Today's Session

1. Do my best to supplement technical details of Dr Jamie Costley's presentation of our article on collaborative note-taking behaviors on Google docs and effects on student note-taking completeness and performance.

2. Present a R Shiny app, autopsych, that I built with my colleagues.

3. Share some current quasi-experimental work Jamie and I are doing on collaborative note-taking (maybe next time if time short…)

4. Meet more cool people from Russia.

The interaction of collaboration, note-taking completeness, and performance over 10 weeks of an online course

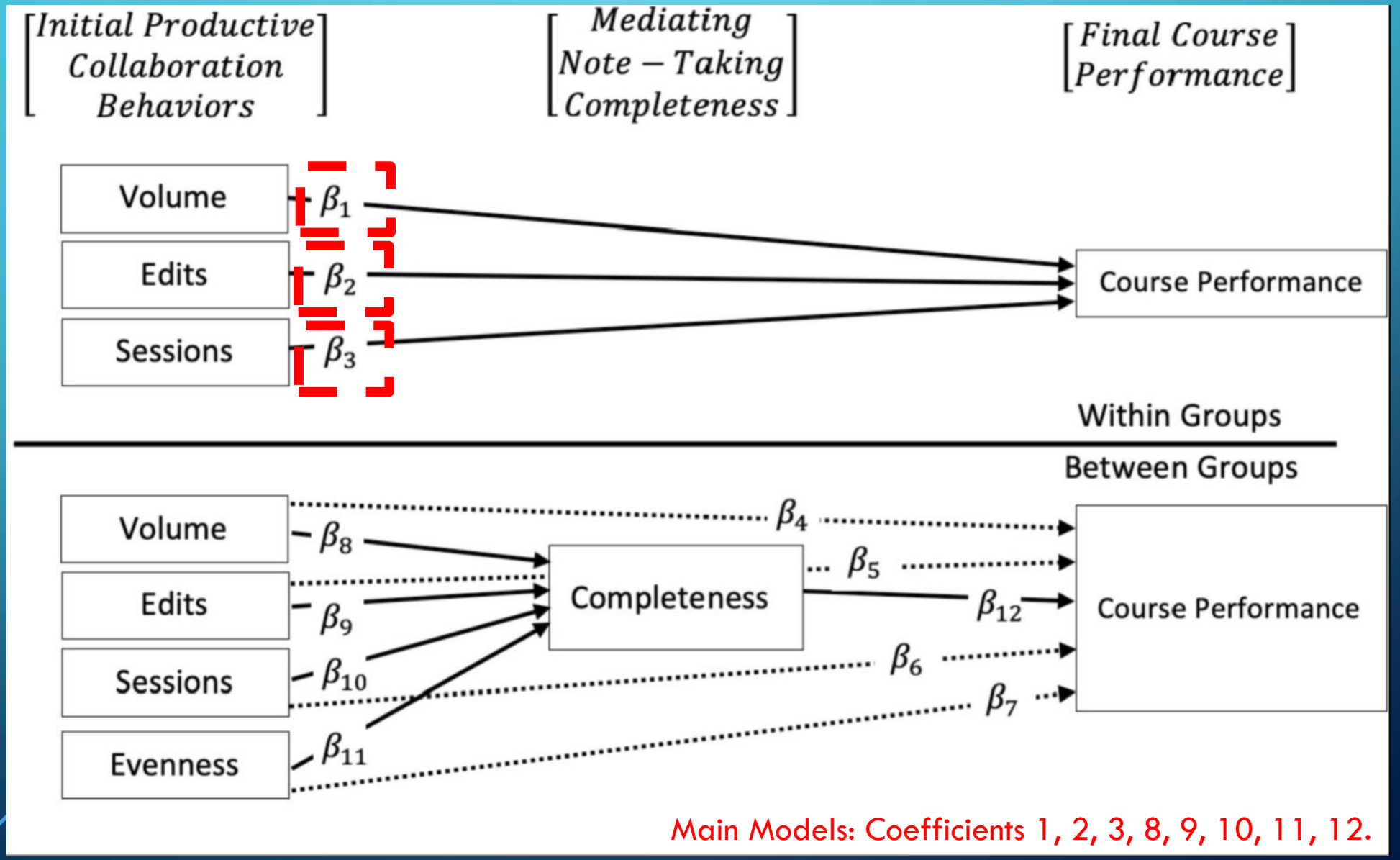
RQ1: How do within-group level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, and (c) number of log-ins affect students' weekly course performance?

RQ2: How do group-level collaborative behaviors such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness affect students' weekly group course performance?

RQ3: How do group-level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness of volume affect the completion of weekly group notes?

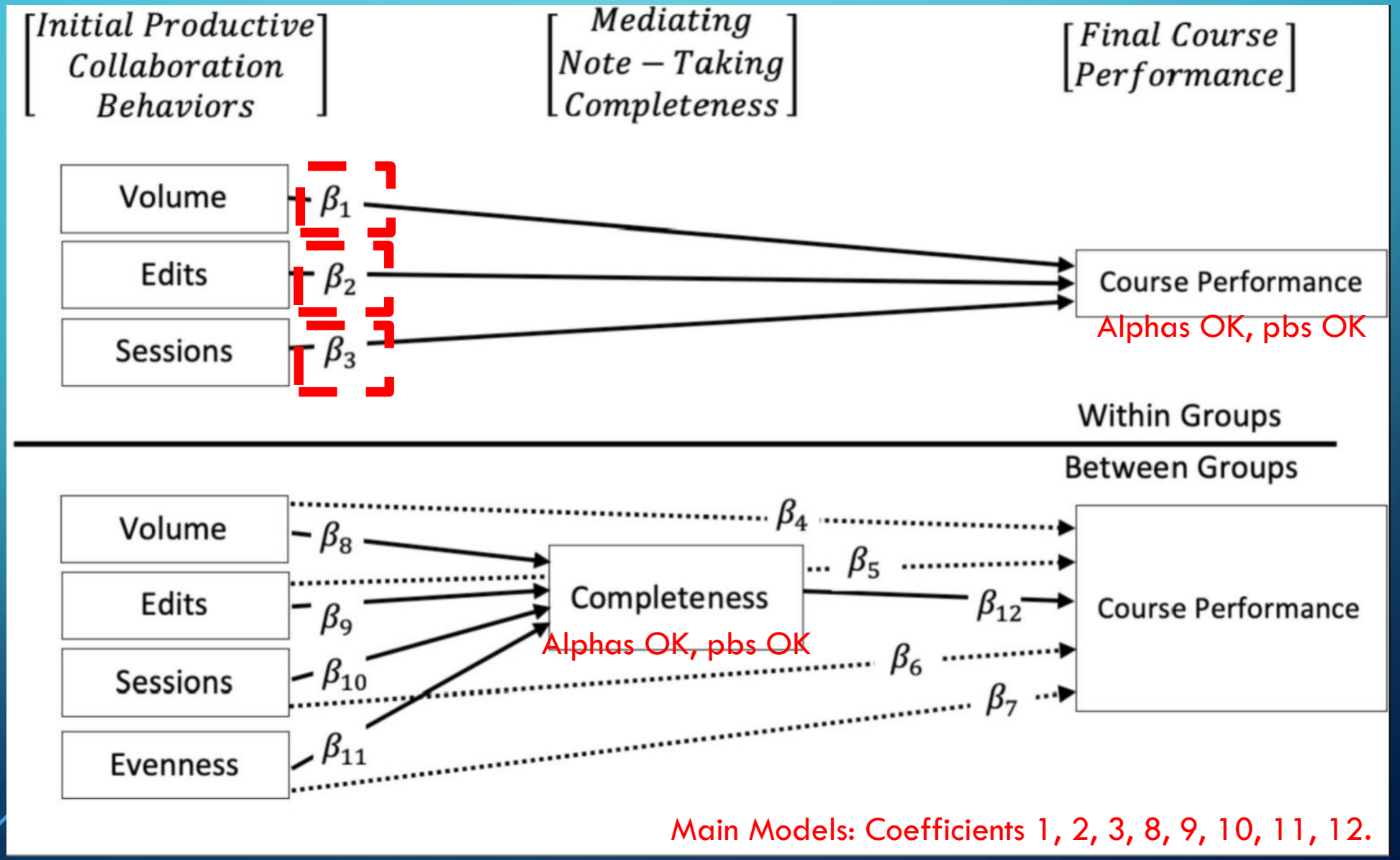
RQ4: How does the completion of group notes contribute to weekly student performance?

