

# Lab 2 - Competing Path Models

Structural Equation Modeling ED 216F - Instructor: Karen Nylund-Gibson

*Adam Garber*

April 09, 2020

## Contents

<b>1</b>	<b>Lab preparation</b>	<b>1</b>
1.1	Creating a version-controlled R-Project by downloading repository from Github . . . . .	2
1.2	Data source: . . . . .	3
1.3	List of over 1000 datasets available in R packages . . . . .	3
<b>2</b>	<b>Begin lab 2 exercise</b>	<b>3</b>
<b>3</b>	<b>Explore the data</b>	<b>4</b>
<b>4</b>	<b>Specifying path models using {MplusAutomation}</b>	<b>7</b>
4.1	Estimate model 1 . . . . .	8
4.2	Estimate model 2 . . . . .	10
4.3	Estimate model 3 . . . . .	12
4.4	Estimate model 4 . . . . .	13
4.5	Estimate model 5 . . . . .	14
<b>5</b>	<b>Compare model fit</b>	<b>16</b>
<b>6</b>	<b>End of Lab 2</b>	<b>17</b>
<b>7</b>	<b>References</b>	<b>17</b>

---

## 1 Lab preparation

---

## 1.1 Creating a version-controlled R-Project by downloading repository from Github

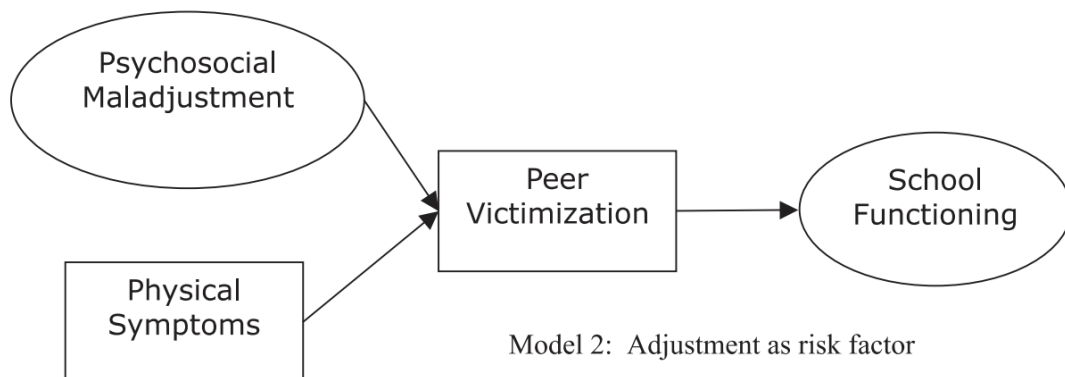
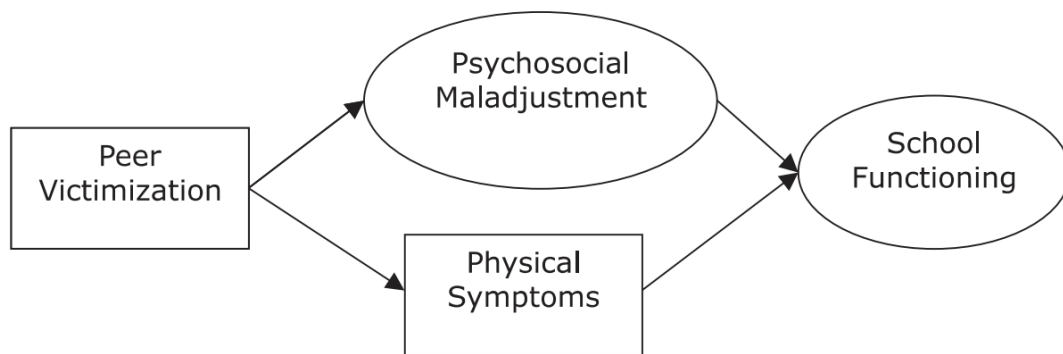
Download repository here: <https://github.com/garberadamc/SEM-Lab2>

On the Github repository webpage:

- a. **fork** your own **branch** of the lab repository
- b. copy the repository web URL address from the **clone** or **download** menu

