AN INTRODUCTION TO BAYESIAN ESTIMATION AND MISSING DATA IMPUTATION FOR EDUCATION RESEARCH

Craig K. Enders & Michael Woller,
UCLA Department of Psychology

WWW.APPLIEDMISSINGDATA.COM/BLIMP

BLIMP 3.0

Blimp 3 offers powerful latent variable modeling and imputation for incomplete data sets with up to three levels. Blimp's unique Bayesian computational architecture allows easy specification of complex analyses that are difficult or impossible to fit in other software packages.

Download Now

User's Guide

WWW.APPLIEDMISSINGDATA.COM/BLIMP-PAPERS

APPLIED MISSING DATA home analysis examples blimp blimp papers centerstat workshop quantitude podcast

Workshops and Training

Enders, C. K., & Woller, M. (2023, April). An introduction to Bayesian estimation and missing data imputation for educational research. Professional development workshop presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Download Workshop Materials

ACKNOWLEDGEMENTS

Brian Keller & Han Du

IES Institute of Education Sciences

WORKSHOP OUTLINE

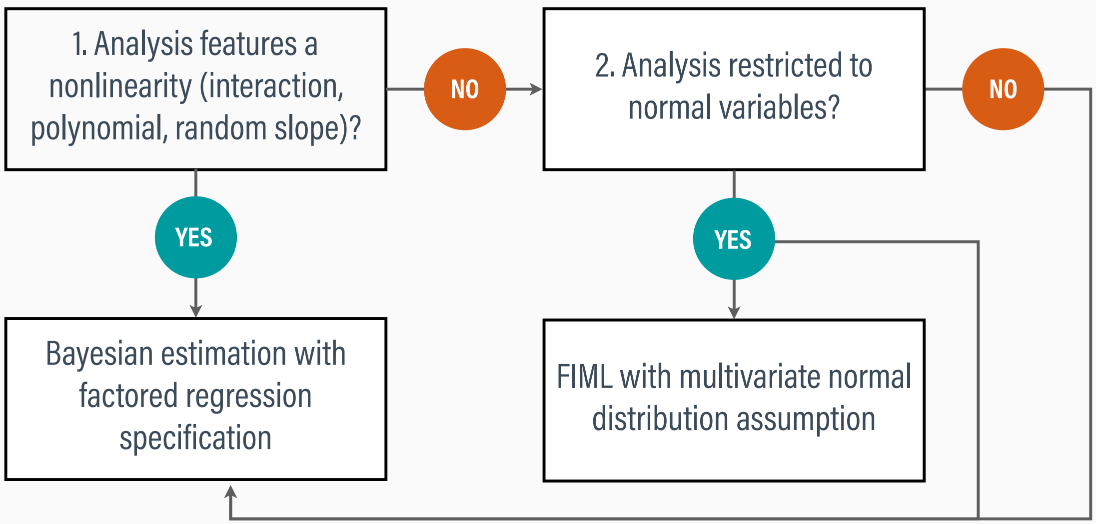
- Missing data processes
- Introduction to Bayesian estimation
- Fitting regression models in Blimp
- Understanding Blimp output
- Incomplete categorical variables
- Interaction effects in Blimp

Why go Bayesian?

THINGS BAYESIAN MCMC IS GOOD AT

- Direct estimation for complex models with missing data
- Mixed metrics (normal, ordinal, nominal, skewed, count, latent)
- Nonlinear effects (interactions, curvilinear effects)
- Multilevel data (random coefficients, interactions, heterogeneous variation)
- Latent variable modeling (interactions, multilevel)

CHOOSING A MISSING DATA METHOD



MISSING DATA MECHANISMS

PARTITIONING THE DATA

Complete			=	Observed			+	Missing			Indicators		
Y_1	Y_2	Y_3		Y_1	Y_2	Y_3		Y_1	Y_2	Y_3	M_1	M_2	M_3
4	4	3		4	4	3					0	0	0
3	**3**	5		3	**NA**	5			**3**		0	**1**	0
7	1	6		7	1	6					0	0	0
2	1	6		**NA**	1	6		**2**			**1**	0	0
5	9	3	=	5	9	3	+				0	0	0
3	**2**	**2**		3	**NA**	**NA**			**2**	**2**	0	**1**	**1**
1	6	7		1	6	7					0	0	0
9	4	9		9	4	9					0	0	0
2	**5**	6		2	**NA**	6			**5**		0	**1**	0

RUBIN'S MISSINGNESS MECHANISMS

- Missing data mechanisms describe different ways in which the pattern of 0s and 1s in M relate to observed or missing values
- Missingness may be independent of the data, or it could relate to the observed or missing parts (or both)
- Mechanisms are essentially models that describe nonresponse

MISSING COMPLETELY AT RANDOM

- The probability of missing values is completely unrelated to the data

$$f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\mathbf{Y}_{\text{mis}},\phi)=f(\mathbf{M}=1|\phi)$$

- Purely haphazard missingness

Relation Between Nonresponse and Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	1	0	3	NA	5		3	
0	0	0	7	1	6			
1	0	0	NA	1	6	2		
0	0	0	5	9	3			
0	1	1	3	NA	NA		2	2
0	0	0	1	6	7			
0	0	0	9	4	9			
0	1	0	2	NA	6		5	

(CONDITIONALLY) MISSING AT RANDOM

- The probability of missing values is unrelated to the missing (latent) data

$$f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\mathbf{Y}_{\text{mis}},\phi)=f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\phi)$$

- Missingness is haphazard after conditioning on observed data

Relation Between Nonresponse and Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	1	0	3	NA	5		3	
0	0	0	7	1	6			
1	0	0	NA	1	6	2		
0	0	0	5	9	3			
0	1	1	3	NA	NA		2	2
0	0	0	1	6	7			
0	0	0	9	4	9			
0	1	0	2	NA	6		5	

MISSING NOT AT RANDOM

- The probability of missing values is related to the missing (latent) data

$$f(\mathbf{M} = 1|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \phi) \text{ or } f(\mathbf{M} = 1|\mathbf{Y}_{\text{mis}}, \phi)$$

- The observed data may or may not additionally determine missingness

Relation Between Nonresponse and Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	**1**	0	3	**NA**	5		**3**	
0	0	0	7	1	6			
1	0	0	**NA**	1	6	**2**		
0	0	0	5	9	3			
0	**1**	**1**	3	**NA**	**NA**		**2**	**2**
0	0	0	1	6	7			
0	0	0	9	4	9			
0	**1**	0	2	**NA**	6		**5**	

EDUCATION EXAMPLES

- MCAR = longitudinal planned missing data design where participants provide data at a subset of the repeated measures
- Conditional MAR = students from low SES households are more likely to have missing achievement scores, but missingness is random after controlling for SES
- MNAR = students with low achievement levels are more likely to have missing achievement test scores

WORKING ASSUMPTIONS

- Like FIML, Bayesian estimation gives consistent estimates (unbiased in large samples) if the process is conditionally MAR
- FIML frequentist inference further assumes that the missingness process replicates across different random samples from the population (missing *always* at random)
- The conditionally MAR assumption is untestable because it involves propositions about the unseen score values

INTRODUCTION TO BAYESIAN ESTIMATION

FREQUENTIST VS. BAYESIAN PARADIGMS

Frequentist

- The parameter is a fixed quantity, estimates vary across different samples
- Statements about probability, precision, and confidence refer to estimates
- Probability = long run frequency of outcomes across many samples

Bayesian

- Parameters are random variables with a distribution of plausible realizations
- Statements about probability, precision, and intervals refer to the parameter
- Probability = our degree of certainty about a parameter after analyzing data

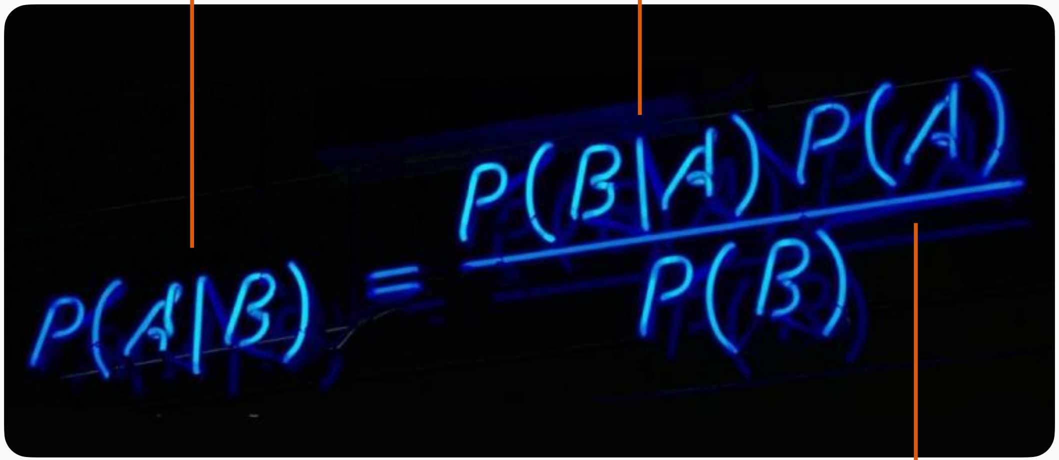
PARAMETERS AS VARIABLES

- Parameters are unknown variables, some realizations (values) are more plausible than others, given our data
- Bayes' theorem is the mathematical tool that converts fixed parameters (frequentist) into variable ones (Bayesian)
- Bayes' theorem is a rule for obtaining conditional probabilities

BAYES' THEOREM

Posterior = parameters (A) given the data (B)

Likelihood = data (B) given the parameters (A)



Prior = a priori belief about parameters (A)

PRIOR DISTRIBUTIONS

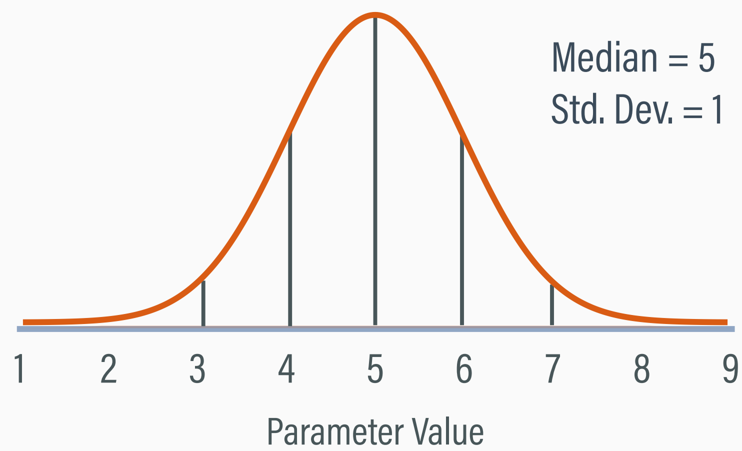
- Bayesian analyses require prior distributions that encode our beliefs about the parameter values prior to analyzing the data
- Blimp adopts non-informative (diffuse) priors that impart as little information as possible (i.e., let the data do the talking)
- e.g., A diffuse prior for means and coefficients conveys that all possible parameter values are equally likely (a flat distribution)

POSTERIOR DISTRIBUTIONS

- The prior and likelihood function as two data sources that merge to define the posterior distribution of the parameters
- The posterior describes a distribution of plausible parameter values that could have produced our particular data set
- Instead of estimates varying around a fixed parameter (frequentist), parameters vary around a fixed data set

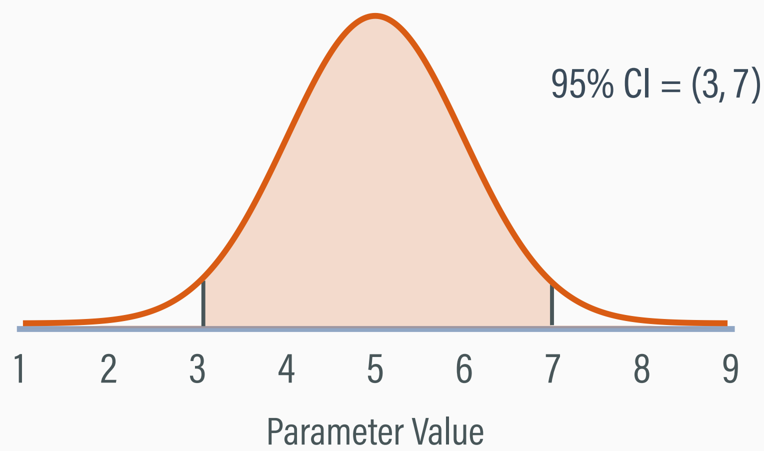
MEDIAN AND STANDARD DEVIATION

- The posterior median and standard deviation quantify the most likely parameter value and uncertainty
- Analogous to a point estimate and standard error, sans repeated sampling



95% CREDIBLE INTERVALS

- The 95% credible interval gives limits spanning 95% of the parameter's range
- Akin to a confidence interval, but references a range of highly plausible parameter values



AERA WORKSHOP DATA

Predictors

Outcome

Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
ESL	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

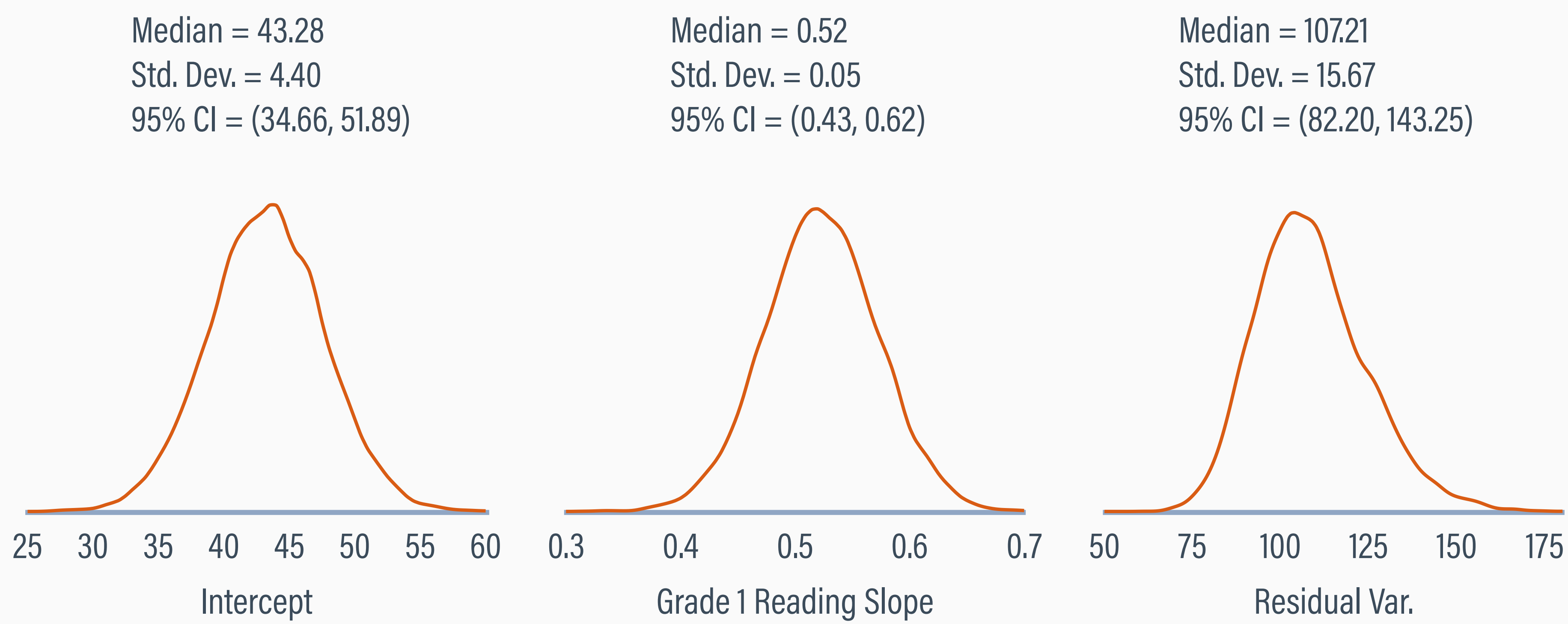
SIMPLE REGRESSION ILLUSTRATION

- Study that seeks to determine whether reading levels in 1st grade predict 9th grade reading achievement in middle school

$$READ_9 = \beta_0 + \beta_1(READ_1) + \varepsilon$$

- Both reading tests have missing values to be imputed

POSTERIOR DISTRIBUTIONS



ESTIMATOR COMPARISON

The two estimators are numerically equivalent!!!