RQ1: How do within-group level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, and (c) number of log-ins affect students' weekly course performance?



Main Models: Coefficients 1, 2, 3, 8, 9, 10, 11, 12.

RQ1: How do within-group level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, and (c) number of log-ins affect students' weekly course performance?



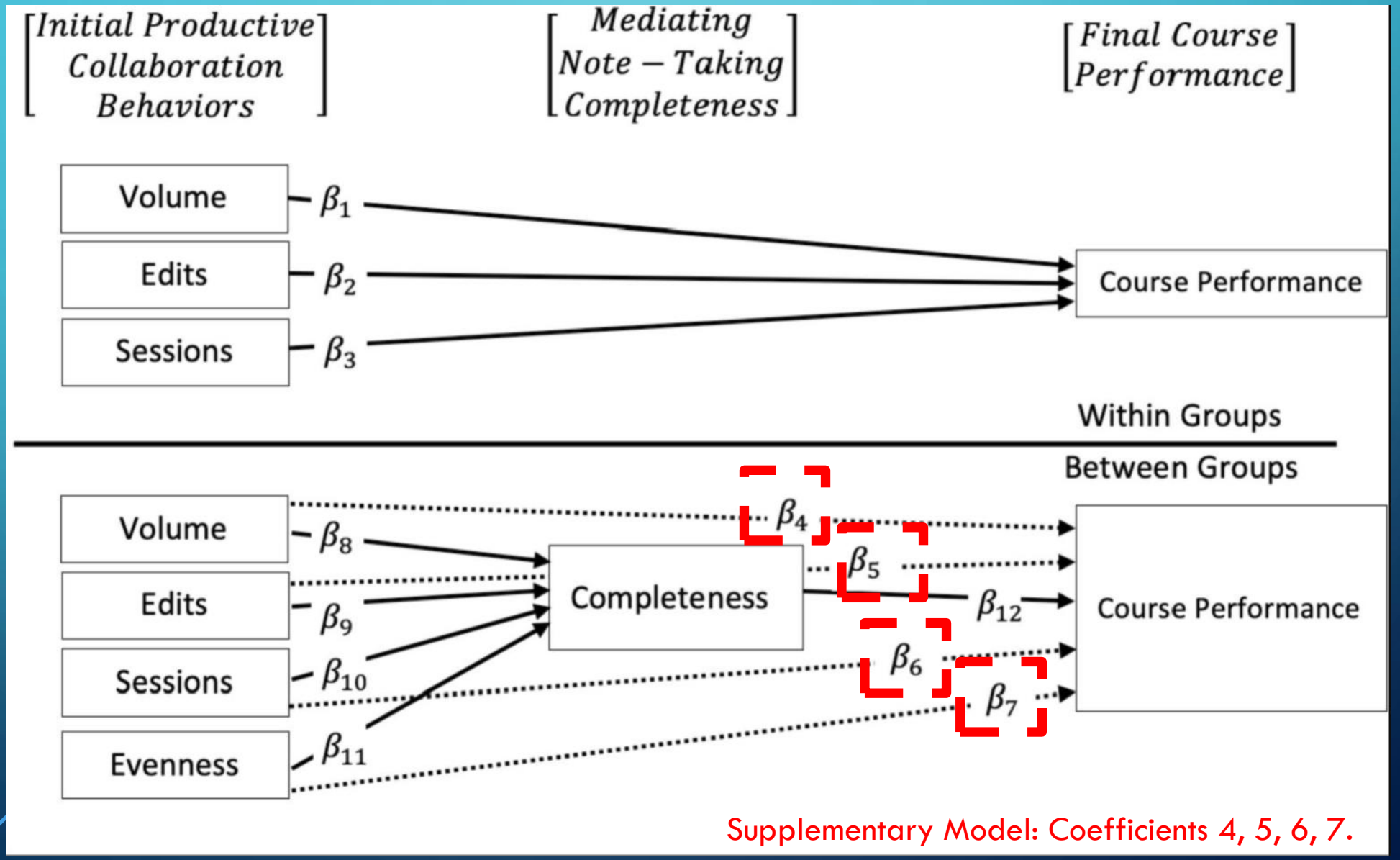
RQ1: How do within-group level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, and (c) number of log-ins affect students' weekly course performance?[coefficients 1, 2, 3]

Table 3

Summary of Effects from Main Multilevel Temporal Models for Weekly Group Completeness and Course Performances

Independent Variables	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Initial Online Note-Taking Behavior Effect on Course Performances [Within-Group Effects]										
Volume of Words (1)	.085	.147*	.223*	.081	.122	.110	**.187*** **	.149*	.137*	**.229*** **
Edits (of others) (2)	-.047	-.083	-.037	.036	-.063	-.028	-.068	.041	-.006	-.083
Session Logins (3)	.102	.000	-.104	-.042	.099	.065	.081	-.006	.000	-.037
$R^2(f^2)$	.021(.021)	.022(.22)	.041(.043)	.009(.009)	.029(.030)	.017(.017)	.039(.041)	.028(.029)	.018(.018)	.044(.046

RQ2: How do group-level collaborative behaviors such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness affect students' weekly group course performance?