Within R-Studio:

- c. click “NEW PROJECT” (upper right corner of window)
- d. choose option **Version Control**
- e. choose option **Git**
- f. paste the repository web URL path copied from the **clone** or **download** menu on Github page
- g. choose location of the R-Project (**too many nested folders will result in filepath error**)



*figure.* Picture adapted from Nishina, Juvonen, Witkow (2005)

---

## 1.2 Data source:

This lab exercise utilizes the *California Test Score Data Set 1998-1999* from the California Department of Education (Stock, James, and Watson, 2003) [See documentation here](#)

This dataset is available via the R-package {Ecdat} and can be directly loaded into the R environment.

---

## 1.3 List of over 1000 datasets available in R packages

This list was compiled by Vincent Arel-Bundock and can be found here:

<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

---

Install the “rhdf5” package to read gh5 files

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("rhdf5")
```

---

Load packages

```
library(MplusAutomation)
library(haven)
library(rhdf5)
library(tidyverse)
library(here)
library(corrplot)
library(kableExtra)
library(reshape2)
library(janitor)
library(ggribes)
library(DiagrammerR)
library(semPlot)
library(sjPlot)
library(Ecdat)
library(gt)
library(gtsummary)
```

---

## 2 Begin lab 2 exercise

---

Read the dataframe into your R-environment from package {Ecdat}

```
data(Caschool)

ca_schools <- as.data.frame(Caschool)
```

Look at the data with `glimpse`

```
glimpse(ca_schools)
```

---

Subset variables to use in path model analyses with `select`

```
path_vars <- ca_schools %>%
  select(str, expnstu, compstu, elpct, mealpct,
         readscr, mathscr, testscr)
```

---

### 3 Explore the data

K through 8th grade schools in California ( $N = 420$ )

Take a look at focal variables, make a `tribble` table

```
var_table <- tribble(
  ~"Name",    ~"Labels",
  #-----/-----/,
  "str"       , "student teacher ratio"      ,
  "expnstu"   , "expenditure per student"    ,
  "compstu"   , "computer per student"       ,
  "elpct"     , "percent of English learners"   ,
  "mealpct"   , "percent qualifying for reduced-price lunch" ,
  "readscr"   , "average reading score"       ,
  "mathscr"   , "average math score"         ,
  "testscr"   , "average test score (read.scr+math.scr)/2" )

var_table %>%
  kable(booktabs = T, linesep = "") %>%
  kable_styling(latex_options = c("striped"),
                full_width = F,
                position = "left")
```

Name	Labels
str	student teacher ratio
expnstu	expenditure per student
compstu	computer per student
elpct	percent of English learners
mealpct	percent qualifying for reduced-price lunch
readscr	average reading score
mathscr	average math score
testscr	average test score (read.scr+math.scr)/2

---

check some basic descriptives with the `{gtsummary}` package

```
table1 <- tbl_summary(path_vars,
  statistic = list(all_continuous() ~ "{mean} ({sd})"),
  missing = "no" ) %>%
  bold_labels()