	Bayes			FIML		
Parameter	Median	SD	95% CI	Est.	SE	95% CI
Intercept	43.28	4.40	(34.66, 51.89)	43.16	4.28	(34.78, 51.54)
1st Grade Reading	0.52	0.05	(0.43, 0.62)	0.52	0.05	(0.43, 0.62)
Residual variance	107.21	15.67	(82.20, 143.25)	102.59	14.19	(74.76, 140.40)
R^2	.50	.06	(.37, .61)	.49	.07	(.35, .63)

MARKOV CHAIN MONTE CARLO (MCMC)

- A Bayesian analysis involves estimating and sampling from distributions of model parameters and the missing values
- MCMC breaks a complex problem involving multiple unknowns (parameters and missing values) into separate steps
- Each step estimates one unknown at a time, treating the current values of all other quantities as known constants

MCMC ESTIMATION

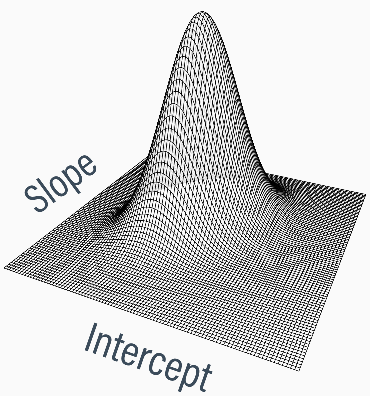
- Do for $t = 1$ to T iterations
  - Estimate focal model parameters, conditional on filled-in data
  - Estimate predictor model parameters, conditional on filled-in data
  - Impute outcome scores, conditional on the focal model parameters
  - Impute predictors, conditional on the focal and predictor model parameters
- Repeat

MEANING OF ESTIMATION

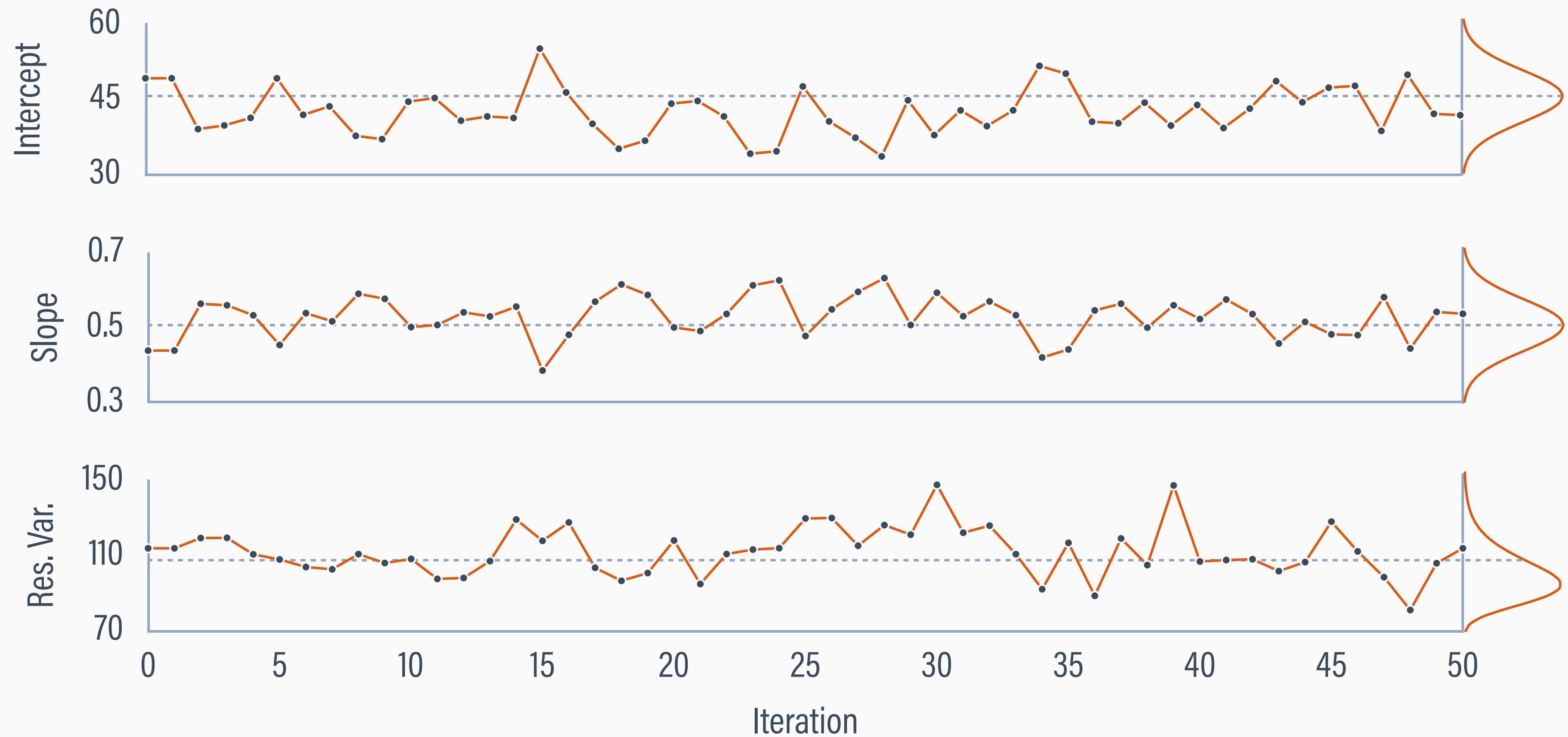
- A posterior is a distribution of plausible parameter values that could have produced our particular data set
- MCMC uses computer simulation (random number generation) to "sample" plausible parameter values
- Parameter values and missing data imputations are in a constant state of flux across a long MCMC sequence

PARAMETER-GENERATING DISTRIBUTIONS

- MCMC draws coefficients from a multivariate normal distribution, with OLS estimates defining its shape
- MCMC draws variances from an inverse gamma distribution with its shape determined by the df and residual SS



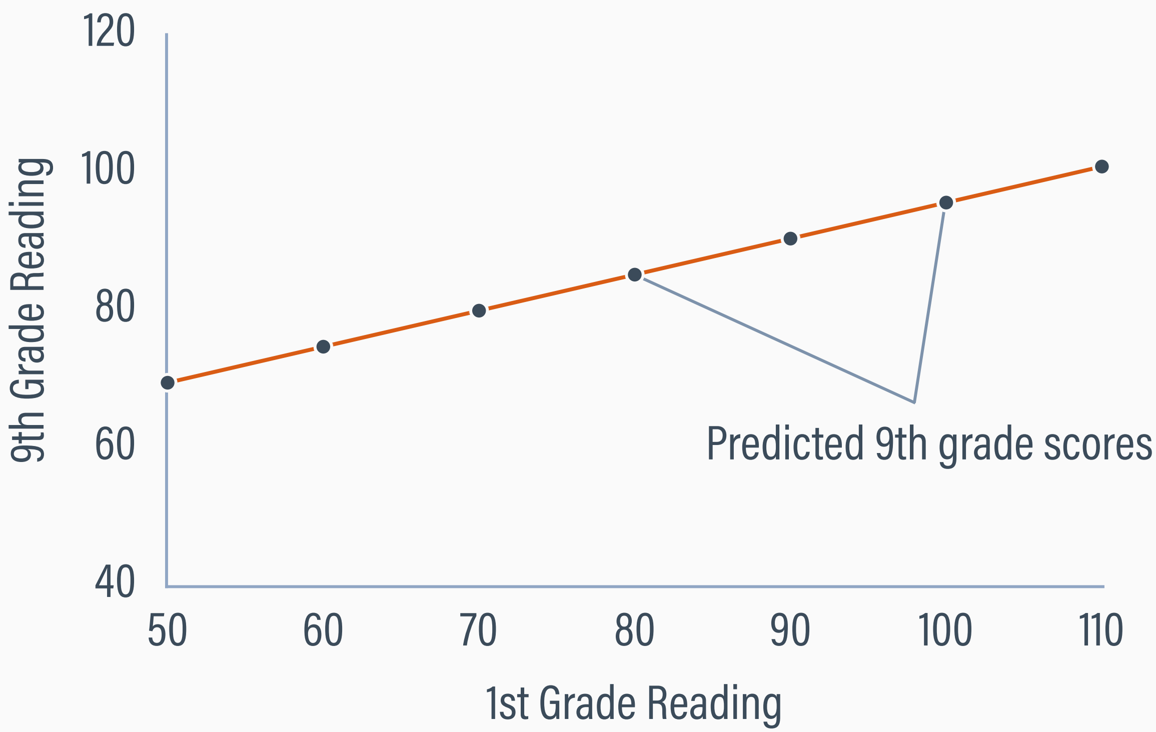
PARAMETERS FROM 50 MCMC CYCLES



MISSING DATA IMPUTATION

- Missing scores are imputed by drawing replacement scores at random from a distribution of plausible values
- Each unique set of parameter values combine to define the center and spread of the imputation distributions
- Imputing predictors is more complex than imputing outcomes

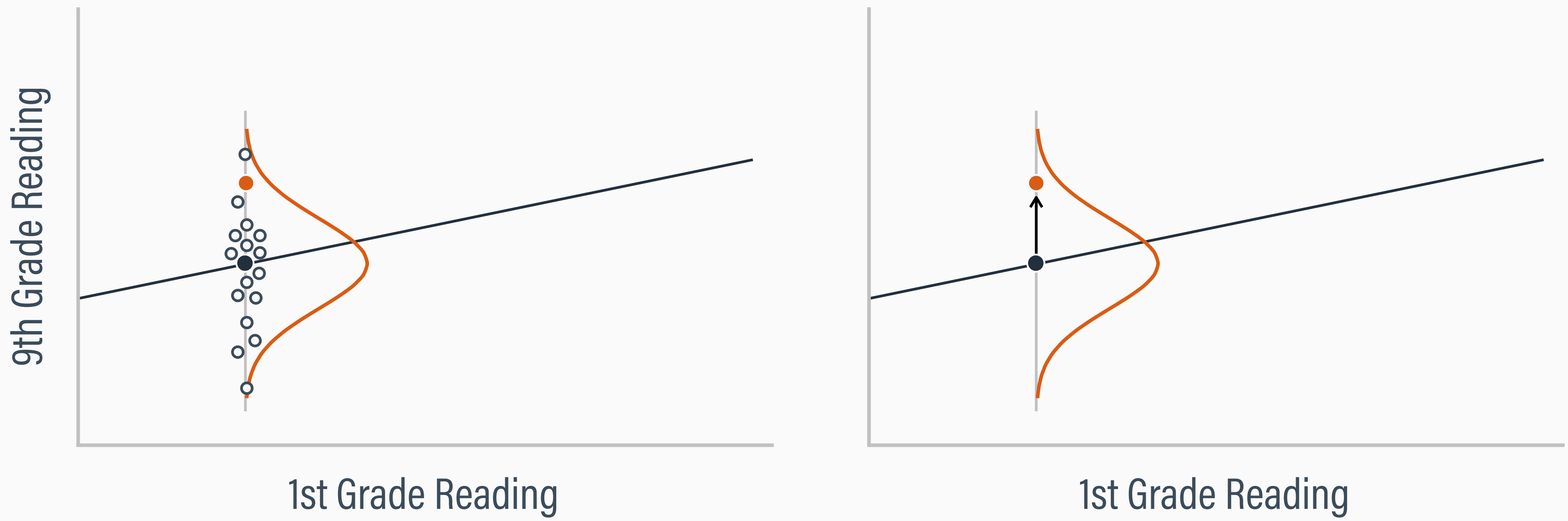
PREDICTED OUTCOME SCORES



DISTRIBUTIONS OF IMPUTATIONS

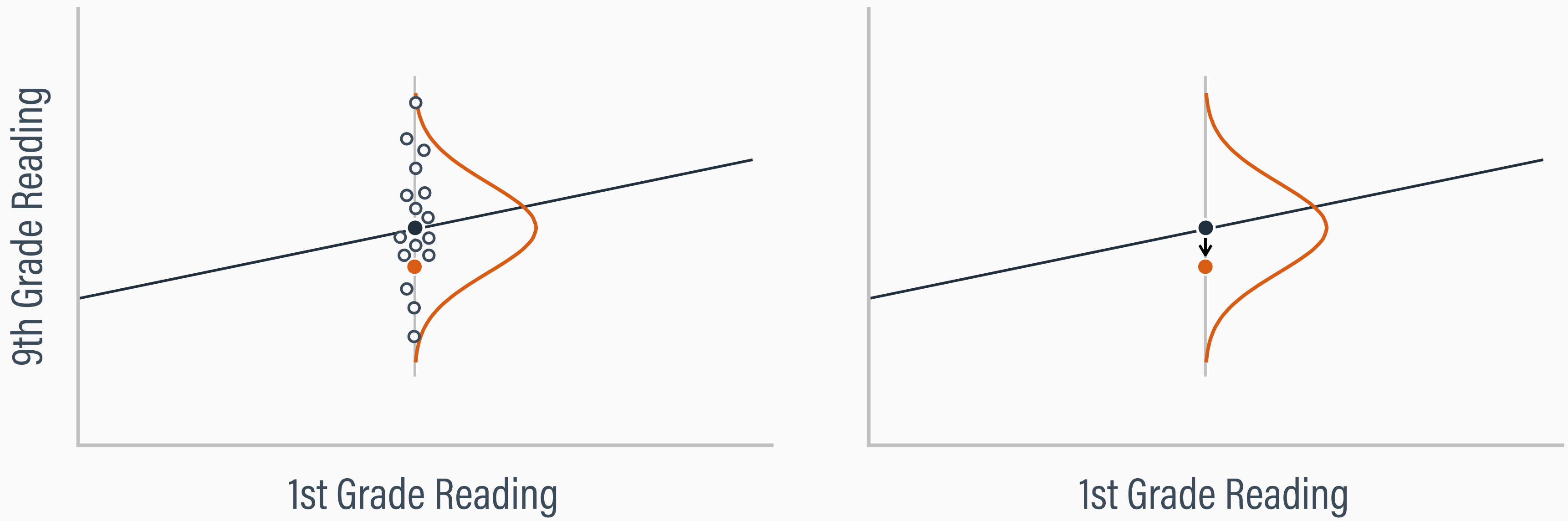
SAMPLING AN IMPUTATION

Imputation = predicted value + random normal noise



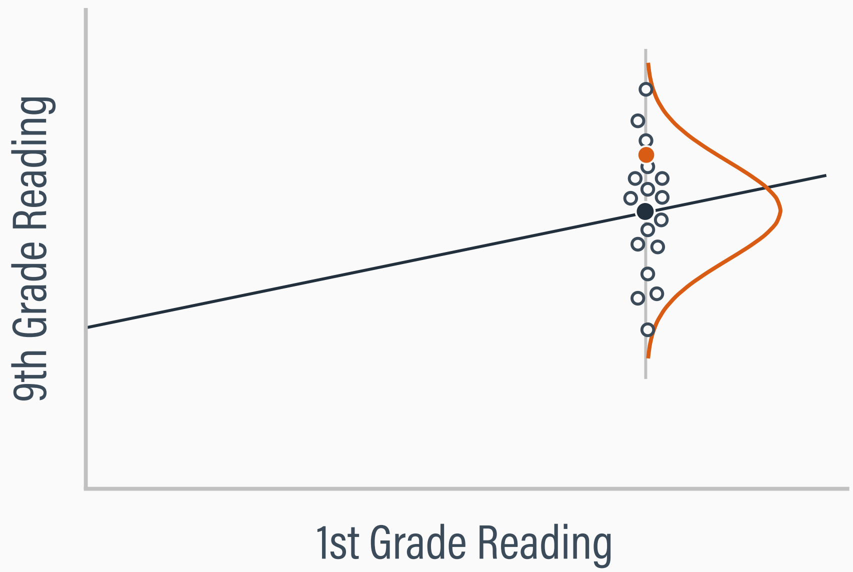
SAMPLING AN IMPUTATION

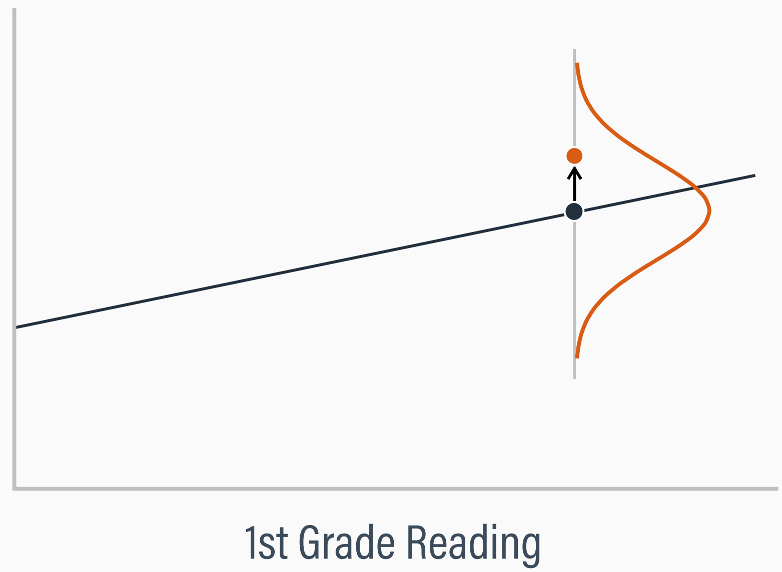
Imputation = predicted value + random normal noise