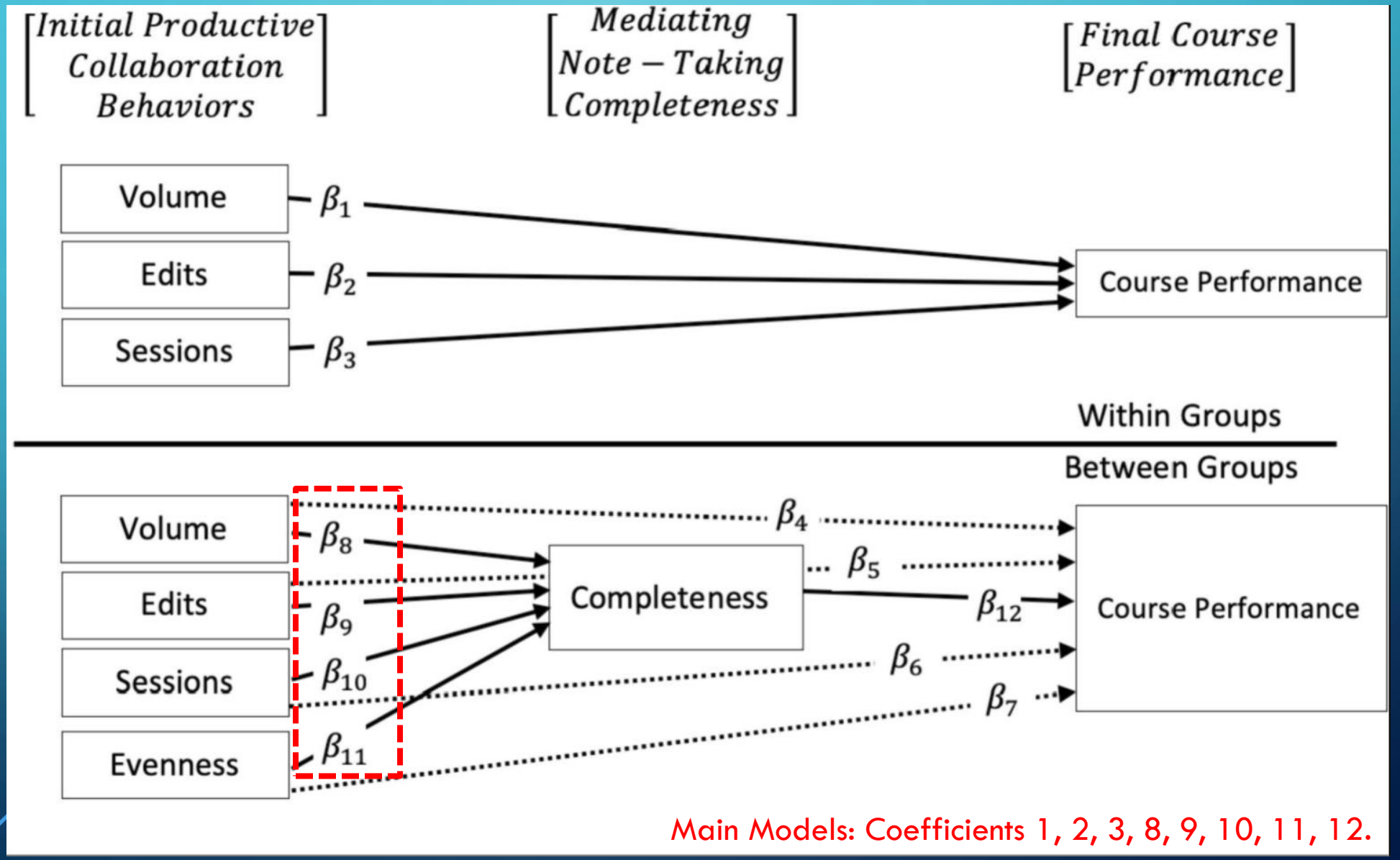
Supplementary Model: Coefficients 4, 5, 6, 7.

RQ2: How do group-level collaborative behaviors such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness affect students' weekly group course performance? [part 2 for each of the ten models]

Summary of Effects from Supplementary Model for Weekly Group Completeness and Course Performance

Independent Variables	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Initial Online Note-Taking Behavior Effect on Course Performance [Between-Group Effects]										
Intercept	0.021	-0.062	0.043	-0.040	0.018	-0.028	-0.063	0.104	0.000	-.073
Volume of Words (4)	-.382	.282	.067	.329	-.203	.267	.269	-.228	.376	-.001
Edits (of others) (5)	-.404	.104	-.097	-.058	**-.383***	-.354	-.304	-.293	-.276	-.189
Session Logins (6)	**.584***	.161	-.296	-.024	.137	.056	-.182	.270	-.043	-.008
Evenness of Group Vol. (7)	-.529	-.088	.048	.137	-.608	-.008	-.318	-.106	-.090	-.293
$R^2(f^2)$	.494(.976)	.269(.368)	.106(.119)	.064(.068)	.417(.715)	.162(.193)	.258(.348)	.203(.255)	.198(.282)	.112(.126)
Group and Student										

RQ3: How do group-level productive collaboration behaviors, such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness of volume affect the completion of weekly group notes?