table1
```

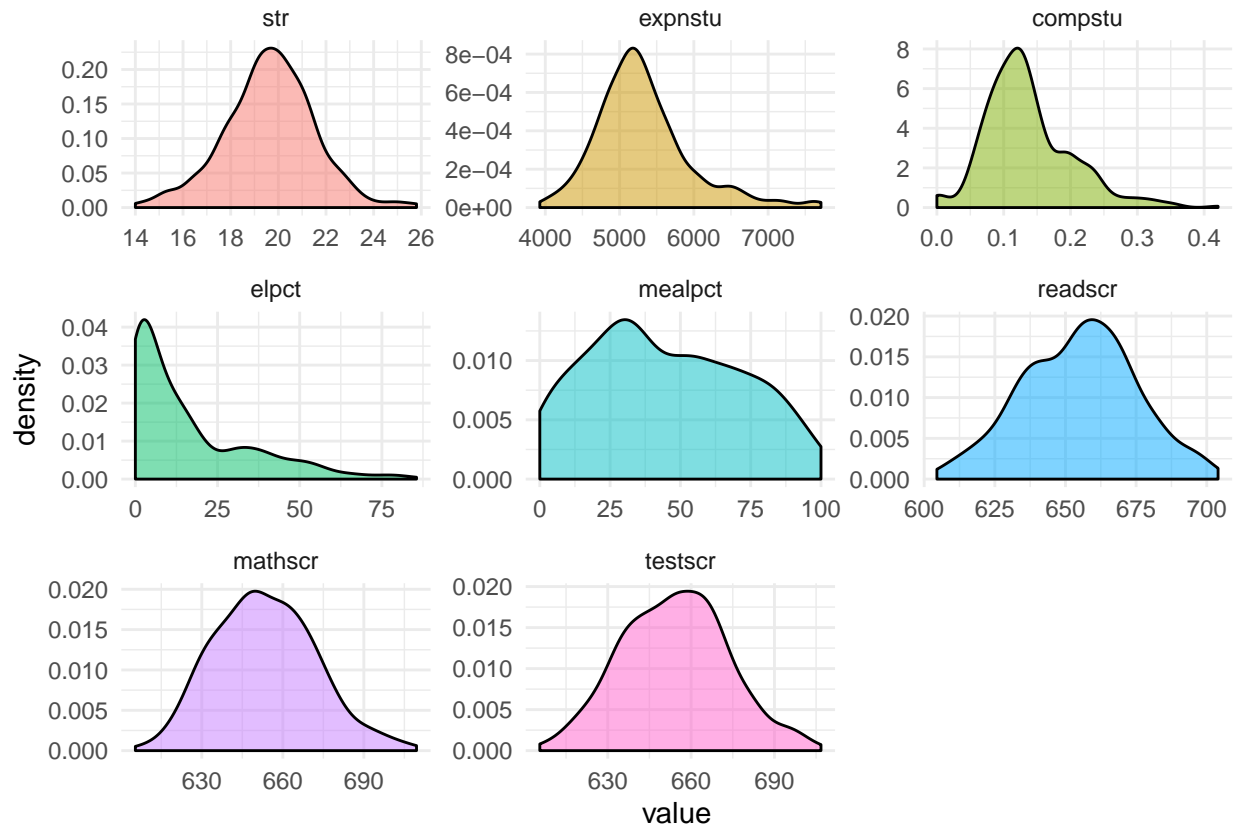
Characteristic	N = 420 <sup>1</sup>
str	19.64 (1.89)
expnstu	5312 (634)
compstu	0.14 (0.06)
elpct	16 (18)
mealpct	45 (27)
readscr	655 (20)
mathscr	653 (19)
testscr	654 (19)

<sup>1</sup>Statistics presented: mean (SD)

---

look at shape of variable distributions

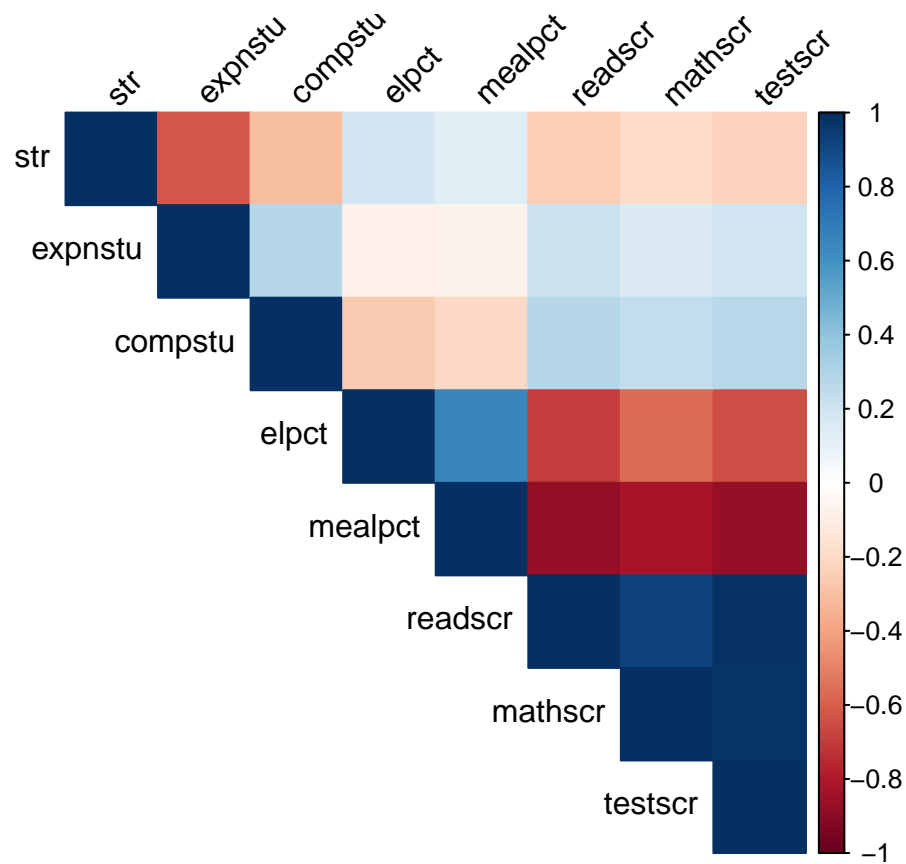
```
melt(path_vars) %>%
  ggplot(., aes(x=value, label=variable)) +
  geom_density(aes(fill = variable),
    alpha = .5, show.legend = FALSE) +
  facet_wrap(~variable, scales = "free") +
  theme_minimal()
```



look at correlation matrix with {corrplot}