SAMPLING AN IMPUTATION

Imputation = predicted value + random normal noise



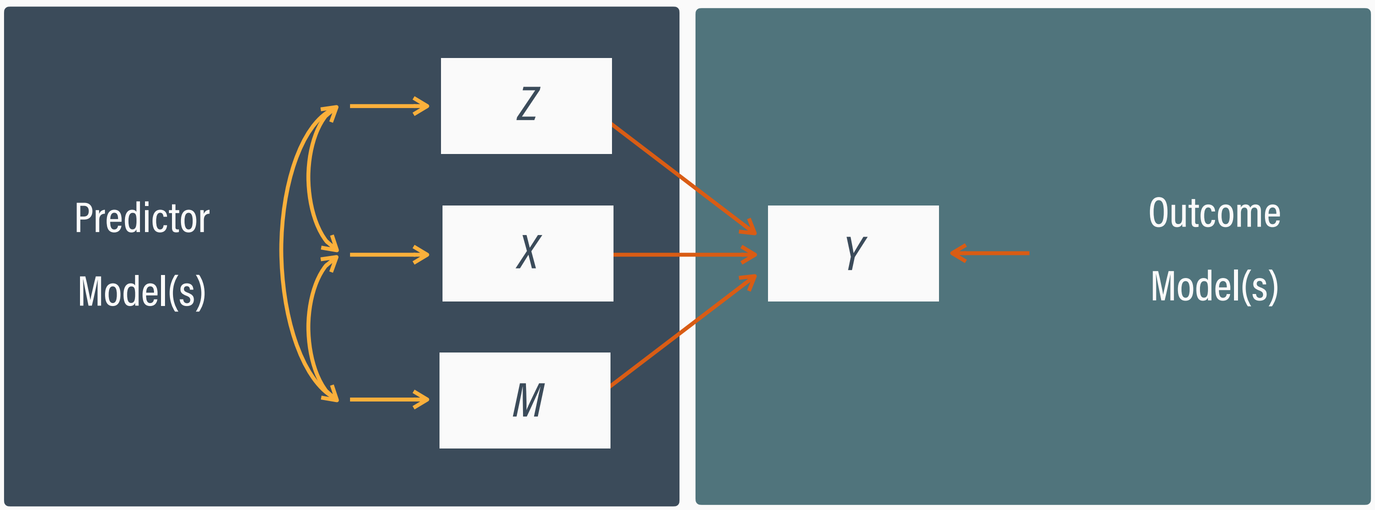


INCOMPLETE PREDICTORS

- Incomplete predictors require a model and distribution
- Multivariate normal methods (e.g., FIML) can mis-specify the data distributions in a way that introduces bias
- Factored regression uses a modular specification where a sequence of models replaces a general multivariate model

FACTORED REGRESSION OVERVIEW

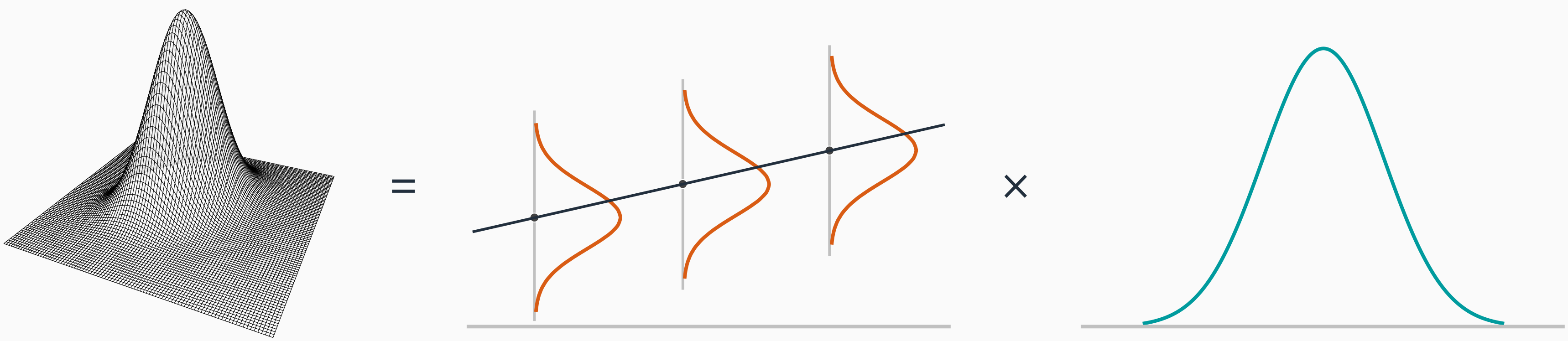
- Factored regression specifications invoke separate models (and distributions) for incomplete predictors and outcomes



BIVARIATE FACTORED REGRESSION

Bivariate Distribution = Univariate Outcome Model × Univariate Predictor Model

$$f(Y,X) = f(Y|X) \times f(X)$$



SIMPLE REGRESSION EXAMPLE

Factorization

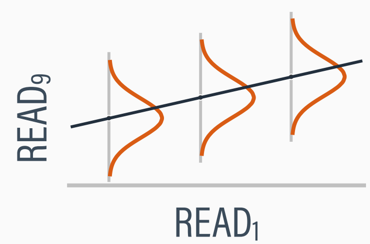
$$f(READ_1, READ_9) = f(READ_9|READ_1) \times f(READ_1)$$

Fitted models

$$READ_1 = \gamma_0 + \epsilon$$

$$READ_9 = \beta_0 + \beta_1(READ_1) + \varepsilon$$

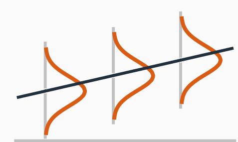
Distributions



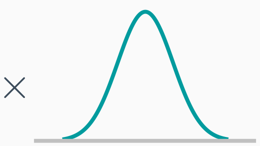
IMPUTING MISSING PREDICTORS

- A missing predictor always appears in two models: as a regressor in the focal model and an outcome in its own model

$$f(READ_1|READ_9) \propto f(READ_9|READ_1) \times f(READ_1) =$$



$$\times$$

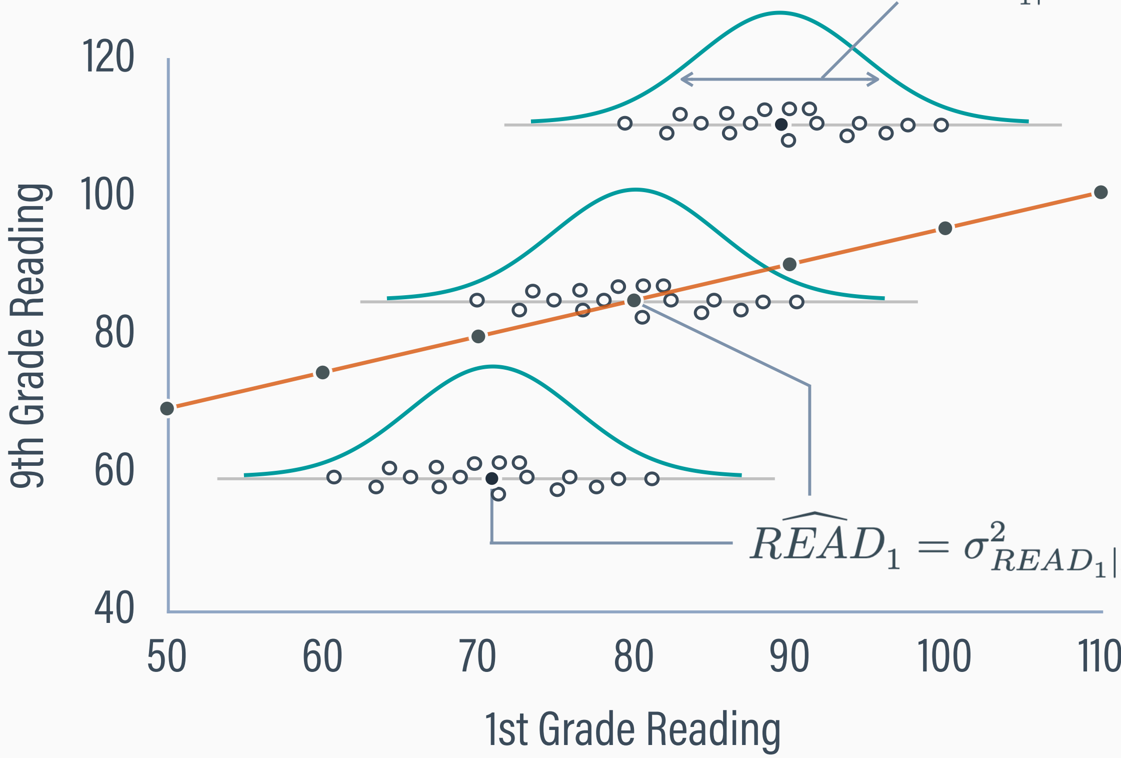


- The distribution of missing $READ_1$ scores is a composite function that depends on both model-implied distributions

DISTRIBUTIONS OF IMPUTATIONS

DISTRIBUTIONS OF IMPUTATIONS

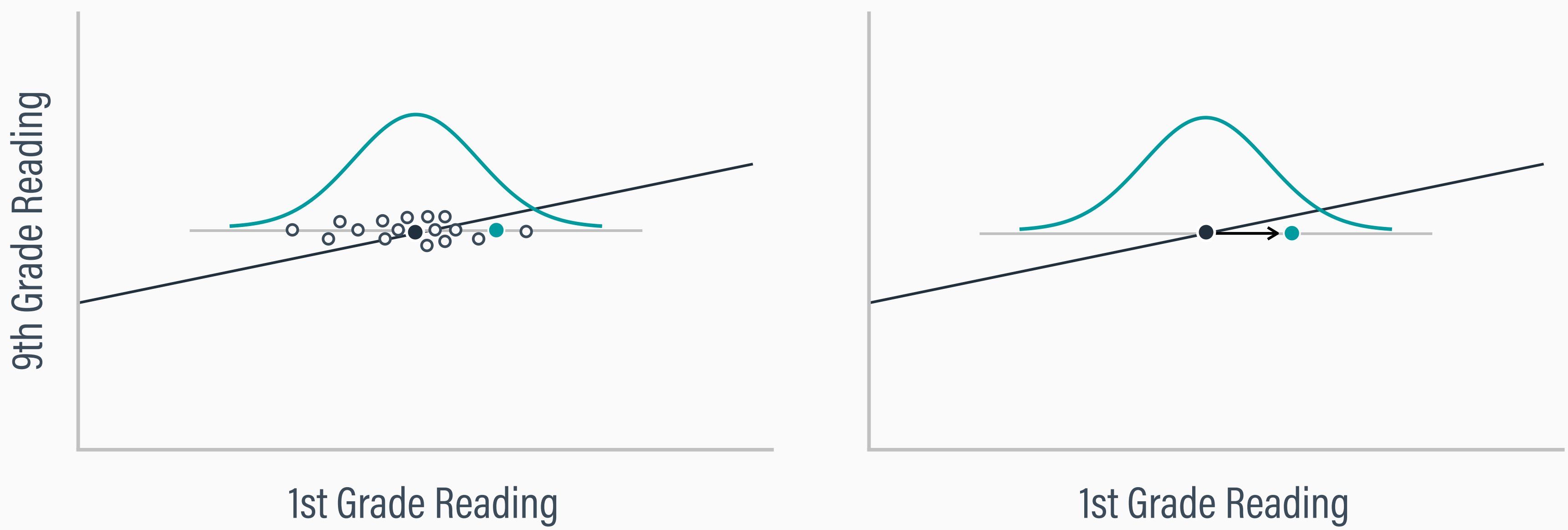
$$\sigma^2_{READ_1|READ_9} = \left(\frac{1}{\sigma^2_{\epsilon}} + \frac{\beta_1^{\,2}}{\sigma^2_{\varepsilon}}\right)^{-1}$$



$$\widehat{READ}_1 = \sigma^2_{READ_1|READ_9} \times \left(\frac{\gamma_0}{\sigma^2_{\epsilon}} + \frac{\beta_1(READ_9 - \beta_0)}{\sigma^2_{\varepsilon}}\right)$$

IMPUTATION EXAMPLE

Imputation = predicted value + random normal noise



FITTING REGRESSION MODELS IN BLIMP

AERA WORKSHOP DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
ESL	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

ANALYSIS MODEL

- Multiple regression where reading performance and teacher-rated learning problems in 1st grade predict 9th grade reading

$$READ_9 = \beta_0 + \beta_1(READ_1) + \beta_2(LRNPROB_1) + \beta_3(MALE) + \varepsilon$$

- Both reading achievement tests have missing values

BLIMP SCRIPT

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male;
BURN: 1000;
ITERATIONS: 10000;
SEED: 90291;
```

DATA AND VARIABLES

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male;
BURN: 1000;
ITERATIONS: 10000;
SEED: 90291;
```

MODEL DETAILS

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male;
BURN: 1000;
ITERATIONS: 10000;
SEED: 90291;
```

COMPUTATIONAL DETAILS

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male;
BURN: 1000;
ITERATIONS: 10000;
SEED: 90291;
```

DATA AND VARIABLES

```
DATA: aeraworkshop.dat;                          # ascii text data
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9
   read9grp stanread7;                           # variable order
MISSING: 999;                                    # missing value code
NOMINAL: male;                                   # binary or multicategorial predictor
```

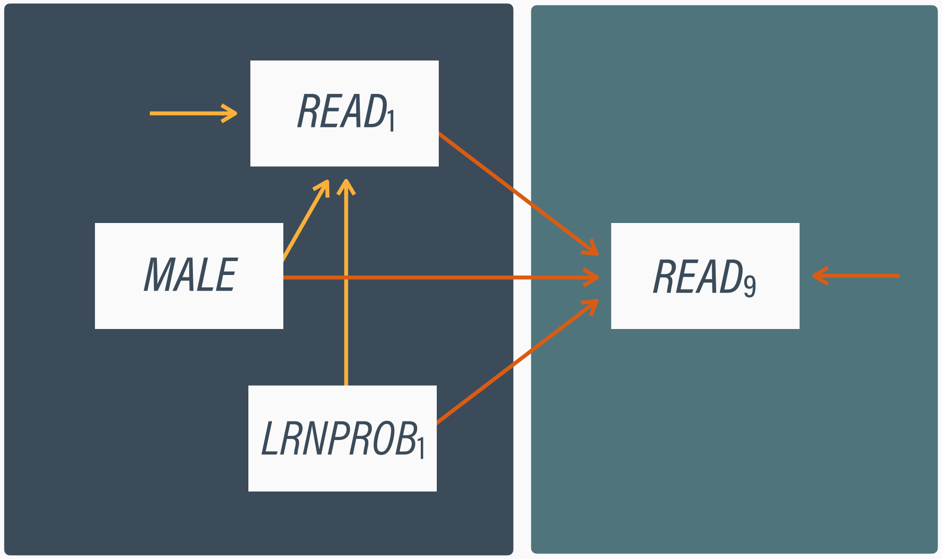
MODEL DETAILS

```
FIXED: male lrnprob1;                    # complete predictors
MODEL:                                   # regression equations
read9 ~ read1 lrnprob1 male;             # DV ~ predictors
```

PATH DIAGRAM AND MODEL COMMAND

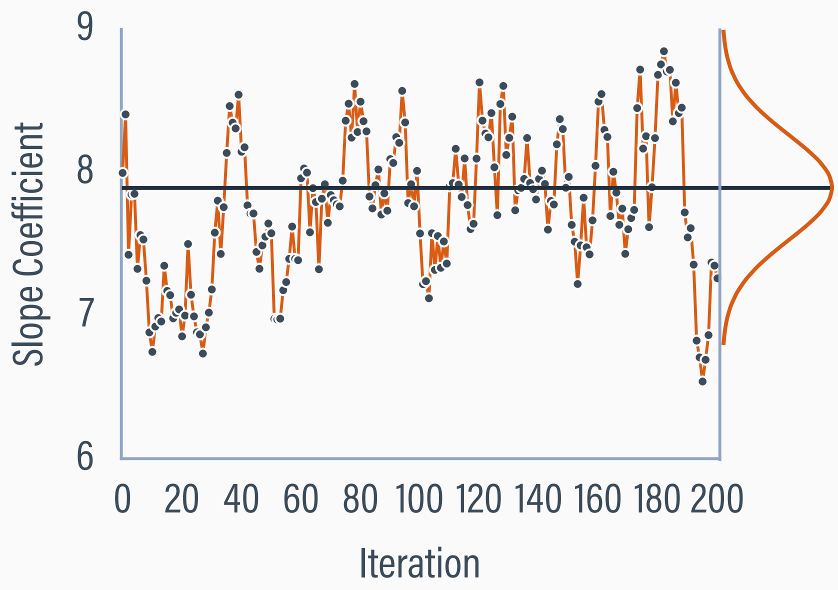
Multivariate Distribution = Mutivariate Predictor Distributions × Univariate Outcome Distribution

