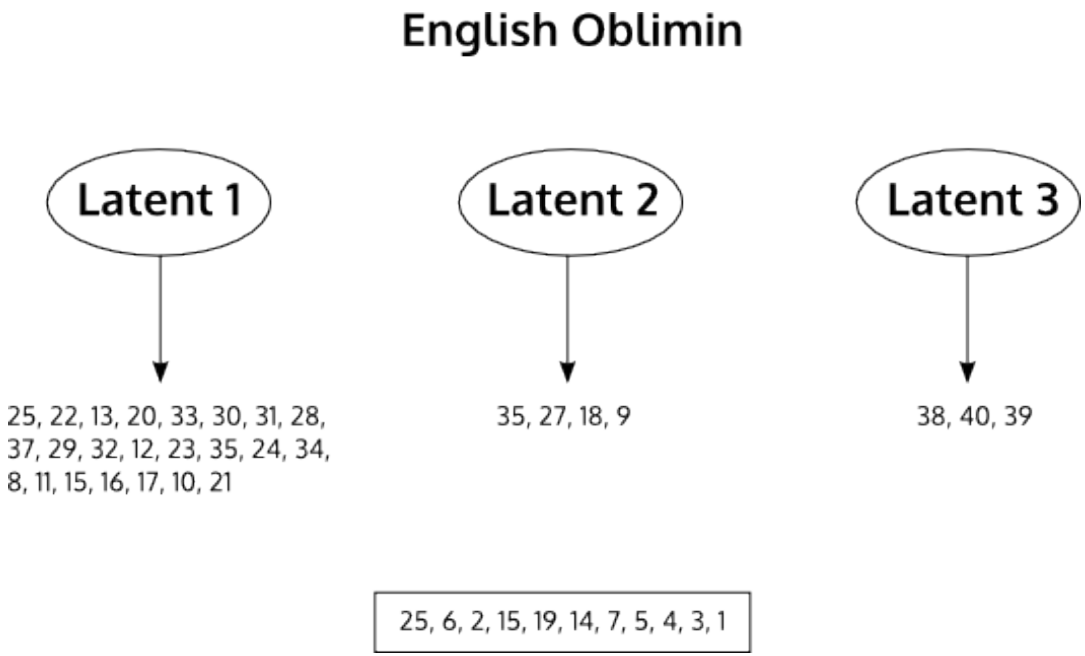
- NNFI:

$$1 - \frac{\frac{\chi^2_t}{df_t}}{\frac{\chi^2_n}{df_n}} \tag{3}$$

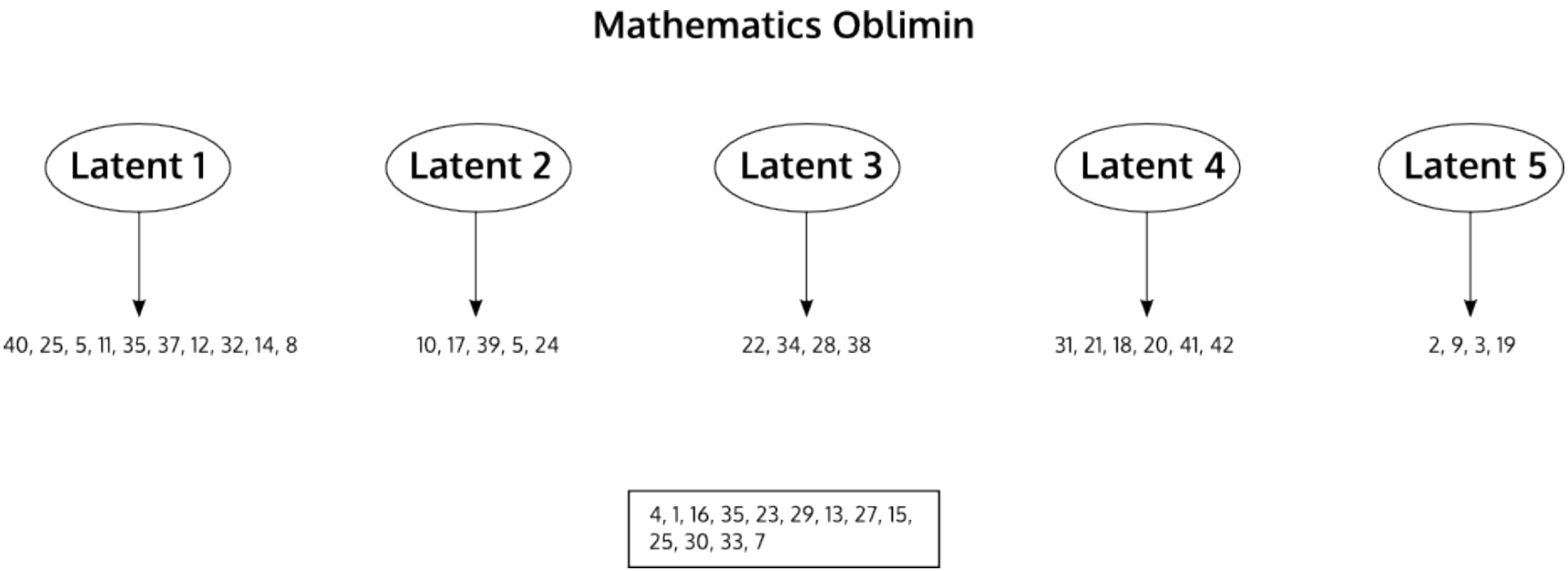
- RMSEA:

$$\sqrt{\frac{\chi^2_{df} - df_t}{df_t(N-1)}} \tag{4}$$

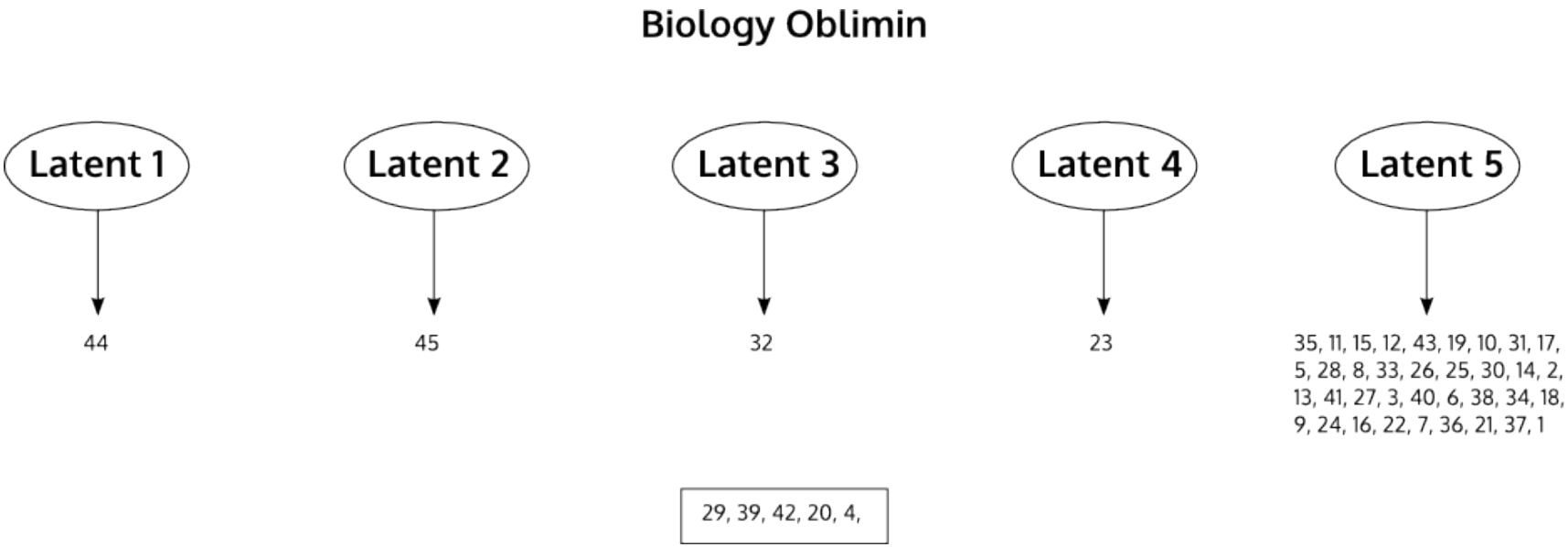
Results


English Oblimin
Latent 1
Latent 2
Latent 3
25, 22, 13, 20, 33, 30, 31, 28, 37, 29, 32, 12, 23, 35, 24, 34, 8, 11, 15, 16, 17, 10, 21
35, 27, 18, 9
38, 40, 39
25, 6, 2, 15, 19, 14, 7, 5, 4, 3, 1