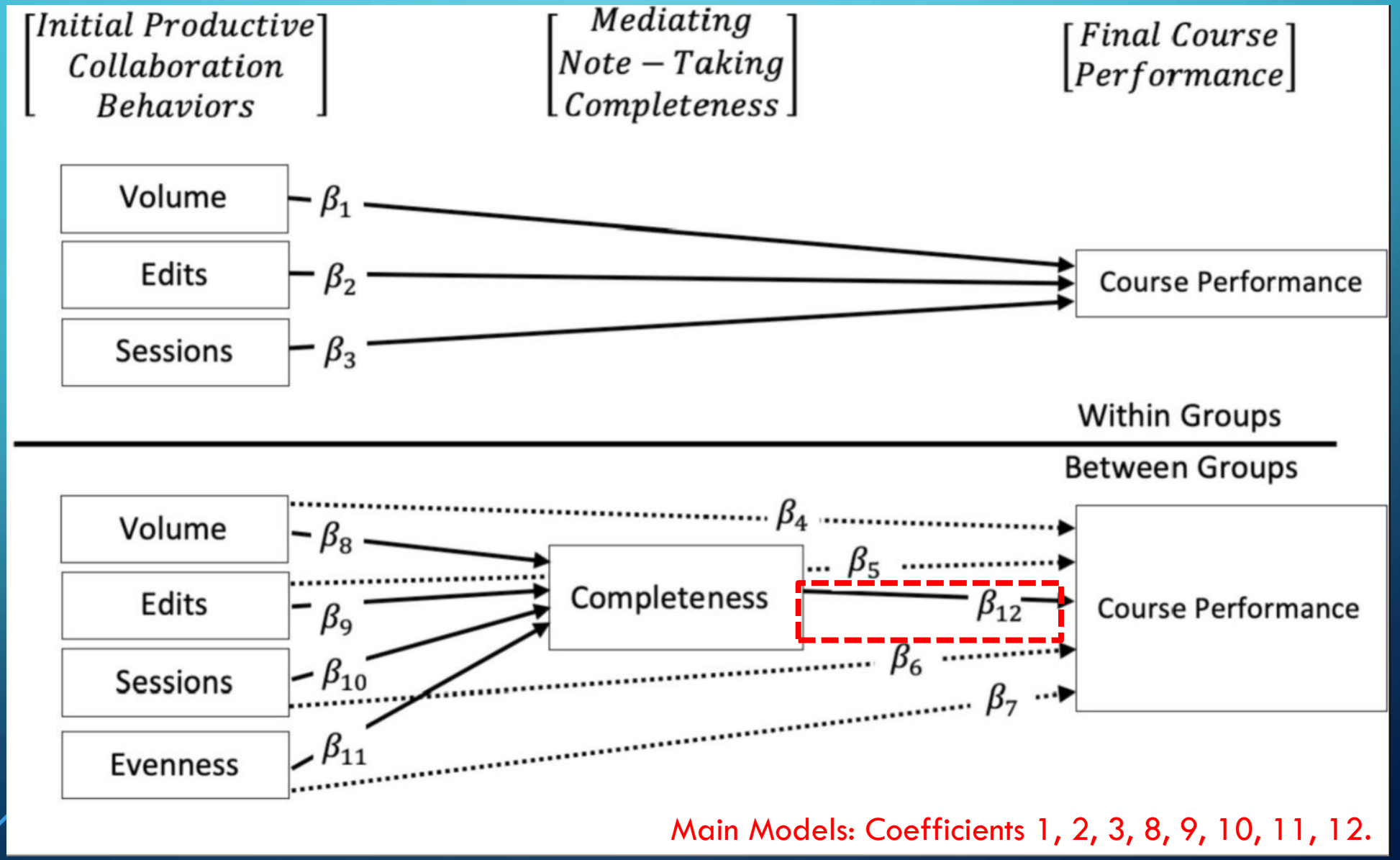
Main Models: Coefficients 1, 2, 3, 8, 9, 10, 11, 12.

RQ3: How do group-level collaborative behaviors such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness affect students' weekly group course performance?

Initial Online Note-Taking Behavior Effect on Note Completeness [Between-Group Effects]

Intercept	.004	-.049	-.015	-.071	-.030	.011	-.035	.107	-.035	-.047
Volume of Words ($_{8}$)	**.331****	**.488****	**.637*****	.215	**.386****	.335*	**.570*****	**.525*****	**.694*****	**.841*****
Edits (of others) ($_{9}$)	-.049	.119	-.145	.122	-.126	.067	-.129	.073	-.028	.059
Session Logins ($_{10}$)	.118	-.037	.061	-.148*	-.022	.137	.116	.066	.001	-.115
Volume Evenness ($_{11}$)	**-.392****	-.078	.093	-.306	-.284	-.046	.132	-.079	-.022	.061
$R^2(f^2)$	.441(.789)	.337(.508)	.309(.447)	.261(.353)	.315(.460)	.193(.239)	.249(.332)	.391(.642)	.488(.953)	.624(1.66

RQ4: How does the completion of group notes contribute to weekly student performance?



Main Models: Coefficients 1, 2, 3, 8, 9, 10, 11, 12.

RQ3: How do group-level collaborative behaviors such as (a) volume of words, (b) edits of others, (c) number of log-ins, and (d) evenness affect students' weekly group course performance?

Note Taking Completeness Effect on Course Performance [Between-Group Effects]

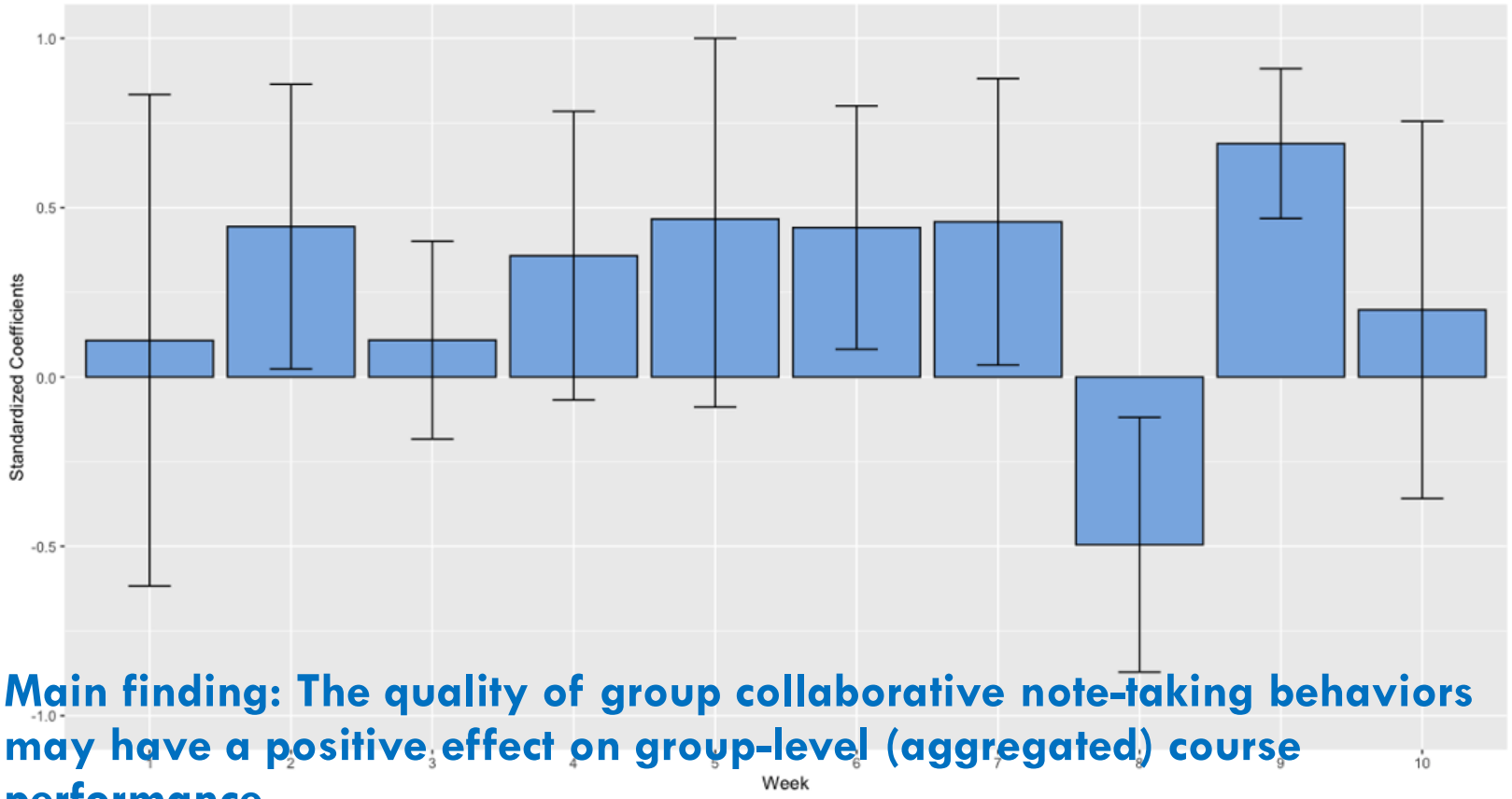
Intercept	.008	-.025	.028	-.030	.041	-.052	-.088	-.117	.014	-.056
Completeness(12)	.108	.444	.109	.358	.466	.441*	.458*	-.495*	**.689*** **	.198
$R^2(f^2)$	.012(.012)	.197(.245)	.012(.012)	.128(.147)	.217(.277)	.194(.241)	.210(.266)	.245(.325)	.475(.905)	.039(.041