```
p_cor <- cor(path_vars, use = "pairwise.complete.obs")

corrplot(p_cor,
  method = "color",
  type = "upper",
  tl.col="black",
  tl.srt=45)
```



#### 4 Specifying path models using {MplusAutomation}

recall what the unrestricted variance-covariance matrix **looks** like

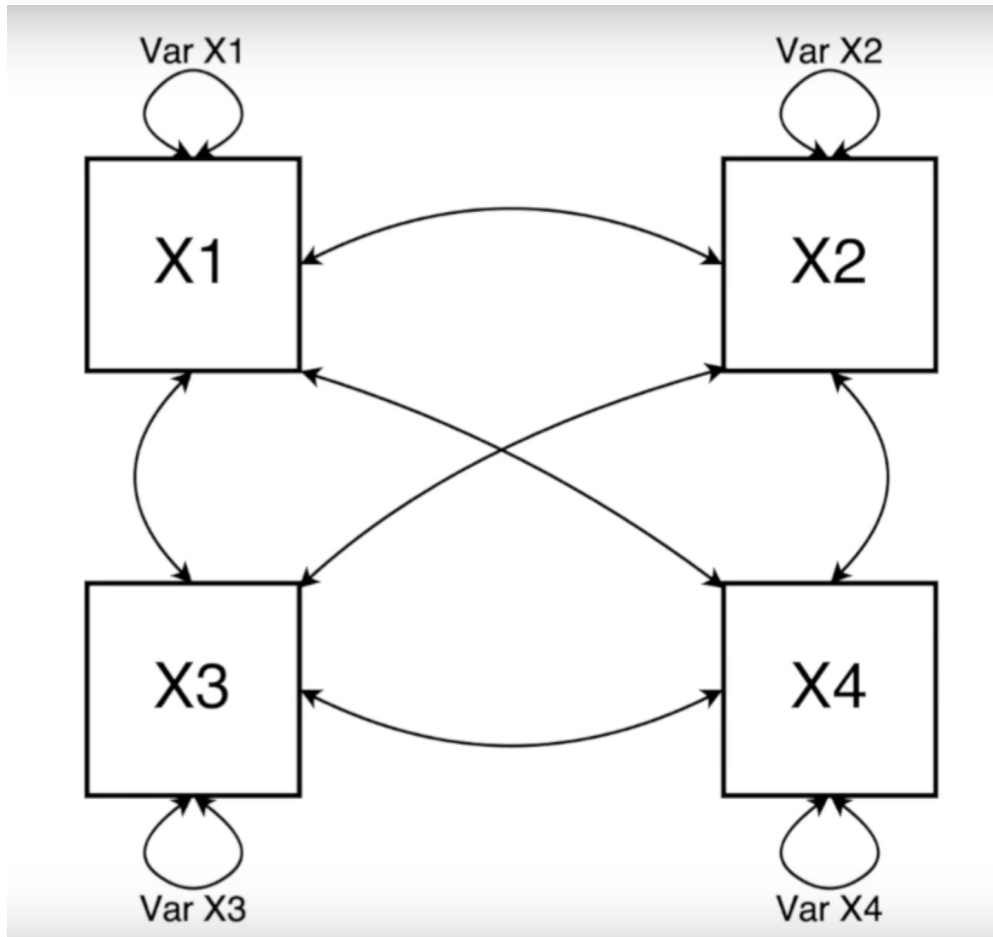


figure. Unrestricted variance covariance matrix picture from {openMX} video tutorial.

---

## 4.1 Estimate model 1

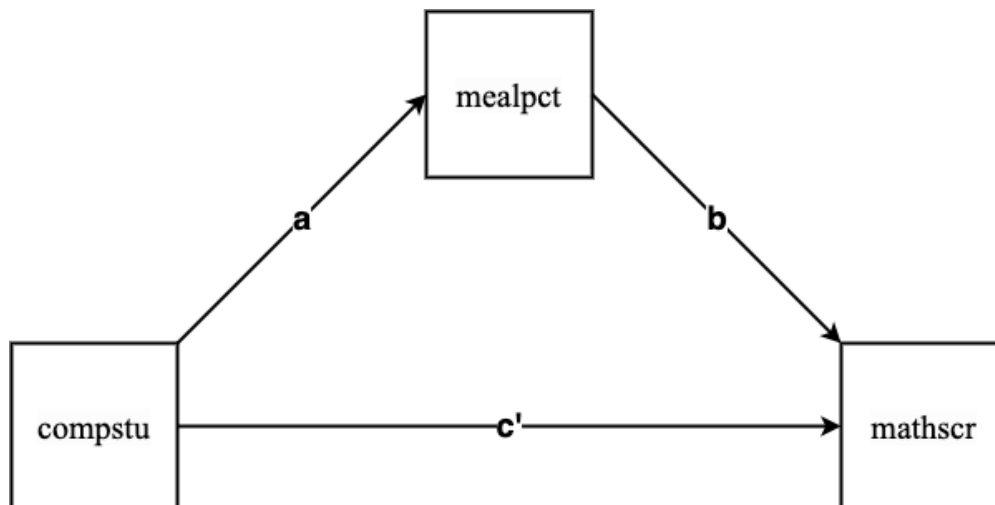
Indirect path model:

1. covariate: ratio of computers to students (`compstu`)
2. mediator: percent qualifying for reduced-price lunch (`mealpct`)