Results


Mathematics Oblimin
Latent 1
Latent 2
Latent 3
Latent 4
Latent 5
40, 25, 5, 11, 35, 37, 12, 32, 14, 8
10, 17, 39, 5, 24
22, 34, 28, 38
31, 21, 18, 20, 41, 42
2, 9, 3, 19
4, 1, 16, 35, 23, 29, 13, 27, 15, 25, 30, 33, 7

Results


Biology Oblimin
Latent 1
Latent 2
Latent 3
Latent 4
Latent 5
44
45
32
23
35, 11, 15, 12, 43, 19, 10, 31, 17,
5, 28, 8, 33, 26, 25, 30, 14, 2,
13, 41, 27, 3, 40, 6, 38, 34, 18,
9, 24, 16, 22, 7, 36, 21, 37, 1
29, 39, 42, 20, 4,

Results

English	χ^2	CFI	TLI	RMSEA
Comparison	-1670.297	0.030	0.034	-0.006
Strict fit indices	6007.770	0.918	0.912	0.030

Results

Mathematics	χ^2	CFI	TLI	RMSEA
Comparison	-998.348	.014	.017	.005
Strict fit indices	4130.561	.949	.946	.033

Results

Biology	χ^2	CFI	TLI	RMSEA
Comparison	-2132.914	.026	.03	.006
Strict fit indices	7686.782	.926	.922	.031

Discussion

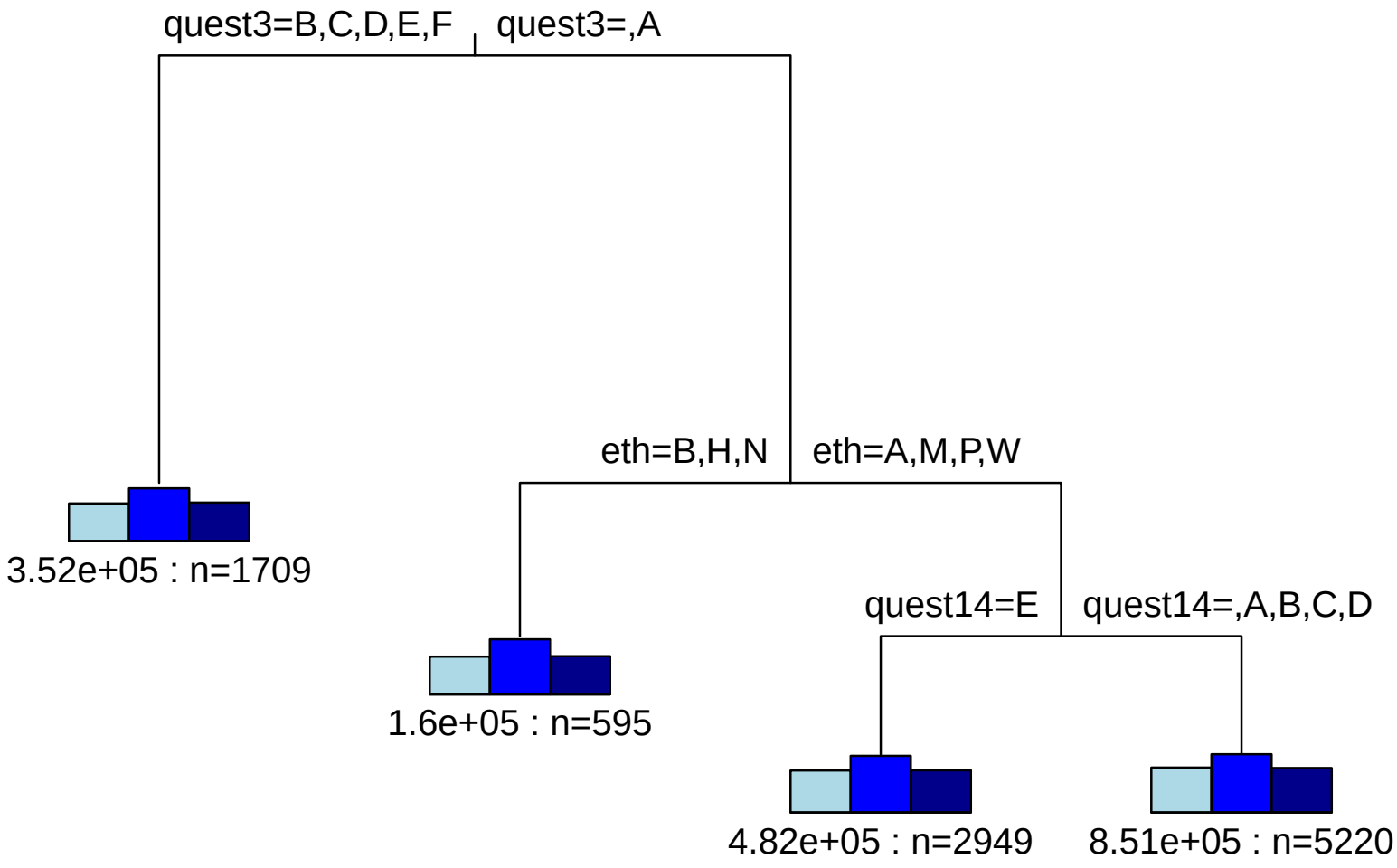
Strict invariance: what does it mean for these tests?