Note. R^2= *total variance explained in outcome variables;* $f^2 = R^2/(1-R^2)$; $^*p < .05$, $^{**}p < .01$ in **bold**; $^{***}p < .001$ **bold and underlined**; all values, unless stated otherwise, represent standardized beta coefficients (see Figure 1); group and student sample sizes per week given in Table 2.

Main finding: The quality of group collaborative note-taking behaviors may have a positive effect on group-level (aggregated up) course performance.

The interaction of collaboration, note-taking completeness, and performance over 10 weeks of an online course: MAIN FINDING

Figure 1. Effect of Between-Group Completeness on Test Scores across 10 Weeks



Main finding: The quality of group collaborative note-taking behaviors may have a positive effect on group-level (aggregated) course performance.

Note. Standardized 95% confidence intervals generated with the assistance of the lavaan standardized solution function (Rosseel, 2012).

autopsych: An R Shiny Tool for the Reproducible Rasch Analysis, Differential Item Functioning, Equating, and Examination of Group Effects

Courtney, M. G. R., Chang, K., Mei, E., Meissel, K., Rowe, L., & Issayeva, L. (2021). autopsych: An R Shiny Tool for the Reproducible Rasch Analysis, Differential Item Functioning, Equating, and Examination of Group Effects. ***PLOS ONE***. doi:10.1371/journal.pone.0257682

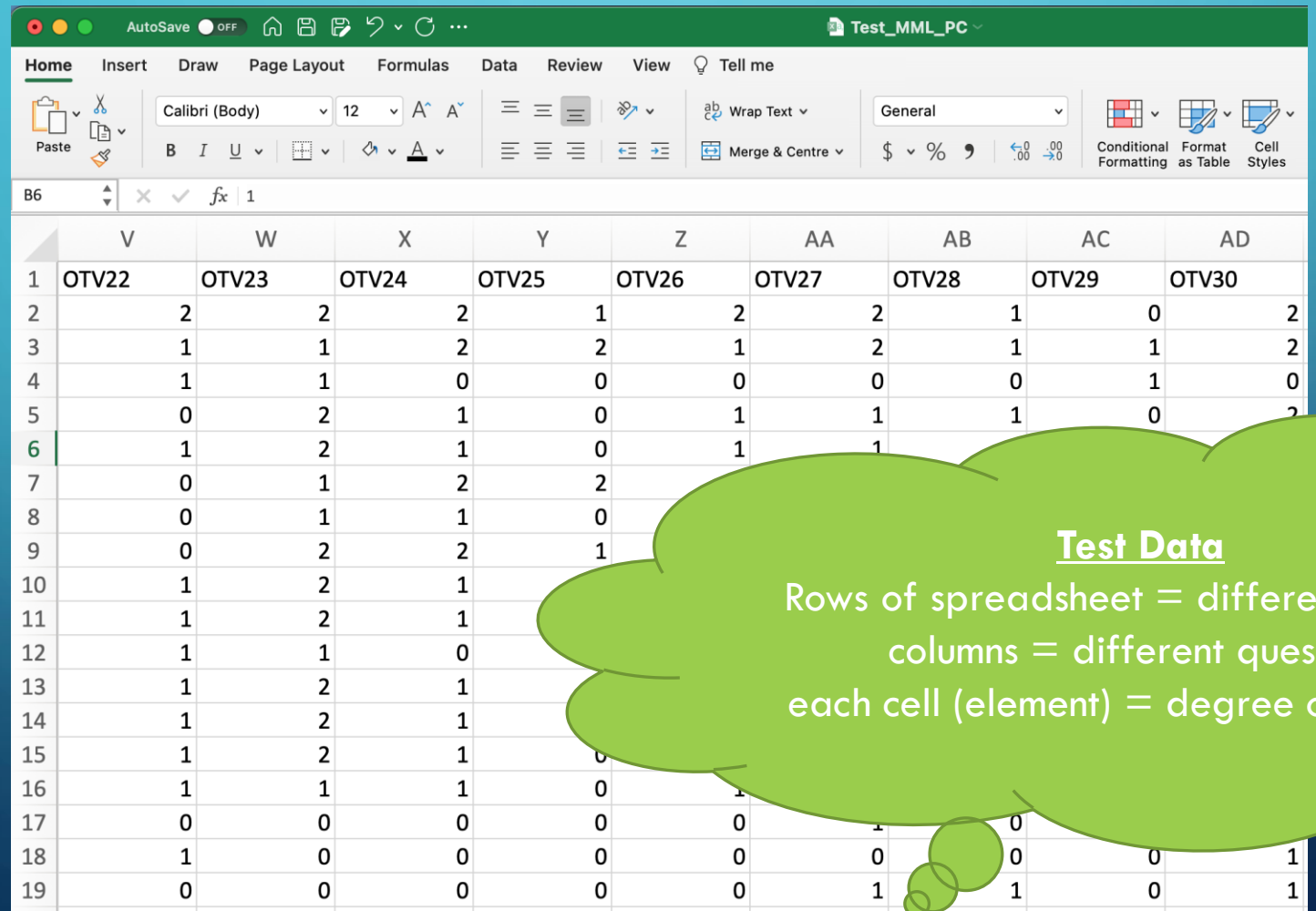
autopsych: An R Shiny Tool for the Reproducible Rasch Analysis, Differential Item Functioning, Equating, and Examination of Group Effects