FIXED: male;
MODEL:
read9 ~ read1 lrnprob1 male;

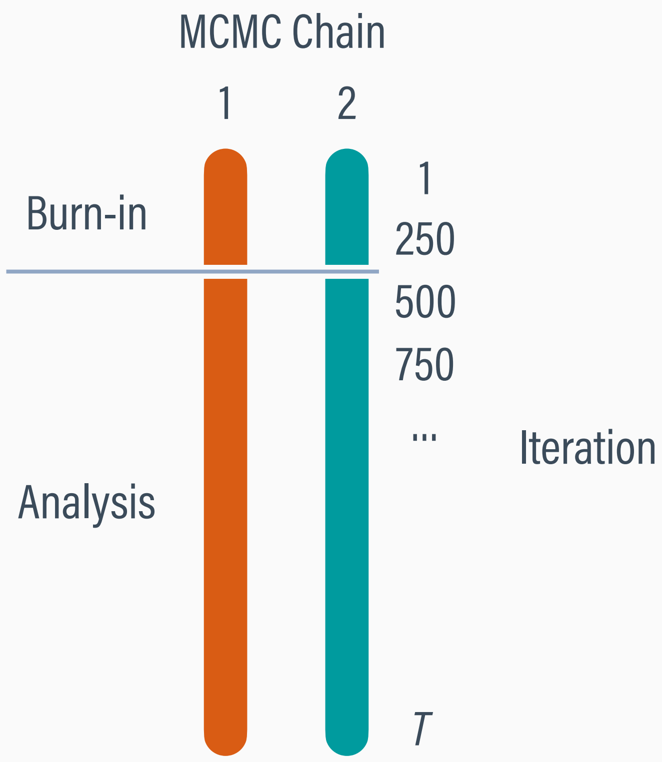


MCMC ESTIMATION

- MCMC uses computer simulation to "draw" or "sample" parameters from a distribution of plausible values
- Estimates continually vary across iterations in a random pattern

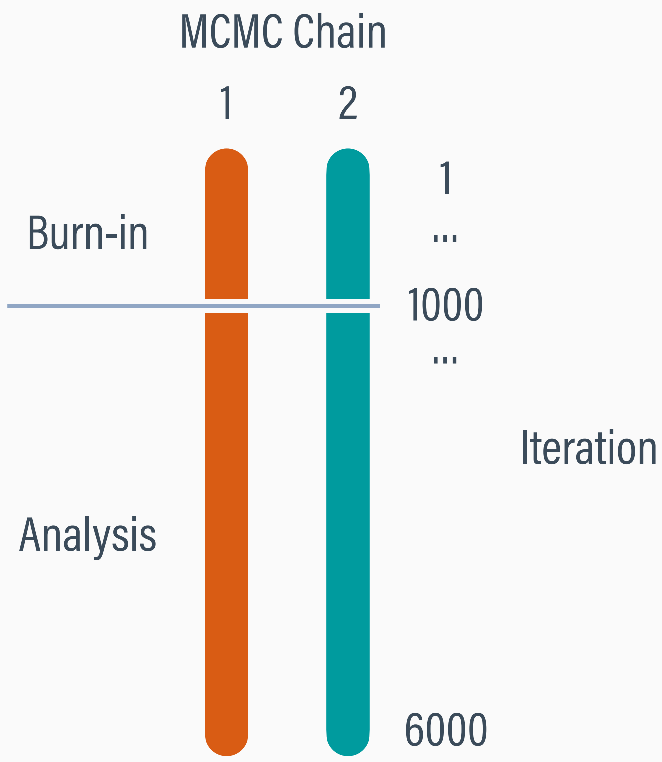


MCMC ALGORITHM



COMPUTATIONAL DETAILS

```
BURN: 1000;            # burn-in iterations
ITERATIONS: 10000;     # analysis iterations
SEED: 90291;           # random number seed
```



UNDERSTANDING BLIMP OUTPUT

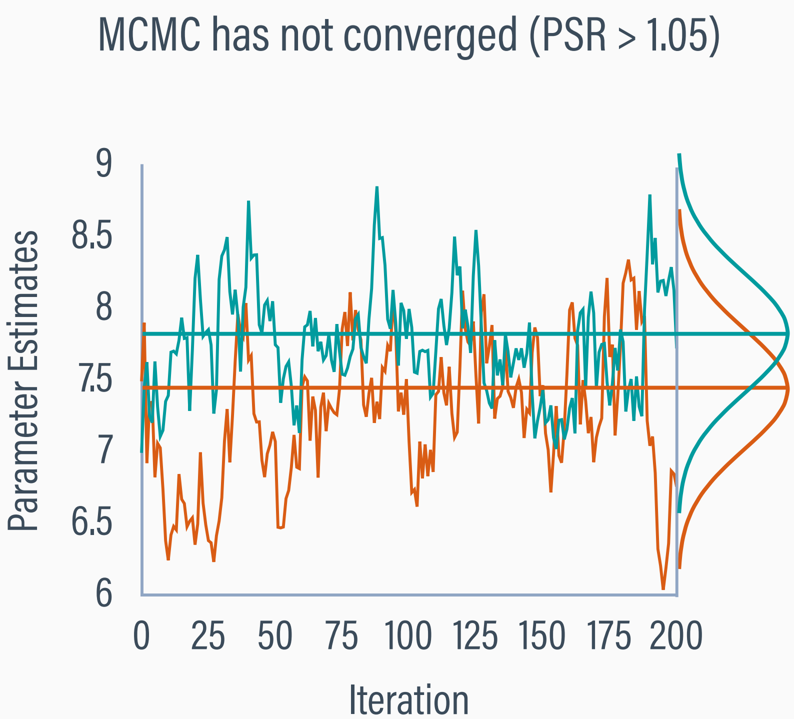
MCMC CONVERGENCE

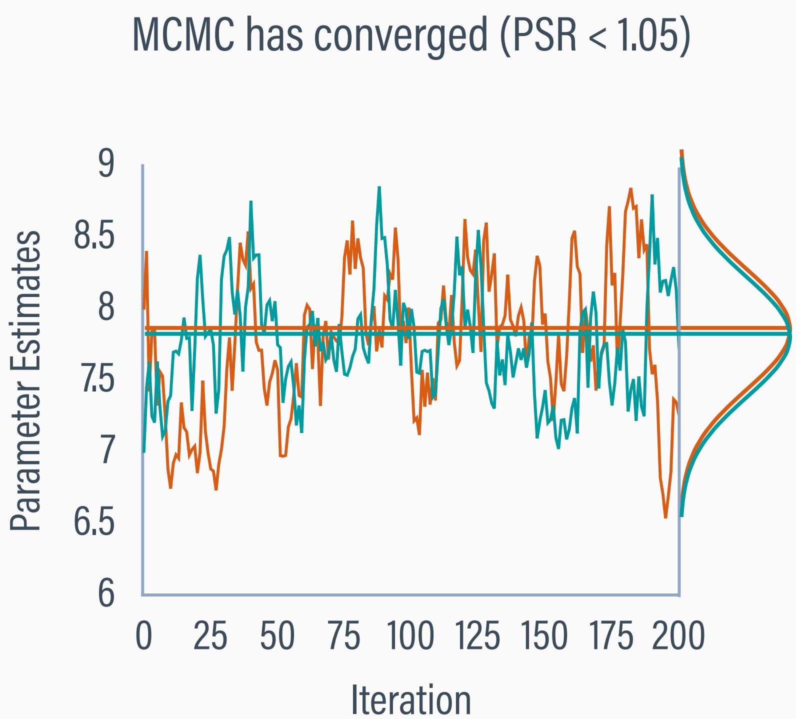
- MCMC converges when posterior distributions are stationary
- Parameter estimates oscillate around a stable mean, and variation doesn't change with additional iterations
- Set burn-in cycles > number of iterations needed to converge

POTENTIAL SCALE REDUCTION

- The potential scale reduction (PSR) factor compares parameter distributions generated from two unique MCMC processes
- MCMC converges when the two chains give estimates with same mean and spread
- PSRs for all parameters should be < 1.05

PSR GRAPHIC


MCMC has not converged (PSR > 1.05)
Parameter Estimates
9
8.5
8
7.5
7
6.5
6
0
25
50
75
100
125
150
175
200
Iteration


MCMC has converged (PSR < 1.05)
Parameter Estimates
9
8.5
8
7.5
7
6.5
6
0
25
50
75
100
125
150
175
200
Iteration

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE:  Split chain PSR is being used. This splits each chain's
         iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                        26  to 50                1.130             13
                        51  to 100               1.086              7
                        76  to 150               1.042              1
                       101  to 200               1.040              8
                        ...      ...                ...              ...
                       401  to 800               1.008              1
                       426  to 850               1.006              1
                       451  to 900               1.008              1
                       476  to 950               1.008              5
                       501  to 1000              1.008              8
```

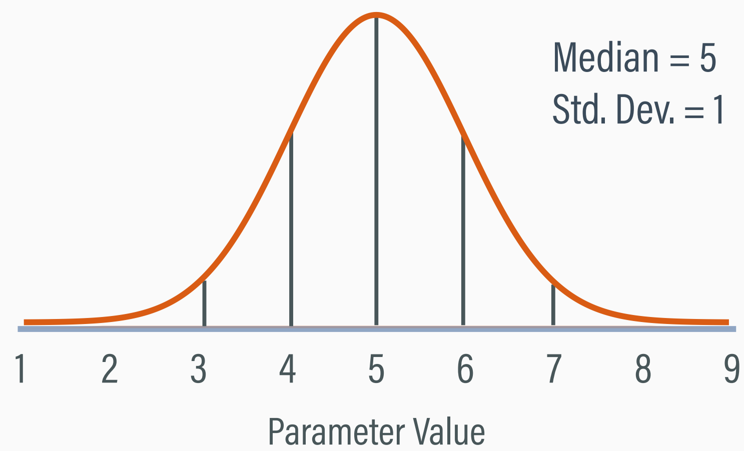
» Worst PSR < 1.05

BAYESIAN POSTERIOR SUMMARIES

- Probability distributions are tools for expressing our knowledge about a parameter in a Bayesian analysis
- The posterior distribution describes plausible parameter values that are consistent with the data (no repeated sampling)
- We use summary statistics to describe parameter distributions

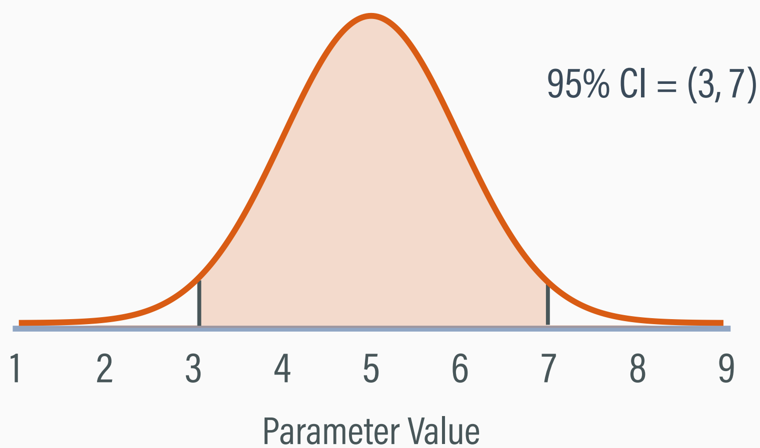
MEDIAN AND STANDARD DEVIATION

- The posterior median and standard deviation quantify the most likely parameter value and uncertainty
- Analogous to a point estimate and standard error, sans repeated sampling

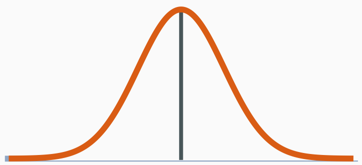


95% CREDIBLE INTERVALS

- The 95% credible interval gives limits spanning 95% of the parameter's range
- Akin to a confidence interval, but references a range of highly plausible parameter values



POINT ESTIMATES



Parameter Values

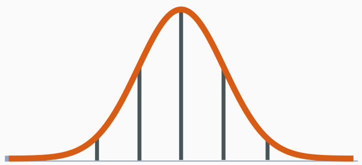
OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: read9

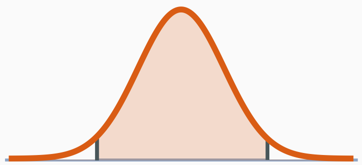
Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	91.781	13.202	70.514	123.094	1.000	5084.608
Coefficients:						
Intercept	62.976	7.140	48.902	76.899	1.001	4079.222
read1	0.516	0.048	0.421	0.611	1.000	6397.445
lrnprob1	-0.398	0.095	-0.587	-0.212	1.000	3457.935
male.1	1.887	1.946	-1.925	5.703	1.000	5824.649
Standardized Coefficients:						
read1	0.696	0.046	0.597	0.775	1.000	4649.513
lrnprob1	-0.293	0.067	-0.420	-0.158	1.001	3354.968
male.1	0.064	0.066	-0.065	0.192	1.000	5801.597
Proportion Variance Explained						
by Coefficients	0.575	0.052	0.462	0.663	1.000	5842.932
by Residual Variation	0.425	0.052	0.337	0.538	1.000	5842.932

MEASURES OF UNCERTAINTY




```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:   read9

Parameters                        Median    StdDev      2.5%     97.5%       PSR     N_Eff
                               -----------------------------------------------------------------
Variances:
  Residual Var.                   91.781    13.202    70.514   123.094     1.000  5084.608

Coefficients:
  Intercept                       62.976     7.140    48.902    76.899     1.001  4079.222
  read1                            0.516     0.048     0.421     0.611     1.000  6397.445
  lrnprob1                        -0.398     0.095    -0.587    -0.212     1.000  3457.935
  male.1                           1.887     1.946    -1.925     5.703     1.000  5824.649

Standardized Coefficients:
  read1                            0.696     0.046     0.597     0.775     1.000  4649.513
  lrnprob1                        -0.293     0.067    -0.420    -0.158     1.001  3354.968
  male.1                           0.064     0.066    -0.065     0.192     1.000  5801.597

Proportion Variance Explained
  by Coefficients                  0.575     0.052     0.462     0.663     1.000  5842.932
  by Residual Variation            0.425     0.052     0.337     0.538     1.000  5842.932

                               -----------------------------------------------------------------
```

95% CREDIBLE INTERVALS



Parameter Values

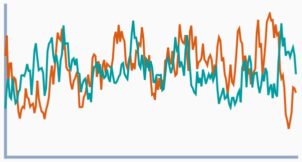
OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **read9**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	91.781	13.202	70.514	123.094	1.000	5084.608
Coefficients:						
Intercept	62.976	7.140	48.902	76.899	1.001	4079.222
read1	0.516	0.048	0.421	0.611	1.000	6397.445
lrnprob1	-0.398	0.095	-0.587	-0.212	1.000	3457.935
male.1	1.887	1.946	-1.925	5.703	1.000	5824.649
Standardized Coefficients:						
read1	0.696	0.046	0.597	0.775	1.000	4649.513
lrnprob1	-0.293	0.067	-0.420	-0.158	1.001	3354.968
male.1	0.064	0.066	-0.065	0.192	1.000	5801.597
Proportion Variance Explained						
by Coefficients	0.575	0.052	0.462	0.663	1.000	5842.932
by Residual Variation	0.425	0.052	0.337	0.538	1.000	5842.932

DIAGNOSTICS



OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **read9**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	91.781	13.202	70.514	123.094	1.000	5084.608
Coefficients:						
Intercept	62.976	7.140	48.902	76.899	1.001	4079.222
read1	0.516	0.048	0.421	0.611	1.000	6397.445
lrnprob1	-0.398	0.095	-0.587	-0.212	1.000	3457.935
male.1	1.887	1.946	-1.925	5.703	1.000	5824.649
Standardized Coefficients:						
read1	0.696	0.046	0.597	0.775	1.000	4649.513
lrnprob1	-0.293	0.067	-0.420	-0.158	1.001	3354.968
male.1	0.064	0.066	-0.065	0.192	1.000	5801.597
Proportion Variance Explained						
by Coefficients	0.575	0.052	0.462	0.663	1.000	5842.932
by Residual Variation	0.425	0.052	0.337	0.538	1.000	5842.932

VARIANCES AND COVARIANCES

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **read9**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	91.781	13.202	70.514	123.094	1.000	5084.608
Coefficients:						
Intercept	62.976	7.140	48.902	76.899	1.001	4079.222
read1	0.516	0.048	0.421	0.611	1.000	6397.445
lrnprob1	-0.398	0.095	-0.587	-0.212	1.000	3457.935
male.1	1.887	1.946	-1.925	5.703	1.000	5824.649
Standardized Coefficients:						
read1	0.696	0.046	0.597	0.775	1.000	4649.513
lrnprob1	-0.293	0.067	-0.420	-0.158	1.001	3354.968
male.1	0.064	0.066	-0.065	0.192	1.000	5801.597
Proportion Variance Explained						
by Coefficients	0.575	0.052	0.462	0.663	1.000	5842.932
by Residual Variation	0.425	0.052	0.337	0.538	1.000	5842.932

COEFFICIENTS

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------
Variances:
  Residual Var.                     91.781     13.202     70.514    123.094      1.000   5084.608

Coefficients:
  Intercept                         62.976      7.140     48.902     76.899      1.001   4079.222
  read1                              0.516      0.048      0.421      0.611      1.000   6397.445
  lrnprob1                          -0.398      0.095     -0.587     -0.212      1.000   3457.935
  male.1                             1.887      1.946     -1.925      5.703      1.000   5824.649

Standardized Coefficients:
  read1                              0.696      0.046      0.597      0.775      1.000   4649.513
  lrnprob1                          -0.293      0.067     -0.420     -0.158      1.001   3354.968
  male.1                             0.064      0.066     -0.065      0.192      1.000   5801.597

Proportion Variance Explained
  by Coefficients                    0.575      0.052      0.462      0.663      1.000   5842.932
  by Residual Variation              0.425      0.052      0.337      0.538      1.000   5842.932

                                ------------------------------------------------------------------
```