- SES: Lack of strict invariance typically indicates that an unmeasured latent variable is differentially affecting the test outcome. In other words, the existence of unmeasured effects is specified by the existence of error variance heterogeneity. This is despite the invariance of both the factor model, the intercepts, and slopes.
- Meredith (1993): When using measures to make decisions strict factorial invariance is essential.

Regression Tree

- CART allows for complex relationships to be modelled within data without parametric assumptions or linearity.
- Multivariate regression trees are ideal in this instance because it can model both the complex relationships, but also non-response, in relation to a multiple vectored response surface: English, Mathematics, & Science.

Results



Error : 0.903 CV Error : 0.904 SE : 0.0113

Discussion

- Missing data: Question 3 deals with after college goals. These results indicate that students who either intend to go to attain a baccalaureate degree OR skip the question perform better on the three sub-tests.
- Ethncity: Amongst students who do plan on a baccalaureate degree, ethnicity efficiently segments the science scores.
- Quest14: How often do you use scientific instruments?
- Confirmation: Each of the responses were regressed onto multiply imputed forms of the questionnaire, which had equivalent R^2, indicating effects were linear only.

Conclusions

- Fails strict invariance

Conclusions

- Fails strict invariance
- Regression tree: split on ethnicity

References I

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Cronbach, L. J. (1947). Test reliability: Its meaning and determination. *Psychometrika*, *12*(1), 1–16.

De'Ath, G. (2002). Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology*, *83*(4), 1105–1117.

DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, *46*, 137–149.

Gaudet, R. D. (2000). Effective school districts in massachusetts. *The Donahue Institute, Boston*.

References II

Kim, K. H., & Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, *67*(4), 609–623.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2013). semtools: Useful tools for structural equation modeling. [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=semTools` (R package version 0.4-0)

References III

Revelle, W. (2014). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from http://CRAN.R-project.org/package=psych (R package version 1.4.2)