3. outcome: average math score (`mathscr`)

---

Path diagram model 1






---

```

m1_path <- mplusObject(
  TITLE = "m1 model indirect - Lab 1",
  VARIABLE =
    "usevar =
      compstu      ! covariate
      mealpct      ! mediator
      mathscr;      ! outcome",

  ANALYSIS =
    "estimator = MLR" ,

  MODEL =
    "mathscr on compstu;          ! direct path (c')
    mathscr on mealpct;          ! b path
    mealpct on compstu;          ! a path

    Model indirect:
    mathscr ind compstu;" ,

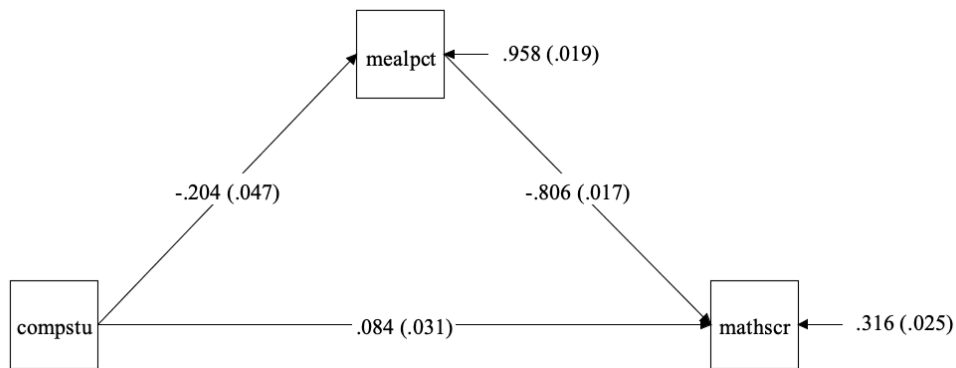
  OUTPUT = "sampstat standardized modindices (ALL)",