Introduce Team :-)

Courtney, M. G. R., Chang, K., Mei, E., Meissel, K., Rowe, L., & Issayeva, L. (accepted). autopsych: An R Shiny Tool for the Reproducible Rasch Analysis, Differential Item Functioning, Equating, and Examination of Group Effects. ***PLOS ONE***. doi:10.1371/journal.pone.0257682

What is the autopsych app?

A web app that you can upload test data to and check for test validity and reliability (autopsych means “automated psychometrics”)



	V	W	X	Y	Z	AA	AB	AC	AD
1	OTV22	OTV23	OTV24	OTV25	OTV26	OTV27	OTV28	OTV29	OTV30
2	2	2	2	1	2	2	1	0	2
3	1	1	2	2	1	2	1	1	2
4	1	1	0	0	0	0	0	1	0
5	0	2	1	0	1	1	1	0	2
6	1	2	1	0	1	1			
7	0	1	2	2					
8	0	1	1	0					
9	0	2	2	1					
10	1	2	1						
11	1	2	1						
12	1	1	0						
13	1	2	1						
14	1	2	1						
15	1	2	1	0					
16	1	1	1	0	1				
17	0	0	0	0	0	1	0		
18	1	0	0	0	0	0	0	0	1
19	0	0	0	0	0	1	1	0	1

Test Data

Rows of spreadsheet = different students

columns = different questions

each cell (element) = degree of mastery

The autopsych app performs five tasks

1. Checks tests for their level of validity and reliability

2. Checks for bias questions in tests

3. Equates two tests (e.g., Grade 3 and Grade 4) to put different groups of students on a single scale

4. Provides a comparison in student performance between classes or experimental conditions

5. Estimates the inter-rater reliability of two raters on the same set of focal students

The autopsych app performs five tasks (tabs)

1 2 3 4 5

Home | Uni-Dim Rasch (MML) | Many-Facets Rasch (DIF) | Rasch Equating | ANOVA | Inter-Rater Reliability | autopsych Version 1.0.0 | Team | Highlights | Contact

Automated Psychometrics

Toward Valid Assessment and Educational Research

Introduction

Welcome to Automated Psychometrics, a novel website that allows teachers, school and university assessment experts, test developers, and researchers to:

(1) Check the general quality of student assessments and developmental rubrics,

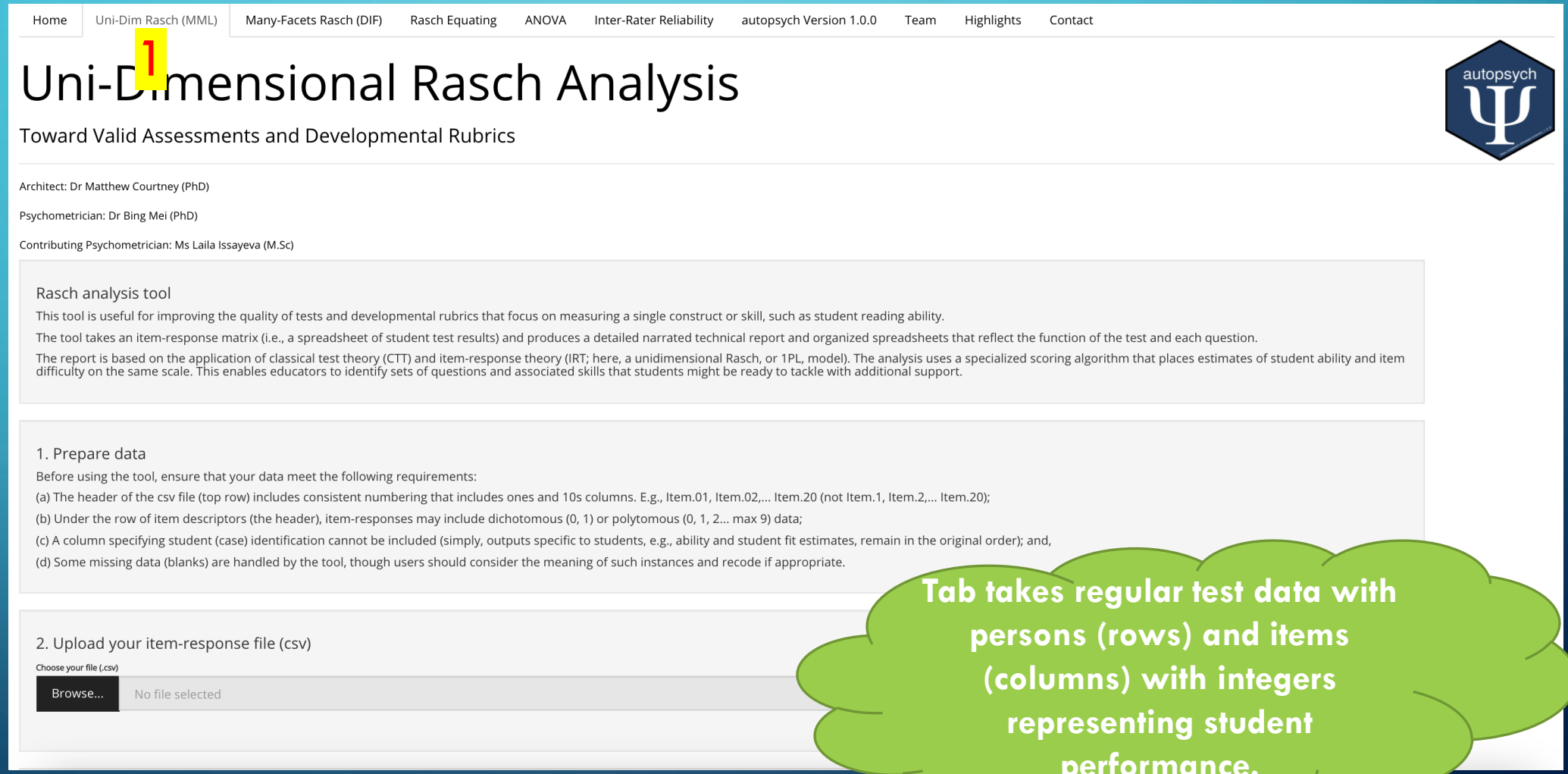
(2) Ensure test questions and developmental criterion are not bias toward any demographic group,

(3) Place students from different year groups on a single developmental scale via test equating,

Waiting for autopsych.shinyapps.io...

https://autopsych.shinyapps.io/version_1_0_0/

Task 1: checks tests for their level of validity and reliability



Tab takes regular test data with persons (rows) and items (columns) with integers representing student performance.