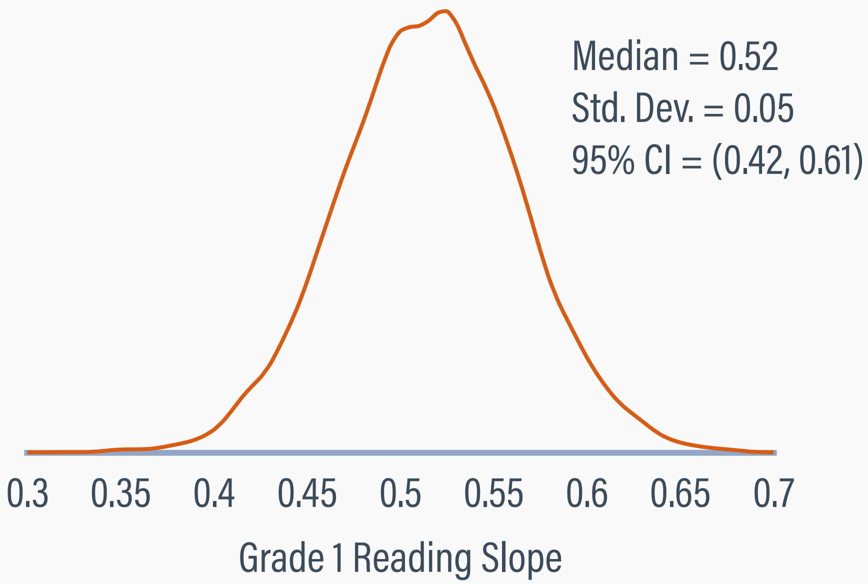
STANDARDIZED COEFFICIENTS

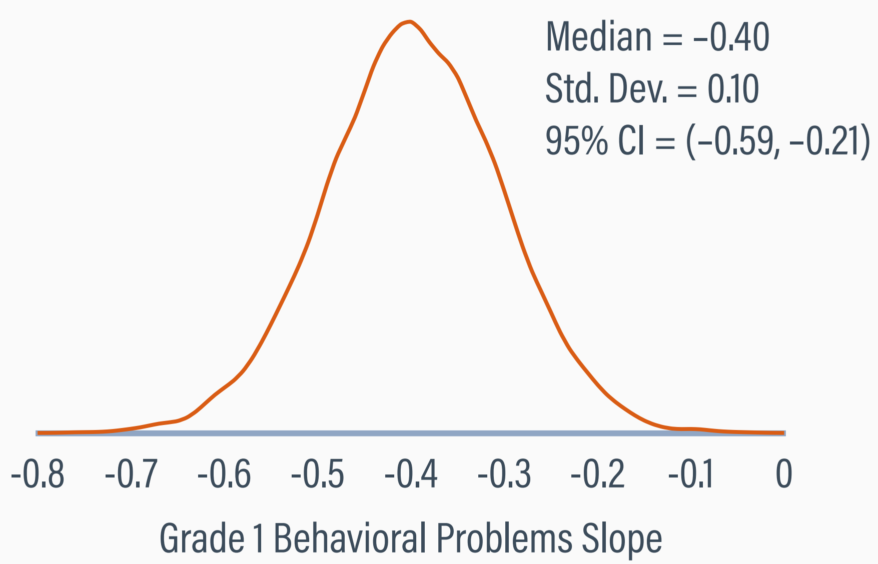
```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Parameters                         Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------

Variances:
  Residual Var.                    91.781     13.202     70.514    123.094      1.000   5084.608

Coefficients:
  Intercept                        62.976      7.140     48.902     76.899      1.001   4079.222
  read1                             0.516      0.048      0.421      0.611      1.000   6397.445
  lrnprob1                         -0.398      0.095     -0.587     -0.212      1.000   3457.935
  male.1                            1.887      1.946     -1.925      5.703      1.000   5824.649

Standardized Coefficients:
  read1                             0.696      0.046      0.597      0.775      1.000   4649.513
  lrnprob1                         -0.293      0.067     -0.420     -0.158      1.001   3354.968
  male.1                            0.064      0.066     -0.065      0.192      1.000   5801.597

Proportion Variance Explained
  by Coefficients                   0.575      0.052      0.462      0.663      1.000   5842.932
  by Residual Variation             0.425      0.052      0.337      0.538      1.000   5842.932

                                ------------------------------------------------------------------
```

EFFECT SIZE ESTIMATES

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:   read9

Parameters                       Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------

Variances:
  Residual Var.                  91.781     13.202     70.514    123.094      1.000   5084.608

Coefficients:
  Intercept                      62.976      7.140     48.902     76.899      1.001   4079.222
  read1                           0.516      0.048      0.421      0.611      1.000   6397.445
  lrnprob1                       -0.398      0.095     -0.587     -0.212      1.000   3457.935
  male.1                          1.887      1.946     -1.925      5.703      1.000   5824.649

Standardized Coefficients:
  read1                           0.696      0.046      0.597      0.775      1.000   4649.513
  lrnprob1                       -0.293      0.067     -0.420     -0.158      1.001   3354.968
  male.1                          0.064      0.066     -0.065      0.192      1.000   5801.597

Proportion Variance Explained
  by Coefficients                 0.575      0.052      0.462      0.663      1.000   5842.932
  by Residual Variation           0.425      0.052      0.337      0.538      1.000   5842.932

                                ------------------------------------------------------------------
```

POSTERIOR DISTRIBUTIONS





COEFFICIENTS

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ---------------------------------------------------------------------
Variances:
  Residual Var.                       91.781     13.202     70.514    123.094      1.000   5084.608

Coefficients:
  Intercept                           62.976      7.140     48.902     76.899      1.001   4079.222
  read1                                0.516      0.048      0.421      0.611      1.000   6397.445
  lrnprob1                            -0.398      0.095     -0.587     -0.212      1.000   3457.935
  male.1                               1.887      1.946     -1.925      5.703      1.000   5824.649

Standardized Coefficients:
  ...
                                ---------------------------------------------------------------------
```

READ1 partial regression slope

LRNPROB1 partial regression slope

MALE partial regression slope

INTERPRETATION EXAMPLE

- For two students who share the same gender and learning problems rating, scoring one point higher on the first grade reading test is associated with a 0.52 increase in grade 9
- The slope is significantly different from 0 because the null value falls outside the 95% credible interval limits (0.42, 0.61)

HOW MANY ITERATIONS?

- MCMC estimates are autocorrelated across iterations
- The effective number of MCMC samples estimates the number of independent estimates after removing autocorrelation
- Gelman et al. (2014, p. 267) recommend at least 100 independent MCMC samples per parameter

EFFECTIVE NUMBER OF MCMC SAMPLES

» Effective number of MCMC samples for every parameter is > 100

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ----------------------------------------------------------------
Variances:
  Residual Var.                     91.781     13.202     70.514    123.094      1.000   5084.608

Coefficients:
  Intercept                         62.976      7.140     48.902     76.899      1.001   4079.222
  read1                              0.516      0.048      0.421      0.611      1.000   6397.445
  lrnprob1                          -0.398      0.095     -0.587     -0.212      1.000   3457.935
  male.1                             1.887      1.946     -1.925      5.703      1.000   5824.649

Standardized Coefficients:
  read1                              0.696      0.046      0.597      0.775      1.000   4649.513
  lrnprob1                          -0.293      0.067     -0.420     -0.158      1.001   3354.968
  male.1                             0.064      0.066     -0.065      0.192      1.000   5801.597

Proportion Variance Explained
  by Coefficients                    0.575      0.052      0.462      0.663      1.000   5842.932
  by Residual Variation              0.425      0.052      0.337      0.538      1.000   5842.932

                                ----------------------------------------------------------------
```

ESTIMATOR COMPARISON

The two estimators are numerically equivalent!!!

	Bayes		FIML	
Parameter	Median	SD	Est.	SE
Intercept	62.98	7.14	62.56	6.92
1st grade reading	0.52	0.05	0.52	0.05
1st grade learning problems	−0.40	0.10	−0.39	0.09
Male dummy code	1.89	1.95	1.95	1.88
Residual variance	91.78	13.20	86.32	11.78
R^2	0.58	0.05	0.59	0.06

BAYESIAN WALD TEST

- Asparouhov and Muthén (2021) proposed a Bayesian Wald test that mimics familiar likelihood-based Wald tests

$$T = (\theta - \theta_0)'\Sigma_\theta^{-1}(\theta - \theta_0)$$

- *T* is the sum of squared standardized differences (chi-square metric) between the posterior means and null hypothesis

BLIMP SCRIPT: TEST STATISTICS

```
MODEL:
read9 ~ read1@beta1 lrnprob1@beta2 male@beta3;    # label coefficients with @
TEST: beta1:beta3 = 0;                            # omnibus test of three slopes
TEST: beta1 = 0;                                  # test of a single slope
TEST: beta2 = 0;                                  # test of a single slope
TEST: beta3 = 0;                                  # test of a single slope
```

MODEL FIT OUTPUT

```
MODEL FIT:

  INFORMATION CRITERIA
     …

  WALD TESTS (Asparouhov & Muthén, 2021)

  Test #1

    Full:
      [1]  read9 ~ Intercept read1@beta1 lrnprob1@beta2 male.1@beta3

    Restricted:
      [1]  read9 ~ Intercept read1@beta1 lrnprob1@beta2 male.1@beta3

    Constraints in Restricted:
      [1]  beta1 = 0
      [2]  beta2 = 0
      [3]  beta3 = 0


    Wald Statistic (Chi-Square)                 147.562
    Number of Parameters Tested (df)                  3
    Probability                                   0.000
```

MODEL FIT OUTPUT

```
MODEL FIT:

  INFORMATION CRITERIA
     ...

  WALD TESTS (Asparouhov & Muthén, 2021)
    ...