  usevariables = colnames(path_vars),
  rdata = path_vars)

m1_path_fit <- mplusModeler(m1_path,
  dataout=here("mplus_files", "Lab2.dat"),
  modelout=here("mplus_files", "m1_path_Lab2.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)
  
```

---

View path diagram for model 1 with standardized estimates (using Diagrammer in Mplus)



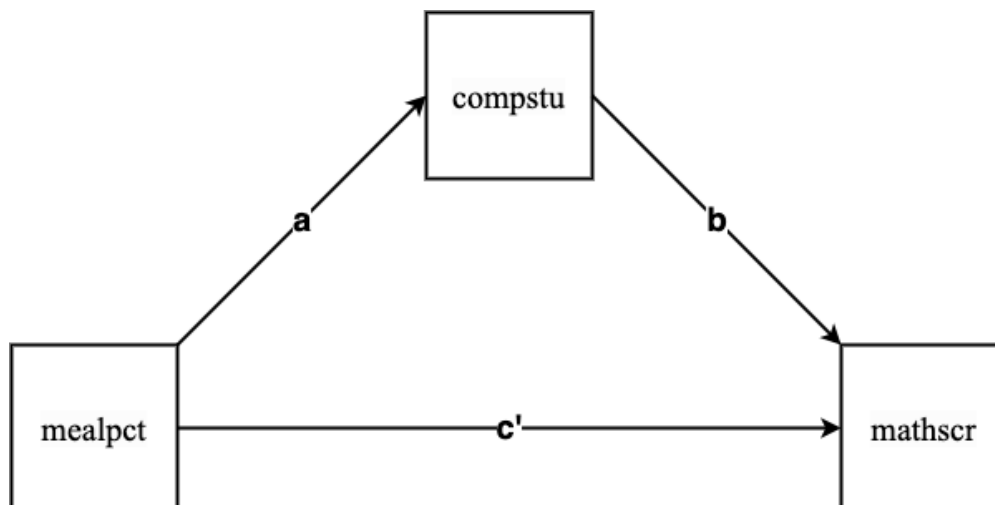
## 4.2 Estimate model 2

change variable status (**switch mediator and covariate variables**)

Indirect path model:

1. covariate: percent qualifying for reduced-price lunch (**mealpct**)
  2. mediator: ratio of computers to students (**compstu**)
  3. outcome: average math score (**mathscr**)
- 

Path diagram model 2



```

m2_path <- mplusObject(
  TITLE = "m1 model indirect - Lab 1",
  VARIABLE =
    "usevar =
      mealpct          ! covariate
      compstu          ! mediator
      mathscr;         ! outcome",

  ANALYSIS =
    "estimator = MLR" ,

  MODEL =
    "mathscr on compstu;          ! direct path (c')
    mathscr on mealpct;          ! b path
    mealpct on compstu;          ! a path