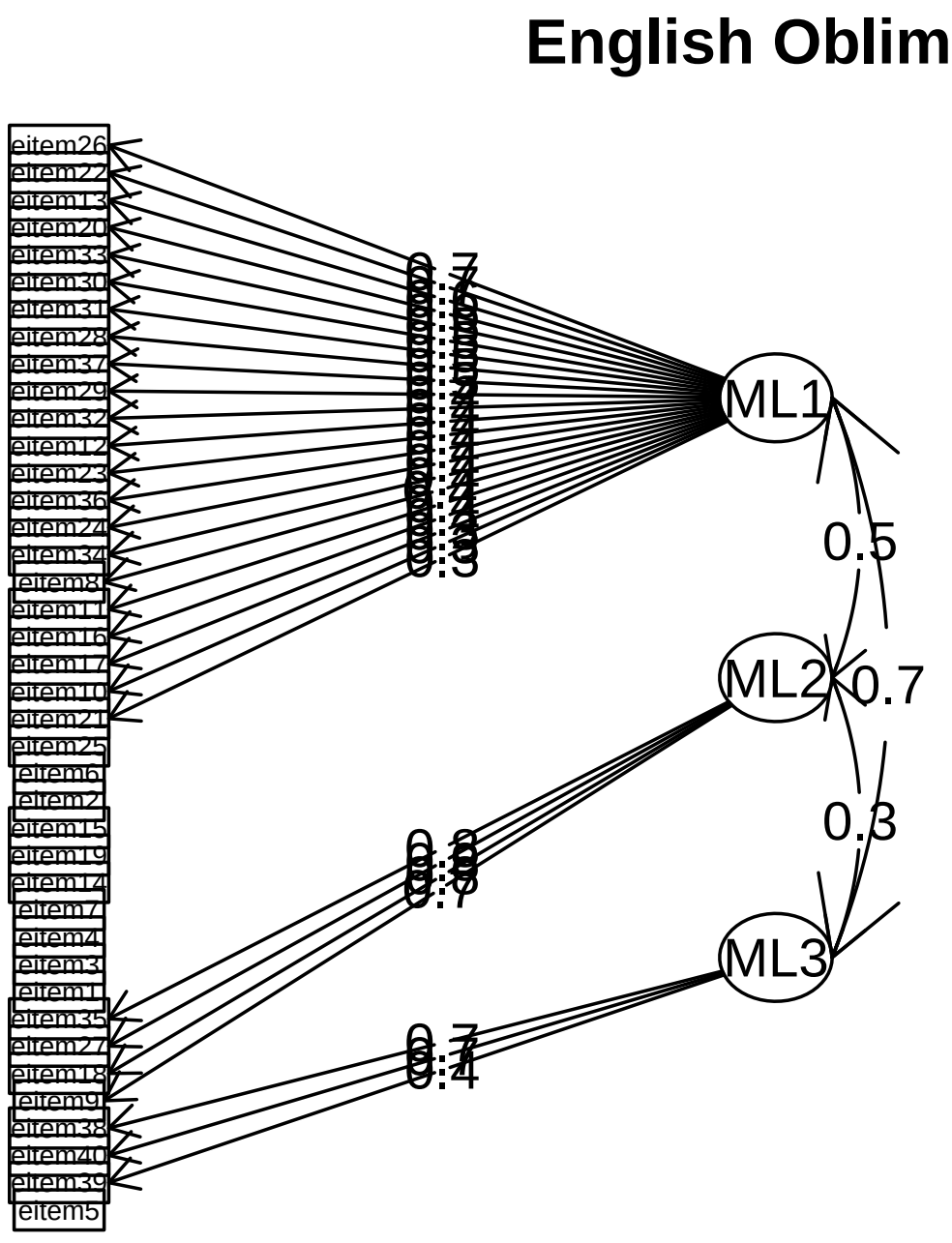
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rubin, D. B. (2009). *Multiple imputation for nonresponse in surveys* (Vol. 307). John Wiley & Sons.