  Test #2

    Full:
      [1]  read9 ~ Intercept read1@beta1 lrnprob1@beta2 male.1@beta3

    Restricted:
      [1]  read9 ~ Intercept read1@beta1 lrnprob1@beta2 male.1@beta3

    Constraints in Restricted:
      [1]  beta1 = 0


    Wald Statistic (Chi-Square)                 115.118
    Number of Parameters Tested (df)                  1
    Probability                                   0.000
```

REPORTING TEMPLATE

We used Bayesian estimation in Blimp 3 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 1,000 iterations was sufficiently conservative. Based on this information, we used two MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

REPORTING TEMPLATE CONTINUED

Table 1 displays the posterior summaries from the analysis. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent (Enders, 2022).

APA TABLE

Table 1

Parameter Summary From the Bayesian Regression Analysis

Parameter	Median	*SD*	LCL	UCL
Intercept	62.98	7.14	48.90	76.90
1st grade reading	0.52	0.05	0.42	0.61
1st grade learning problems	-0.40	0.10	-0.59	-0.21
Male dummy code	1.89	1.95	-1.93	5.70
R^2	0.58	0.05	0.46	0.66

REPORTING TEMPLATE CONTINUED

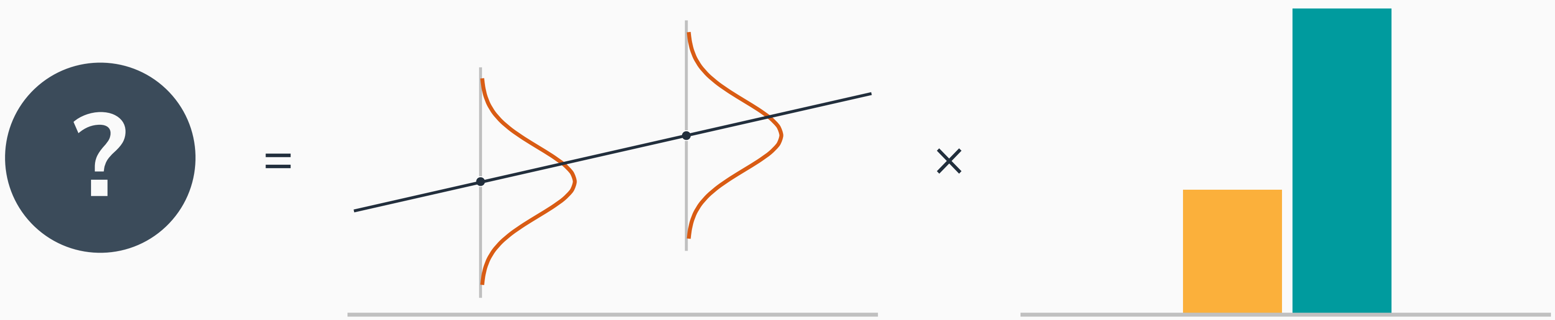
Collectively, the predictors explained approximately 58% of the variation in 9th grade reading scores. The Bayesian Wald test (Asparouhov & Muthén, 2021) of the full model was statistically significant, $\chi^2(3) = 147.56$, $p < .001$. First grade reading exhibited a significant positive association with 9th grade reading performance ($\beta = 0.52$, $SD = 0.05$, $p < .001$), and the measure of teacher-rated learning problems was inversely related to middle-school reading ($\beta = -0.40$, $SD = 0.10$, $p < .001$). Student gender was not predictive of 9th reading performance.

INCOMPLETE CATEGORICAL VARIABLES IN BLIMP

FACTORED REGRESSION

Bivariate Distribution = Univariate Outcome Distribution × Univariate Predictor Distribution

$$f(Y, X) = f(Y|X) \times f(X)$$



LATENT RESPONSE FORMULATION

AERA WORKSHOP DATA

Predictors
Outcome

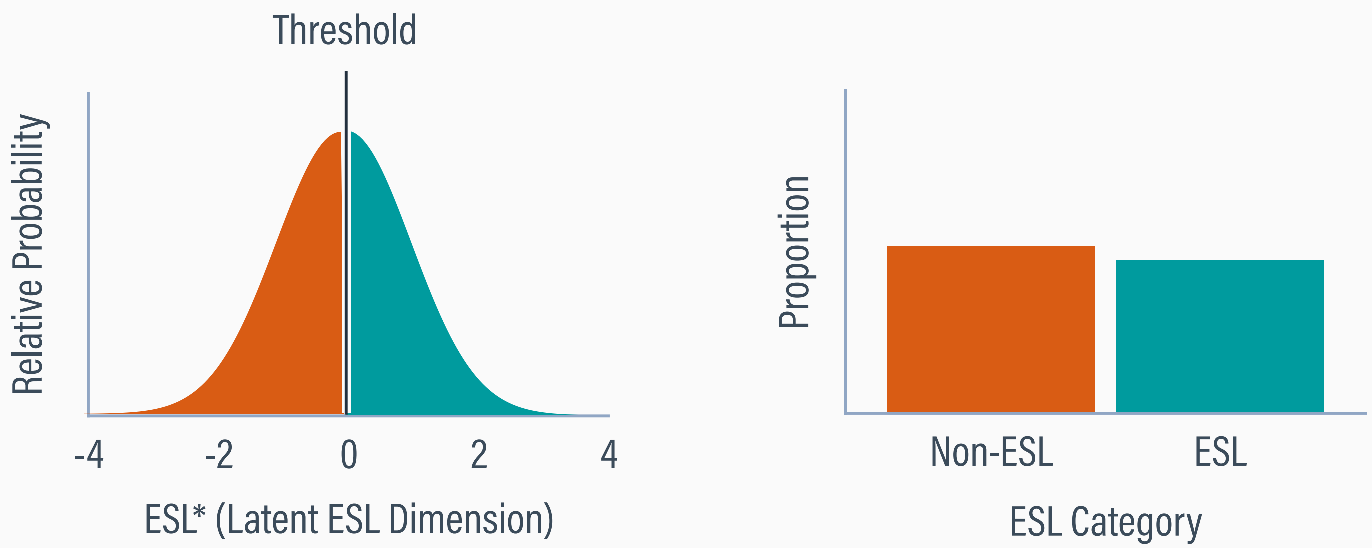
Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
» *ESL*	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

INCOMPLETE BINARY VARIABLES

- Probit regression envisions binary and ordinal variables arising from an underlying normal latent response variable
- Applied to the ESL, the latent variable represents an unobserved, continuous propensity for ESL
- A threshold carves the latent distribution into segments

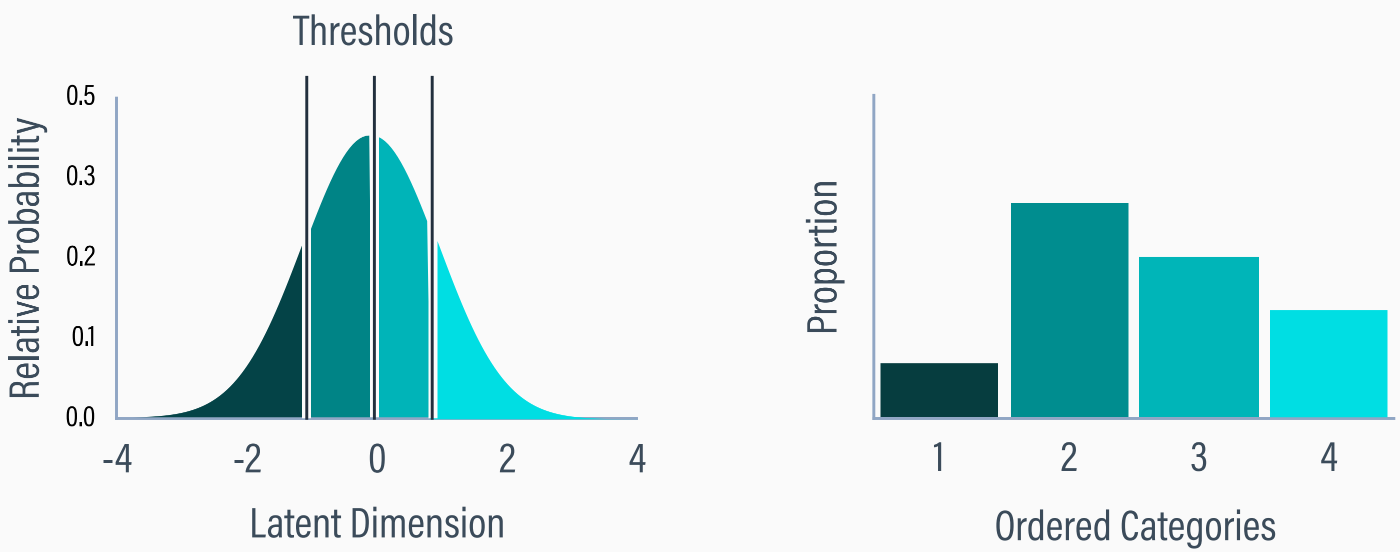
LATENT ESL DISTRIBUTION

- The threshold parameter divides the latent distribution into segments, with areas under the curve matching the bar plot



ORDINAL VARIABLES

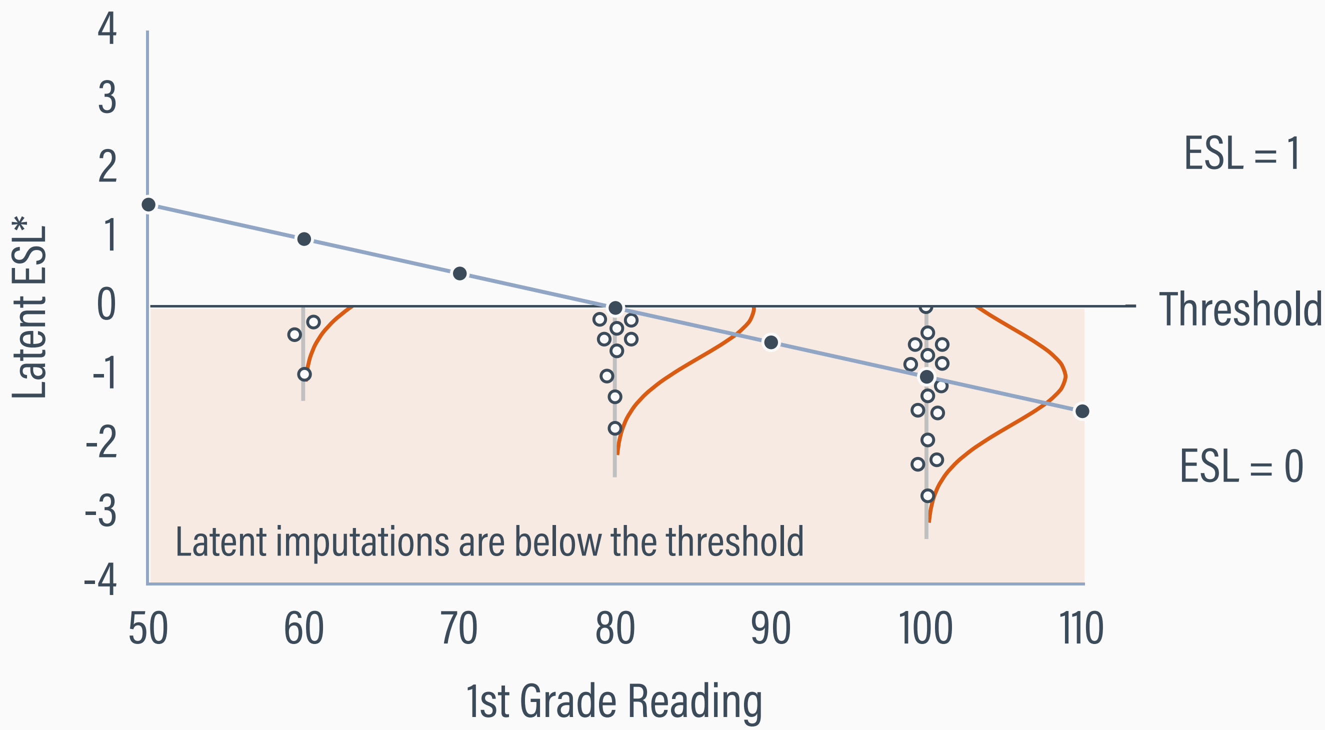
- Multiple threshold parameters divide the latent distribution into segments, with areas under the curve matching the bar plot



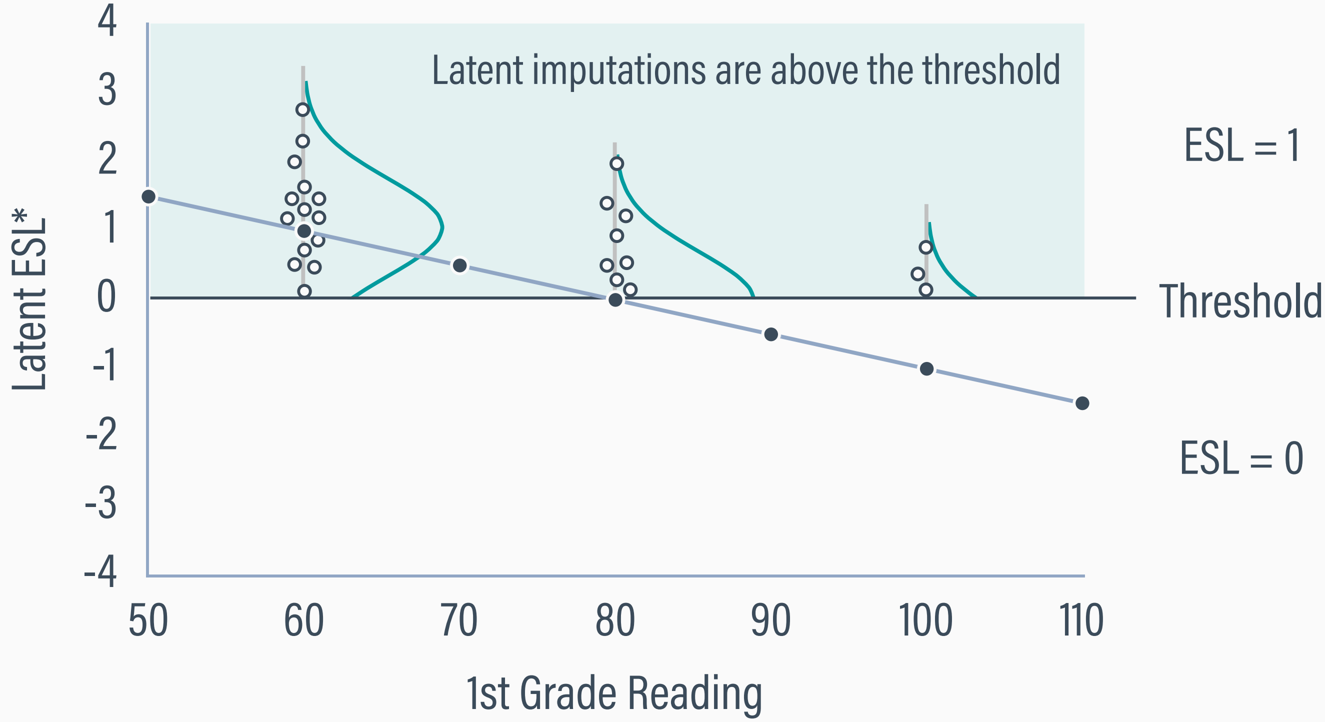
IMPUTING LATENT RESPONSE SCORES

- Latent response scores are missing data to be imputed
- MCMC uses computer simulation to “sample” latent response scores distributions, just like any other incomplete variable
- The imputation procedure restricts the range of latent response scores to align with the areas above and below the threshold

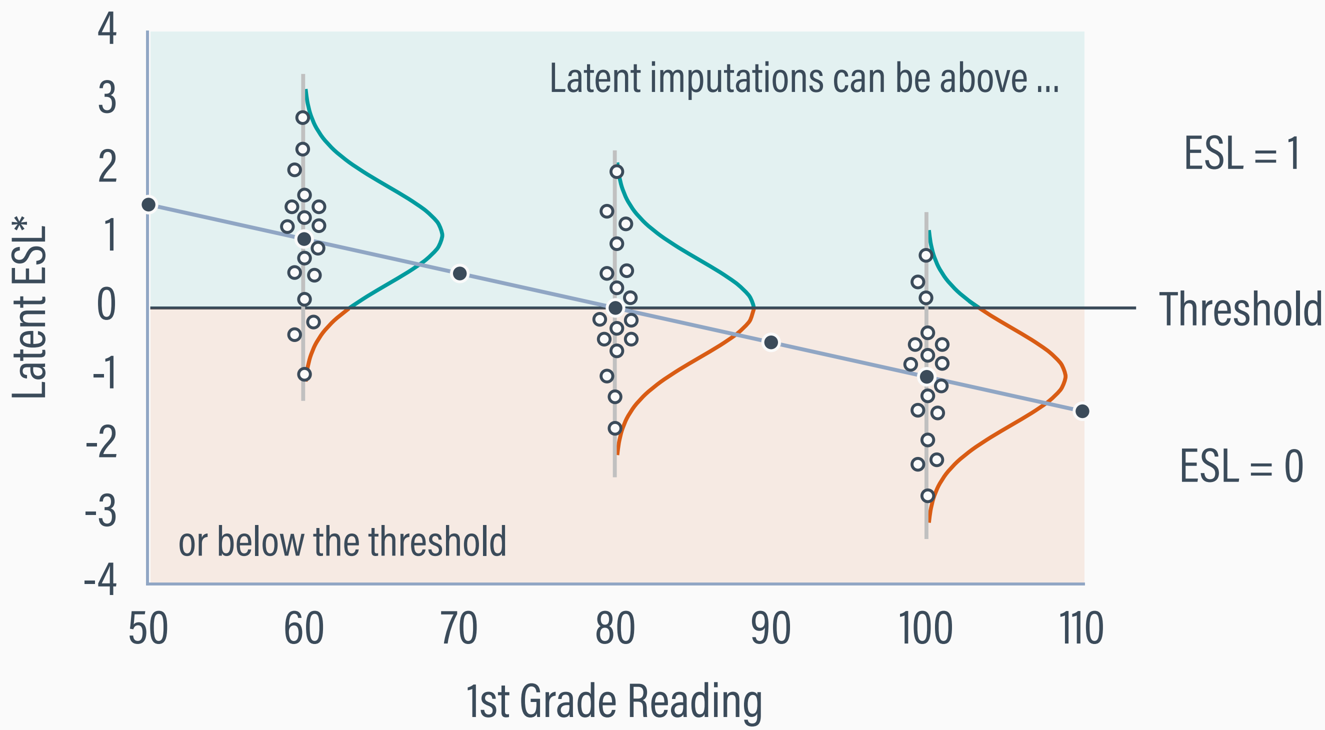
LATENT IMPUTATIONS: ESL = 0



LATENT IMPUTATIONS: ESL = 1



LATENT IMPUTATIONS: ESL = ?



ANALYSIS MODEL

- Multiple regression model with two continuous and two binary predictor variables

$$READ_9 = \beta_0 + \beta_1(READ_1) + \beta_2(LRNPROB_1) + \beta_3(MALE) + \beta_4(ESL) + \varepsilon$$

- ESL is an incomplete binary dummy code that indicates whether English is the student's second language

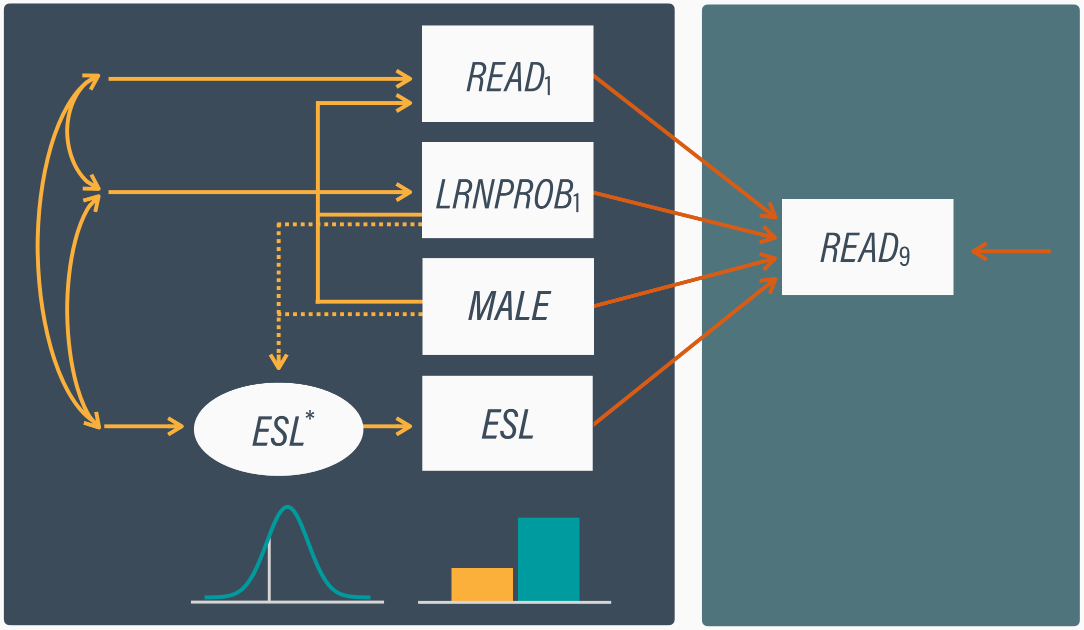
BLIMP SCRIPT

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
ORDINAL: esl;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male esl;
BURN: 1000;
ITERATIONS: 10000;
SEED: 90291;
```

PATH DIAGRAM AND MODEL COMMAND

Joint Distribution = Multivariate Predictor Distribution × Univariate Outcome Distribution

ORDINAL: esl;
NOMINAL: male;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1
male esl;



PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE:  Split chain PSR is being used. This splits each chain's
         iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR    Parameter #
                               26  to 50          1.374              7
                               51  to 100         1.084             15
                               76  to 150         1.060              7
                              101  to 200         1.033             16
                              ...      ...          ...            ...
                              401  to 800         1.018             14
                              426  to 850         1.013             14
                              451  to 900         1.015             14
                              476  to 950         1.018             14
```

» Worst PSR < 1.05 | 501 to 1000 | 1.021 | 16

COEFFICIENTS

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Parameters                      Median     StdDev       2.5%      97.5%        PSR      N_Eff
                              -----------------------------------------------------------------

Variances:
  Residual Var.                 77.639     11.467     59.343    103.993      1.001   4608.173

Coefficients:
  Intercept                     55.844      6.876     42.213     69.103      1.001   3879.836
  read1                          0.568      0.046      0.481      0.659      1.000   5633.633
  lrnprob1                      -0.274      0.094     -0.456     -0.091      1.001   2647.338
  male.1                         2.344      1.804     -1.248      5.844      1.001   5412.478
  esl                           -7.864      1.852    -11.437     -4.173      1.000   4578.354

Standardized Coefficients:
  ...
                              -----------------------------------------------------------------
```

ESL vs. non-ESL mean difference

EXAMPLE INTERPRETATION

- For two students who share the same profile on all other predictors (gender, 1st grade reading, and 1st grade learning problems), speaking English as a second language is associated with a 7.86 decrease in grade 9 reading scores
- The mean difference is significant because 0 falls outside the 95% credible interval limits (–11.43, –4.17)

REPORTING TEMPLATE

We used Bayesian estimation in Blimp 3 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 1,000 iterations was sufficiently conservative. Based on this information, we used two MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

REPORTING TEMPLATE CONTINUED

Table 1 displays the posterior summaries from the analysis. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent. However, Bayesian estimation is preferable because classic FIML estimators in widespread use do not readily accommodate mixtures of discrete and continuous predictors (Enders, 2022).