Task 2: checks for bias questions in tests

Home | Uni-Dim Rasch (MML) | Many-Facets Rasch (DIF) | Rasch Equating | ANOVA | Inter-Rater Reliability | autopsych Version 1.0.0 | Team | Highlights | Contact

Many-Facets 2asch Analysis

Toward Unbiased Test Questions and Developmental Criterion

Architect: Dr Matthew Courtney (PhD)

Psychometrician: Dr Bing Mei (PhD)

Psychometrician: Ms Laila Issayeva (M.Sc)

Many-facets Rasch analysis tool

This tool extends the functionality of the Uni-Dimensional Rasch analysis to include an examination of item (question) bias via the application of many-facets Rasch analysis. This form of analysis provides insight into how some questions (or developmental criteria) might function differently across student groups.

The tool also takes an item-response matrix (i.e., a spreadsheet of student test results). Though, the tool requires that the first column specifies the binary facet of interest (e.g., column header 'gender'). The variable needs to be numeric with coding 1 (representing male, for example) and 2 (representing female, for example). The report includes and produces a detailed narrated technical report and organized spreadsheets that reflect the function of the test and each question, as well as a report on item bias.

The report is based on the application of classical test theory (CTT) and item-response theory (IRT; here, a unidimensional Rasch, or 1PL, model, and extended many-facets analysis). The analysis uses a specialized scoring algorithm that places estimates of student ability and item difficulty on the same scale. This enables educators to identify sets of questions and associated skills that students might be ready to tackle with additional support. Analysts using this tool (as opposed to the JML tool, in production) will be primarily interested in generalizing the results of the analysis to the broader population from which the sample students were drawn. Insights into potential item bias can be helpful for checking that the scale operates in a reasonably similar way across groups of interest.

1. Prepare data

Before using the tool, ensure that your data meet the following requirements:

(a) The first column of the csv file is the binary facet of interest, e.g., gender. The coding is decided by the user such as 1 for male and 2 for female.

(b) The header of the csv file (top row) includes consistent numbering that includes ones and 10s columns. E.g., Item.01, Item.02,... Item.20 (not Item.1, Item.2,... Item.20)

(c) Under the row of item descriptors (the header), item-responses may include dichotomous (0, 1) or polytomous (0, 1, 2... max 9) data;

(d) A column specifying student (case) identification cannot be included (simply, outputs specific to students, e.g., ability and student fit estimates, remain in the original order); and,

(e) Some missing data (blanks) are handled by the tool, though users should consider the meaning of such instances and recode if appropriate.

2. Upload your item-response file (csv)

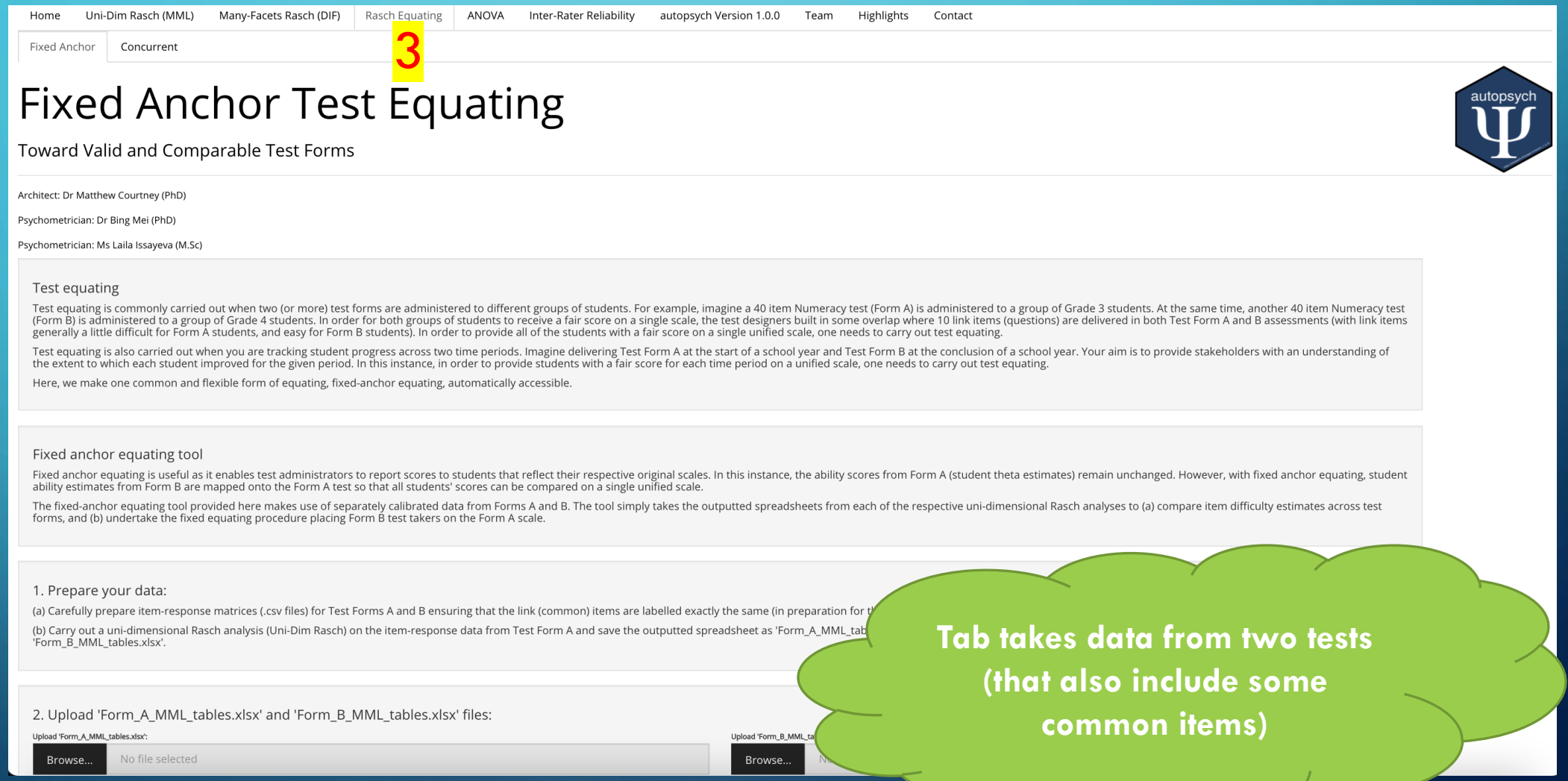
Choose your file (.csv)

Browse... No file selected

3. Specify construct and focal group

Tab takes regular test data with an extra column for student gender, ethnicity, or language group, etc. Checks for item bias.