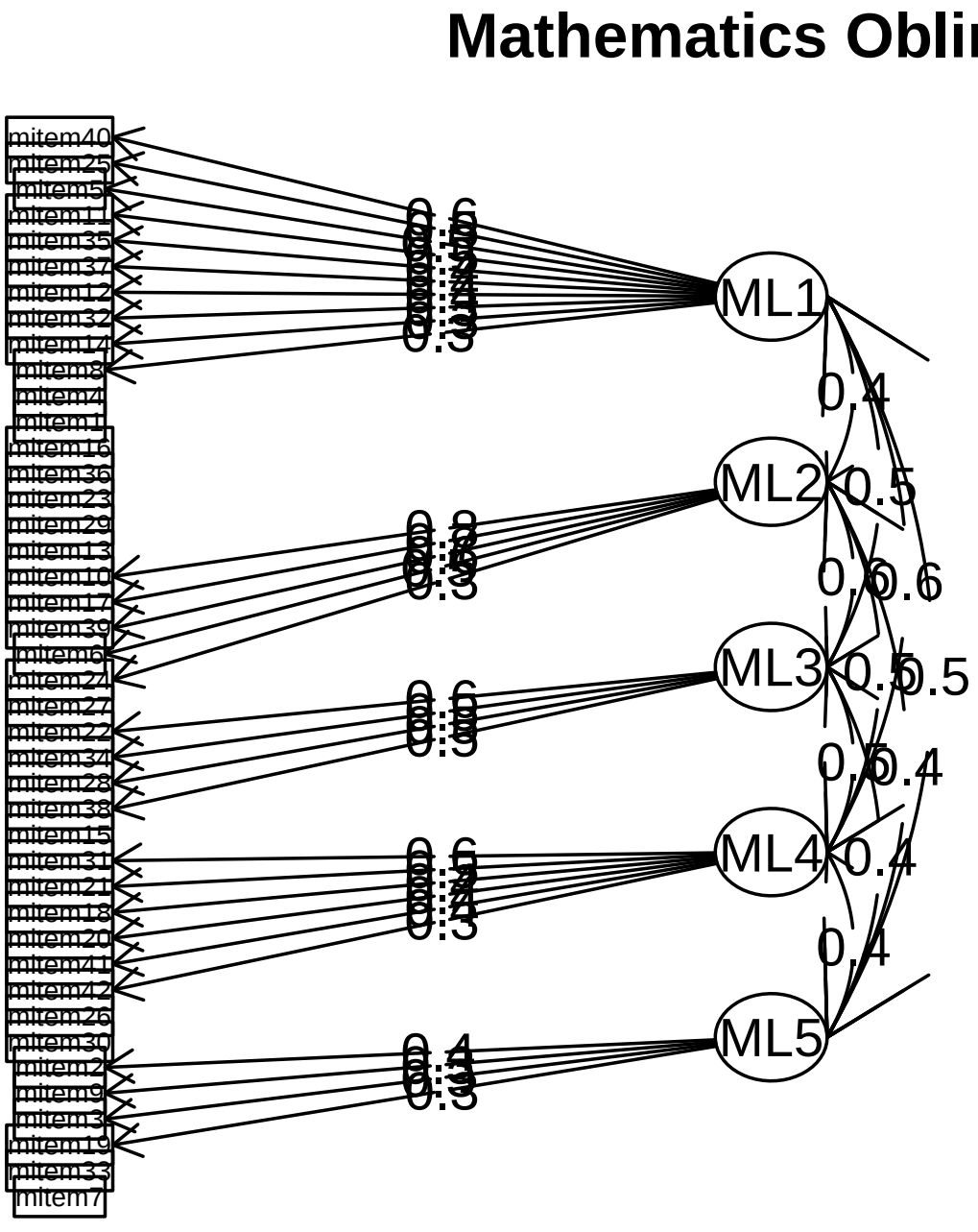
Spring 2009 mcas tests: Summary of state results. (2009, September).

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from http://www.jstatsoft.org/v45/i03/

Appendix A


English Oblimin
eitem26
eitem22
eitem13
eitem20
eitem33
eitem30
eitem31
eitem28
eitem37
eitem29
eitem32
eitem12
eitem23
eitem36
eitem24
eitem34
eitem8
eitem11
eitem16
eitem17
eitem10
eitem21
eitem25
eitem6
eitem2
eitem15
eitem19
eitem14
eitem7
eitem4
eitem3
eitem1
eitem35
eitem27
eitem18
eitem9
eitem38
eitem40
eitem39
eitem5
ML1
ML2
ML3
0.5
0.7
0.3

Appendix A


Mathematics Oblimin
mitem40
mitem25
mitem5
mitem11
mitem35
mitem37
mitem12
mitem32
mitem14
mitem8
mitem4
mitem1
mitem16
mitem36
mitem23
mitem29
mitem13
mitem10
mitem17
mitem39
mitem6
mitem24
mitem27
mitem22
mitem34
mitem28
mitem38
mitem15
mitem31
mitem21
mitem18
mitem20
mitem41
mitem42
mitem26
mitem30
mitem2
mitem9
mitem3
mitem19
mitem33
mitem7
ML1
ML2
ML3
ML4
ML5
0.4
0.5
0.6
0.5
0.4
0.4
0.4

Appendix A