APA TABLE

Table 1

Parameter Summary From the Bayesian Regression Analysis

Parameter	Median	*SD*	LCL	UCL
Intercept	55.84	6.88	42.21	69.10
1st grade reading	0.57	0.05	0.48	0.66
1st grade learning problems	−0.27	0.09	−0.46	−0.09
Male dummy code	2.34	1.80	-1.25	5.84
ESL dummy code	−7.86	1.85	−11.44	−4.17
R^2	.64	.05	.54	.72

REPORTING TEMPLATE CONTINUED

Collectively, the predictors explained approximately 64% of the variation in 9th grade reading scores. First grade reading exhibited a significant positive association with 9th grade reading performance ($\beta = 0.57$, $SD = 0.05$, $p < .05$), and the measure of teacher-rated learning problems was inversely related to middle-school reading ($\beta = -0.27$, $SD = 0.09$, $p < .05$). Students for whom English is a second language scored approximately 7.86 points lower, an average, than native English speakers ($\beta = -7.86$, $SD = 1.85$, $p < .05$). Finally, student gender was not predictive of 9th reading performance.

MULTICATEGORICAL VARIABLES

- Multinomial probit regression envisions a latent normal distribution for each response option
- Like dummy codes, there is a reference group (the lowest category is Blimp's default)
- Latent difference scores compare the latent response scores from the C – 1 categories to that of the reference group

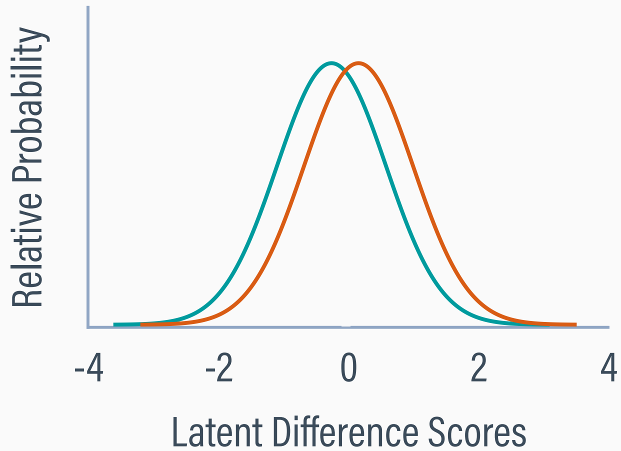
AERA WORKSHOP DATA

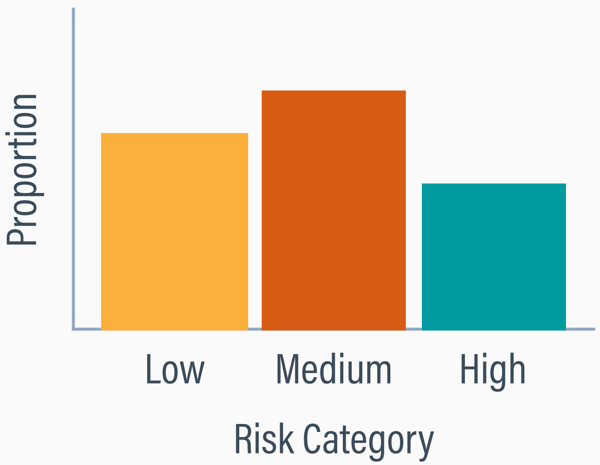
Predictors
Outcome

Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
ESL	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

LATENT RISK GROUP DISTRIBUTIONS

- Latent difference scores compare latent response scores for the Medium and High groups to the Low-risk reference category





ANALYSIS MODEL

- Multiple regression model with two continuous and two binary predictor variables

$$READ_9 = \beta_0 + \beta_1(READ_1) + \beta_2(LRNPROB_1) + \beta_3(MALE) + \beta_4(MEDRSK) + \beta_5(HIGHRSK) + \varepsilon$$

- HIGHRSK and MEDRSK are dummy (1 vs. 0) codes comparing the high and medium risk groups to the low-risk reference

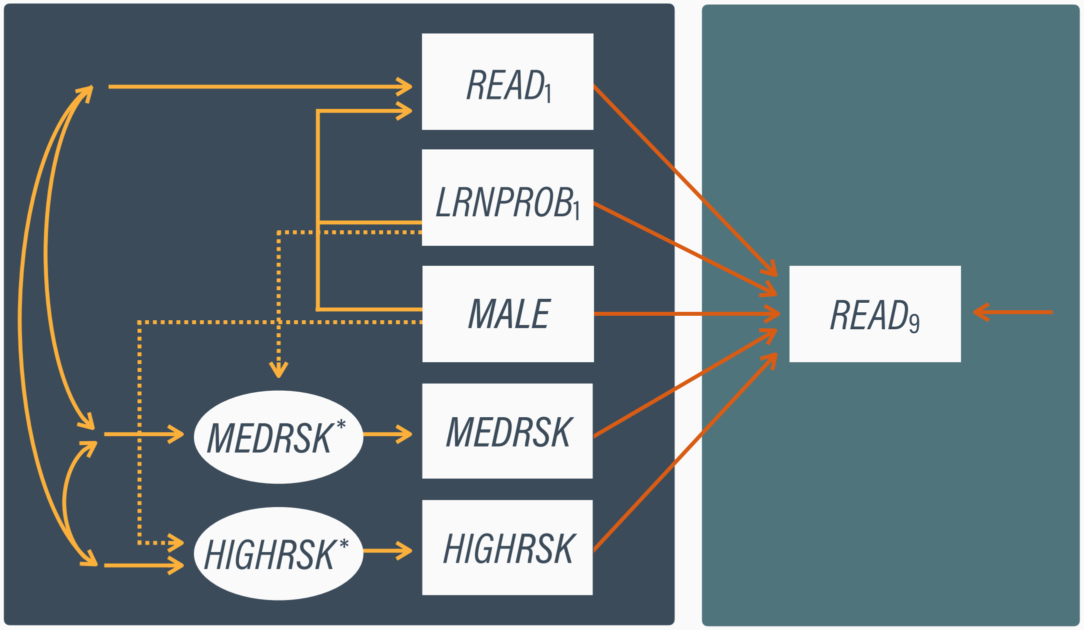
BLIMP SCRIPT

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male riskgrp;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 male riskgrp;
BURN: 2000;
ITERATIONS: 10000;
SEED: 90291;
```

PATH DIAGRAM AND MODEL COMMAND

Joint Distribution = Multivariate Predictor Distribution × Univariate Outcome Distribution

NOMINAL: male rskgrp;
FIXED: male lrnprob1;
MODEL:
read9 ~ read1 lrnprob1
    male rskgrp;



PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE:  Split chain PSR is being used. This splits each chain's
         iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                           51  to 100            1.166             15
                          101  to 200            1.107             16
                          151  to 300            1.058             21
                          201  to 400            1.048             21
                          ...      ...             ...            ...
                          801  to 1600           1.034             22
                          851  to 1700           1.027             22
                          901  to 1800           1.044             22
                          951  to 1900           1.043             22
                         1001  to 2000           1.028             22
```

» Worst PSR < 1.05

COEFFICIENTS

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:   read9

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                     -------------------------------------------------------------------

Variances:
  Residual Var.                       77.639     11.467     59.343    103.993      1.001   4608.173

Coefficients:
  Intercept                           67.276      7.551     52.379     81.912      1.000   4099.229
  read1                                0.483      0.052      0.382      0.586      1.000   6340.240
  lrnprob1                            -0.395      0.096     -0.582     -0.211      1.000   3622.421
  male.1                               2.296      1.918     -1.543      6.039      1.000   5188.817
  riskgrp.2                           -1.831      2.096     -5.913      2.313      1.000   5833.837
  riskgrp.3                           -4.780      2.875    -10.563      0.797      1.000   4587.357

Standardized Coefficients:
  ...
                                     -------------------------------------------------------------------
```

Medium- vs. low-risk mean difference

High- vs. low-risk mean difference

EXAMPLE INTERPRETATION

- For two students who share the same profile on all other predictors (gender, 1st grade reading, and 1st grade learning problems), belonging to the medium- (or high-risk) group is associated with a 1.83 (or 4.78) decrease in grade 9 reading scores relative to the low-risk reference group
- Both mean difference parameters are significant because 0 falls outside their 95% credible interval limits

REPORTING TEMPLATE

We used Bayesian estimation in Blimp 3 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 2,000 iterations was sufficiently conservative. Based on this information, we used two MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

REPORTING TEMPLATE CONTINUED

Table 1 displays the posterior summaries from the analysis. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent. However, Bayesian estimation is preferable because existing FIML estimators cannot accommodate mixtures of continuous and multicategorical nominal predictors (Enders, 2022).

APA TABLE

Table 1

Parameter Summary From the Bayesian Regression Analysis

Parameter	Median	*SD*	LCL	UCL
Intercept	67.28	7.55	52.38	81.91
1st grade reading	0.48	0.05	0.38	0.59
1st grade learning problems	−0.40	0.10	−0.58	−0.21
Male dummy code	2.30	1.92	−1.54	6.04
Medium- vs. low-risk dummy code	−1.83	2.10	−5.91	2.31
High- vs. low-risk dummy code	−4.78	2.88	−10.56	0.80
R^2	.59	.05	.48	.68

REPORTING TEMPLATE CONTINUED

Collectively, the predictors explained approximately 59% of the variation in 9th grade reading scores. First grade reading exhibited a significant positive association with 9th grade reading performance ($\beta = 0.48$, $SD = 0.05$, $p < .05$), and the measure of teacher-rated learning problems was inversely related to middle-school reading ($\beta = -0.40$, $SD = 0.10$, $p < .05$). Although students classified as having moderate or high risk of emotional and behavioral disorders scored lower, on average, than their low-risk peers, neither mean difference was statistically significant. Similarly, student gender was not predictive of 9th reading performance.

CATEGORICAL OUTCOME VARIABLES

- Outcome variables can be binary, ordinal, multicategorical, or count (in addition to normal, latent, and skewed continuous)
- Ordinal outcomes using probit regression, count outcomes using negative binomial regression with over-dispersion
- Binary and multicategorical outcomes can be modeled using either probit or logistic regression

BINARY LOGISTIC MODEL

- Logistic regression model with a binary outcome denoting whether a student was reading at an average level in 9th grade

$$\begin{aligned}\text{logit}(READGRP_9) &= \beta_0 + \beta_1(READ_1) + \beta_2(LRNPROB_1) \\ &+ \ \beta_3(MALE) + \beta_4(ESL) + \varepsilon\end{aligned}$$

- ESL is an incomplete binary dummy code that indicates whether English is the student's second language

AERA WORKSHOP DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
ESL	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

BLIMP SCRIPT

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
NOMINAL: male;
ORDINAL: esl read9grp;
FIXED: male lrnprob1;
MODEL:
logit(read9grp) ~ read1 lrnprob1 male esl;
BURN: 2000;
ITERATIONS: 10000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                        51  to 100              1.101              6
                       101  to 200              1.055              3
                       151  to 300              1.052             14
                       201  to 400              1.045             14
                       ...      ...              ...             ...
                       801  to 1600             1.006             22
                       851  to 1700             1.011             14
                       901  to 1800             1.010             14
                       951  to 1900             1.009             14
```

» Worst PSR < 1.05

```
                      1001  to 2000             1.012             14
```

COEFFICIENTS

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  logit(read9grp)

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                               ------------------------------------------------------------------
Coefficients:
  Intercept                         -3.549      1.397     -6.354     -0.864      1.000   4358.138
  read1                              0.073      0.015      0.047      0.104      1.000   2921.149
  lrnprob1                          -0.052      0.023     -0.099     -0.009      1.001   3005.165
  male.1                             0.442      0.462     -0.462      1.366      1.000   4807.380
  esl                               -0.934      0.505     -1.965      0.023      1.000   3602.590

Odds Ratios:
  Intercept                          0.029      0.141      0.002      0.422      1.000   6928.599
  read1                              1.076      0.016      1.048      1.109      1.000   2919.172
  lrnprob1                           0.950      0.022      0.906      0.991      1.001   3021.609
  male.1                             1.556      0.875      0.630      3.921      1.000   5005.948
  esl                                0.393      0.230      0.140      1.024      1.001   3925.777