Task 3: equates two tests (e.g., Grade 3 and Grade 4) to put different groups of students on a single scale



Tab takes data from two tests (that also include some common items)

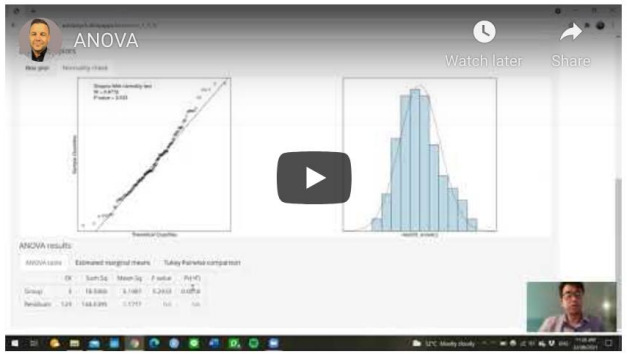
Task 4: provides a comparison in student performance between classes or experimental conditions

Home | Uni-Dim Rasch (MML) | Many-Facets Rasch (DIF) | Rasch Equating | ANOVA | Inter-Rater Reliability | autopsych Version 1.0.0 | Team | Highlights | Contact

4

One-Way ANOVA Analysis

Toward Valid Examinations of Group Differences



Architect: Dr Kevin Chang (PhD)

Psychometrician: Dr Matthew Courtney (PhD)

One-Way ANOVA Tool

This tool provides a convenient way to examine the effect of student grouping (such as student gender, class, or school classification) on student ability or some measured personal

On this tab, users upload their outputted spreadsheet from their Rasch analysis. In addition, users also upload another dataset that includes as many gr
number of rows as the Rasch spreadsheet as each grouping variable, e.g., gender, class, needs to correspond to the same ability estimate

Data inputs

Choose your file from Uni-Dim Rasch tab

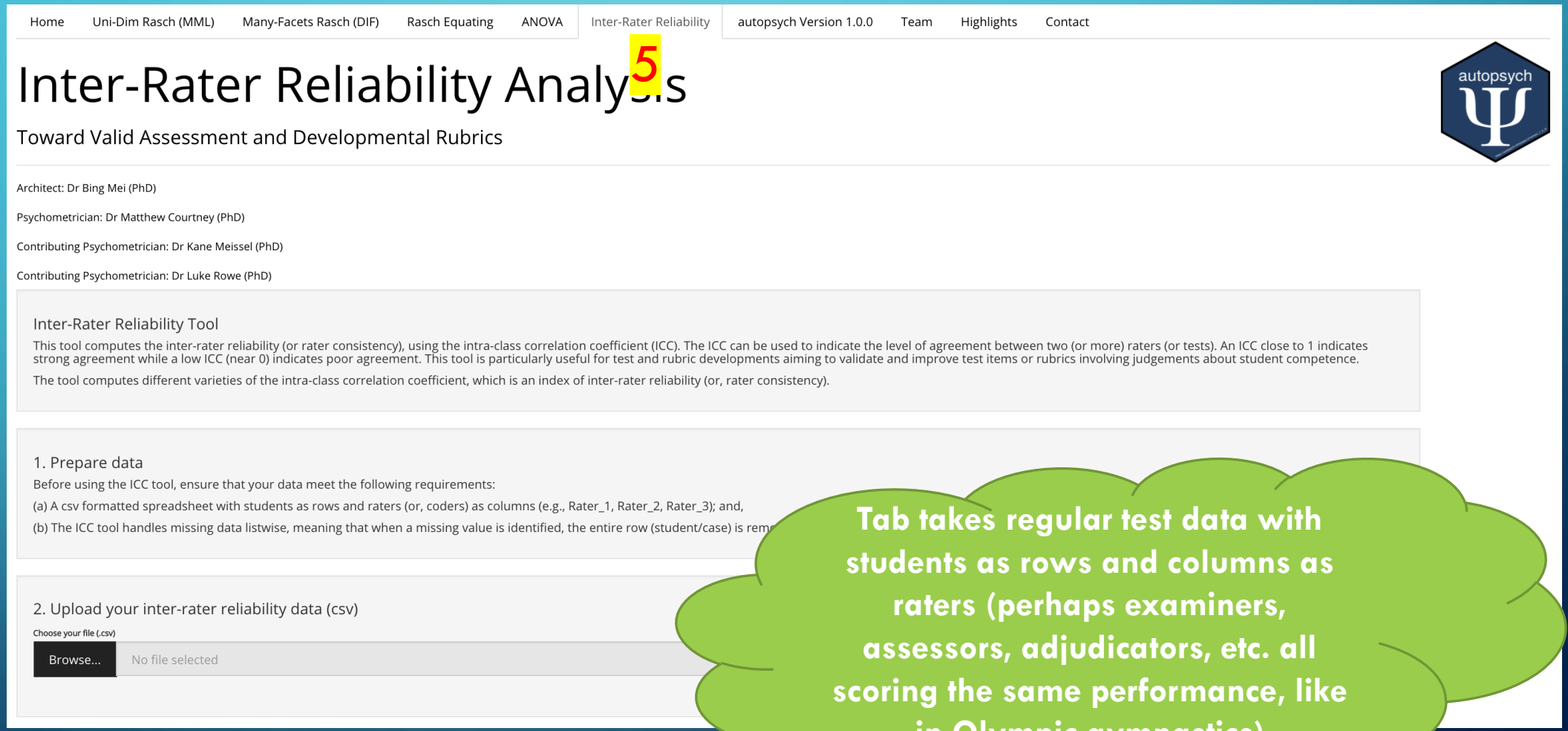
Browse... | No file selected

Choose your file from con

Browse...

Tab takes regular test data and group data (gender, school, or ethnic group). Checks for differences in overall performance.

Task 5: estimates the inter-rater reliability of two raters on the same set of focal students



Tab takes regular test data with students as rows and columns as raters (perhaps examiners, assessors, adjudicators, etc. all scoring the same performance, like in Olympic gymnastics)

We will now provide a video demonstration with real test data for each of the five tasks…

After that, feel free to ask questions…