  ...
```

Odds ratios

Logistic coefficients

INTERACTION EFFECTS IN BLIMP

MODERATED REGRESSION

- Moderation occurs when a focal predictor's influence on an outcome depends on a third variable called a moderator
- Moderated regression answers the question, for whom does an effect apply?

MODERATED REGRESSION MODEL

- β_1 and β_2 are conditional effects: β_1 is the influence of X when M equals 0, and β_2 is the influence of M when X is 0

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 XM + \varepsilon$$

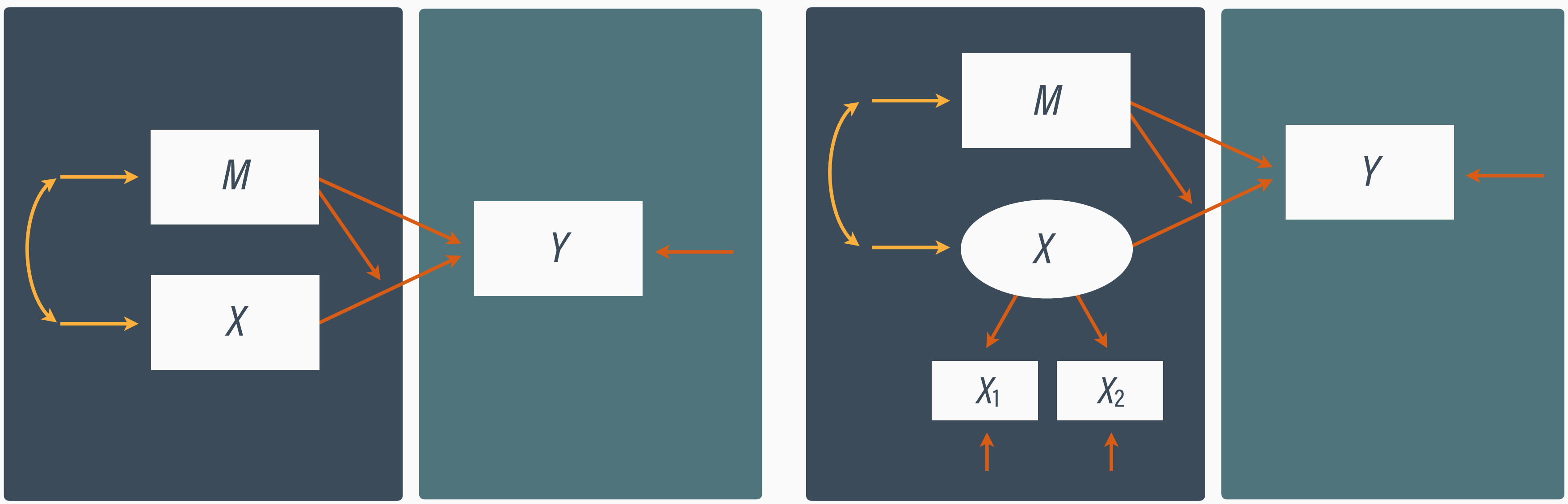
- β_3 captures the change in the β_1 slope for every one-unit increase in M

INCOMPLETE PRODUCT TERMS

- Factored regression specifications readily accommodate incomplete interactive and other nonlinear effects
- Product terms appear as deterministic functions of lower-order terms in the focal regression model rather than free variables
- The models and distributions of the predictors are unchanged

FACTORED SPECIFICATIONS FOR INTERACTIONS

Joint Distribution = Multivariate Predictor Distribution × Univariate Outcome Distribution



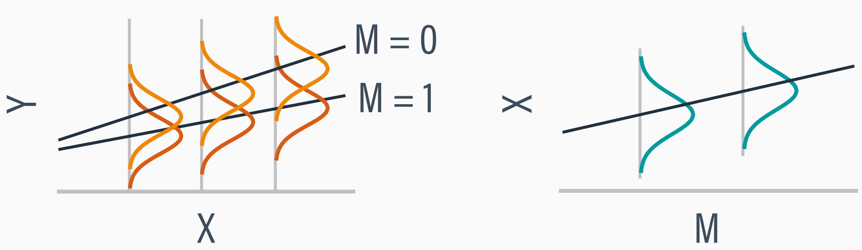
MODERATED REGRESSION EXAMPLE

Factorization $$f(Y, X, M) = f(Y\,|X,M) \times f(X,M)$$

$$X = \gamma_0 + \gamma_1 M + \epsilon$$

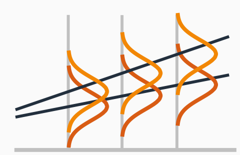
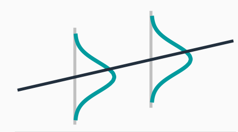
Fitted models $$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 XM + \varepsilon$$

Distributions



IMPUTING MISSING PREDICTORS

- A missing predictor always appears in two models: as a regressor in the focal model and an outcome in its own model

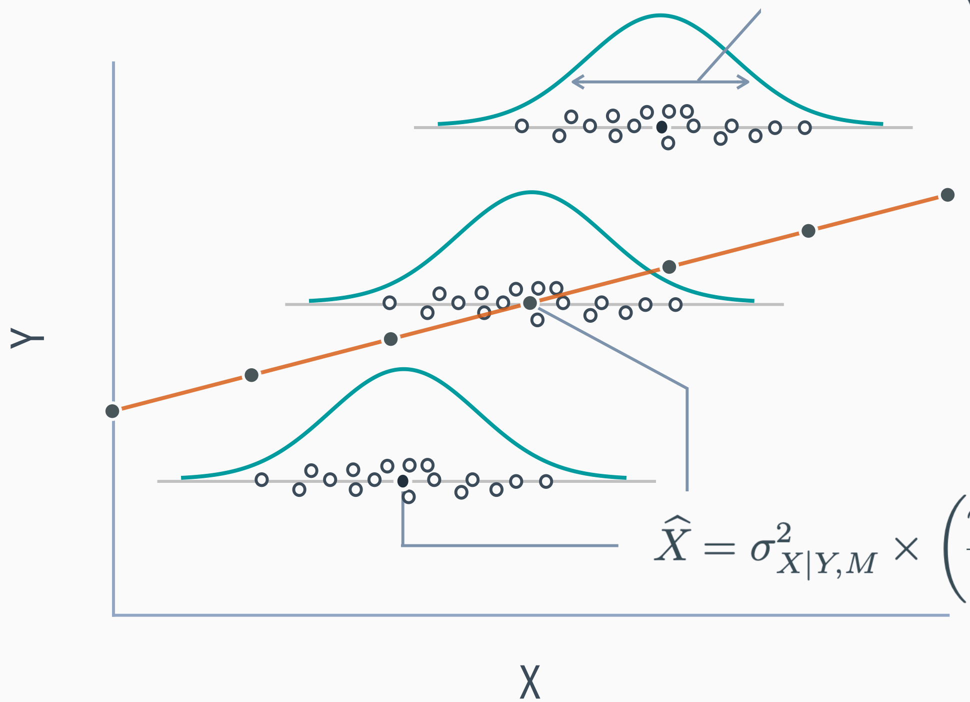
$$f(X|Y,M) \propto f(Y\,|X,M) \times f(X|M) =$$  $\times$ 

- The distribution of missing X scores is a composite function that depends on both model-implied distributions

DISTRIBUTIONS OF IMPUTATIONS

$$\sigma^2_{X|Y,M} = \left(\frac{1}{\sigma^2_\epsilon} + \frac{(\beta_1 + \beta_3 M)^2}{\sigma^2_\varepsilon}\right)^{-1}$$

Heteroscedastic variation!



$$\widehat{X} = \sigma^2_{X|Y,M} \times \left(\frac{\gamma_0 + \gamma_1 M}{\sigma^2_\epsilon} + \frac{(\beta_1 + \beta_3 M)(Y - \beta_0 - \beta_2 M)}{\sigma^2_\varepsilon}\right)$$

AERA WORKSHOP DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
STUDENT	Student identifier	0	Integer index
MALE	Gender code	0	0 = Female, 1 = Male
ESL	English as a second language code	5.1	0 = Non-ESL, 1 = ESL
RISKGRP	Emotional/behavioral disorder risk	2.2	1 = Low, 2 = Medium, 3 = High
ATRISK	Emotional/behavioral risk code	2.2	0 = Low, 1 = Medium/high
BEHSYMP1	1st grade behavioral symptoms	3.6	Numeric (17 to 92)
LRNPROB1	1st grade learning problems	0	Numeric (31 to 88)
READ1	1st grade broad reading composite	6.5	Numeric (39 to 153)
READ9	9th grade broad reading composite	17.4	Numeric (41 to 123)
READGRP9	9th grade reading classification	17.4	0 = Below average, 1 = Average/above
STANREAD7	7th grade standardized reading	19.6	Numeric (100 to 399)

ANALYSIS MODEL

- The influence of 1st grade reading performance on 9th grade test scores is moderated by learning problems in 1st grade

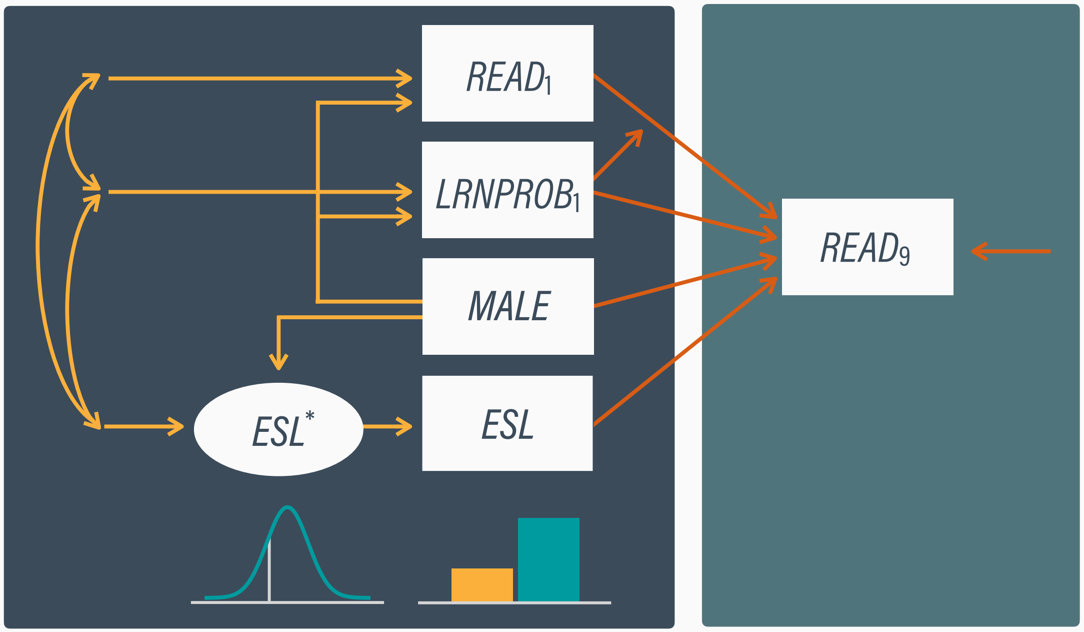
$$READ_9 = \beta_0 + \beta_1(READ_1) + \beta_2(LRNPROB_1) + \beta_3(READ_1)(LRNPROB_1) + \beta_4(MALE) + \beta_5(ESL) + \varepsilon$$

- 1st grade reading and learning problems scores are centered at the grand mean to facilitate interpretation of lower-order terms

PATH DIAGRAM AND MODEL COMMAND

Joint Distribution = Multivariate Predictor Distribution × Univariate Outcome Distribution

FIXED: male;
CENTER: read1 lrnprob1;
MODEL:
read9 ~ read1 lrnprob1
    read1*lrnprob1 male esl;



BLIMP SCRIPT

```
DATA: aeraworkshop.dat;
VARIABLES: id male esl riskgrp atrisk behsymp1 lrnprob1 read1 read9 read9grp stanread7;
MISSING: 999;
ORDINAL: esl;
NOMINAL: male;
FIXED: male;
CENTER: read1 lrnprob1;
MODEL:
read9 ~ read1 lrnprob1 read1*lrnprob1@interaction male esl;
TEST: interaction = 0;
BURN: 2000;
ITERATIONS: 10000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE:  Split chain PSR is being used. This splits each chain's
         iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                           51  to 100           1.461             26
                          101  to 200           1.286             21
                          151  to 300           1.197             26
                          201  to 400           1.142             26
                          ...      ...            ...            ...
                          801  to 1600          1.019             26
                          851  to 1700          1.015             21
                          901  to 1800          1.017             21
                          951  to 1900          1.012             21
                         1001  to 2000          1.018             26
```

» Worst PSR < 1.05 (1001 to 2000: 1.018, parameter 26)

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  read9

Grand Mean Centered: lrnprob1 read1

Parameters                             Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                 -------------------------------------------------------------------
Variances:
  Residual Var.                        68.188      9.878     52.375     90.753      1.000   5133.493

Coefficients:
  Intercept                            90.118      1.755     86.723     93.590      1.002   1163.979
  read1                                 0.602      0.047      0.510      0.696      1.002   3470.201
  lrnprob1                             -0.255      0.094     -0.442     -0.072      1.001   2120.825
  male.1                                3.712      1.700      0.296      6.984      1.000   5492.460
  esl                                  -7.340      1.757    -10.752     -3.918      1.001   4364.216
  read1*lrnprob1                        0.016      0.005      0.008      0.025      1.001   2414.527

Standardized Coefficients:
  ...
```

Change in READ1 slope for a 1-unit increase in LRNPROB1
Conditional effect of LRNPROB1 at READ1 = 0 (the mean)
Conditional effect of READ1 at LRNPROB1 = 0 (the mean)

MODEL FIT OUTPUT

```
MODEL FIT:

  INFORMATION CRITERIA
     …

  WALD TESTS (Asparouhov & Muthén, 2021)

  Test #1

    Full:
      [1]  read9 ~ Intercept read1 lrnprob1 male.1 esl read1*lrnprob1@interaction

    Restricted:
      [1]  read9 ~ Intercept read1 lrnprob1 male.1 esl read1*lrnprob1@interaction

    Constraints in Restricted:
      [1]  interaction = 0


    Wald Statistic (Chi-Square)                  13.206
    Number of Parameters Tested (df)                  1
    Probability                                   0.000
```

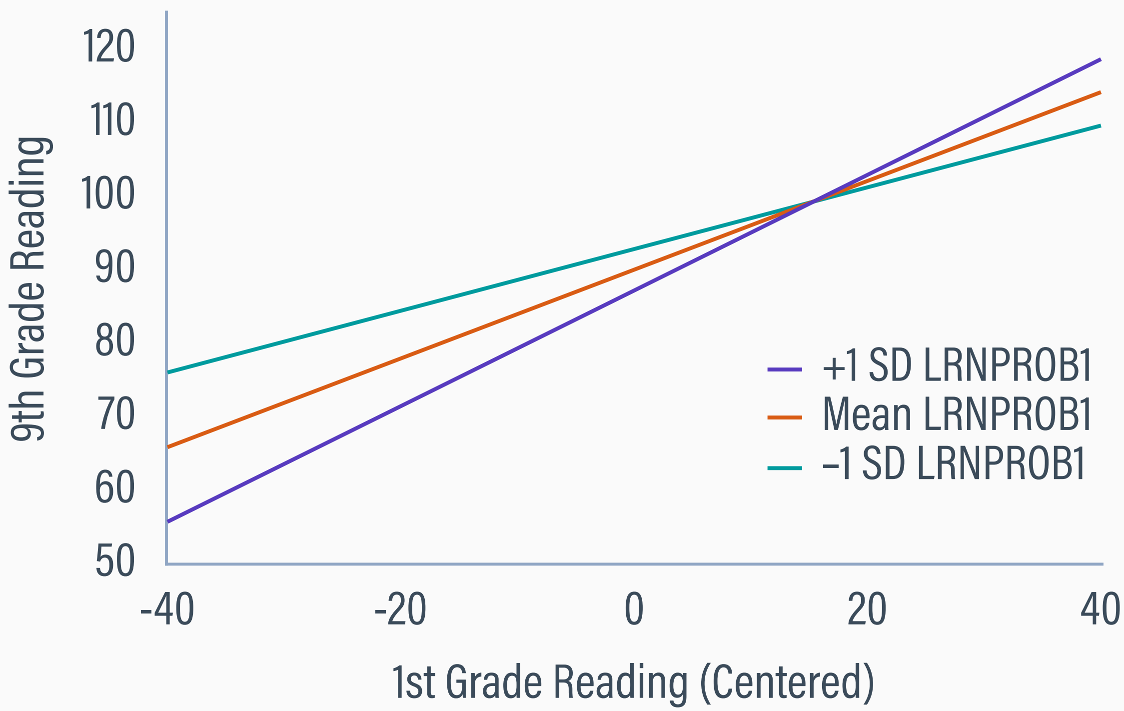
BLIMP SCRIPT: CONDITIONAL EFFECTS

```
CENTER: read1 lrnprob1;
MODEL: read9 ~ read1 lrnprob1 read1*lrnprob1@interaction male esl;
SIMPLE: read1 | lrnprob1;          # focal predictor | moderator
```

CONDITIONAL EFFECT OUTPUT

```
Conditional Effects                Median    StdDev      2.5%     97.5%       PSR     N_Eff
                                -------------------------------------------------------------------
...
  read1 | lrnprob1 @ +1 SD
    Intercept                      87.267     2.377    82.576    91.919     1.000   985.997
    Slope                           0.787     0.076     0.642     0.942     1.001  2811.774

  read1 | lrnprob1 @ 0
    Intercept                      90.100     1.816    86.592    93.650     1.000  1082.863
    Slope                           0.604     0.047     0.512     0.695     1.001  3485.920

  read1 | lrnprob1 @ -1 SD
    Intercept                      92.909     1.807    89.439    96.471     1.001  2013.538
    Slope                           0.421     0.062     0.294     0.538     1.000  4577.480
  ...
                                 -----------------------------------------------------------------
```

CONDITIONAL EFFECTS (SIMPLE SLOPES)



REPORTING TEMPLATE

We used Bayesian estimation in Blimp 3 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 2,000 iterations was sufficiently conservative. Based on this information, we used two MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

REPORTING TEMPLATE CONTINUED

Table 1 displays the posterior summaries from the analysis. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent. However, Bayesian estimation is preferable because classic FIML estimator is known to introduce bias when applied to interactive effects (Enders, 2022).

APA TABLE

Table 1

Parameter Summary From the Bayesian Moderated Regression Analysis

Parameter	Median	*SD*	LCL	UCL
Intercept	90.12	1.76	86.72	93.59
1st grade reading	0.60	0.05	0.51	0.70
1st grade learning problems	−0.26	0.09	−0.44	−0.07
Male dummy code	3.71	1.70	0.30	6.98
ESL dummy code	−7.34	1.76	−10.75	−3.92
Reading by learning problems	0.02	0.01	0.01	0.03
R^2	.69	.04	.60	.76

REPORTING TEMPLATE CONTINUED

Collectively, the predictors explained approximately 69% of the variation in 9th grade reading scores. The Bayesian Wald test (Asparouhov & Muthén, 2021) of the interaction effect was statistically significant, $\chi^2(1) = 13.21$, $p < .001$. At the learning problems mean, first grade reading exhibited a significant positive association with 9th grade reading performance ($\beta = 0.60$, $SD = 0.05$, $p < .001$), controlling for other predictors. The positive interaction coefficient indicates that the first grade reading slope increases as learning problems increase ($\beta = 0.02$, $SD = 0.01$, $p < .001$), such that first grade test scores become increasingly predictive of later achievement. Figure 1 displays the simple slopes at three levels of learning problems.

For more information go to

WWW.APPLIEDMISSINGDATA.COM