    Model indirect:
    mathscr ind compstu;" ,

  OUTPUT = "sampstat standardized modindices (ALL)",

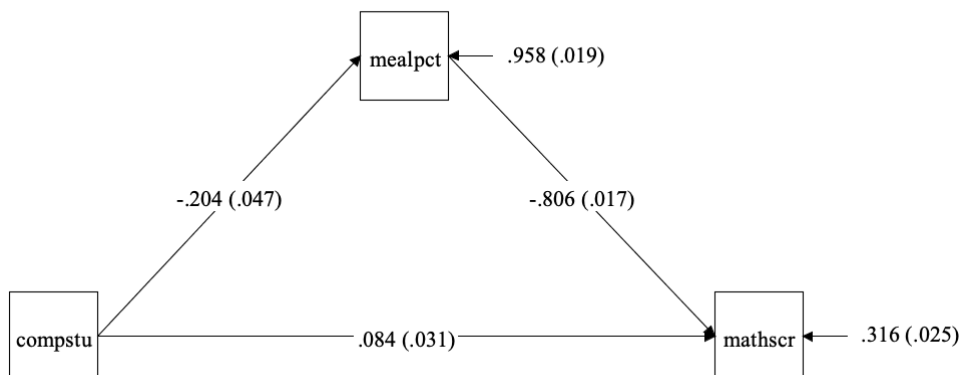
  usevariables = colnames(path_vars),
  rdata = path_vars)

m2_path_fit <- mplusModeler(m2_path,
  dataout=here("mplus_files", "Lab2.dat"),
  modelout=here("mplus_files", "m2_path_Lab2.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

---

View path diagram for model 2 with standardized estimates (using the Diagrammer in Mplus)

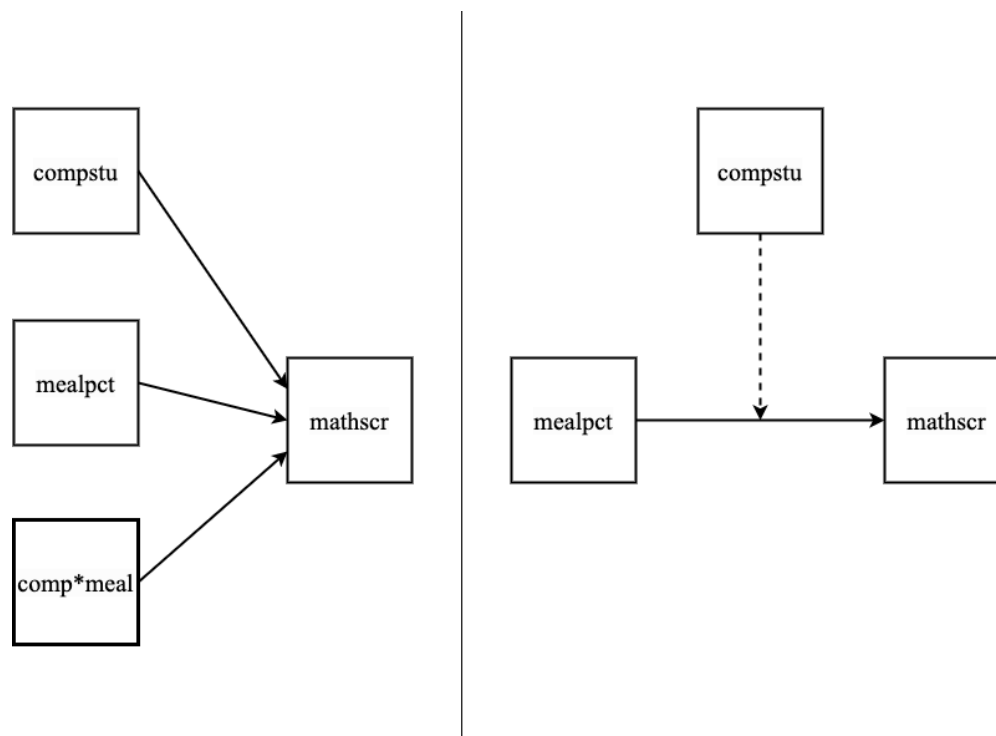


### 4.3 Estimate model 3

Path model with interaction (moderation):

1. covariate-moderator: percent qualifying for reduced-price lunch (`mealpct`)
2. covariate-moderator: ratio of computers to students (`compstu`)
3. outcome: average math score (`mathscr`)

Path diagram model 3



```
m3_path <- mplusObject(  
  TITLE = "m1 model indirect - Lab 1",  
  VARIABLE =  
    "usevar =  
      compstu      ! covariate-moderator  
      mealpct      ! covariate-moderator  
      mathscr      ! outcome  
      int_ab;      ! interaction term ",  
  
  DEFINE =  
    "int_ab = compstu*mealpct; ! create interaction term" ,  
  
  ANALYSIS =  
    "estimator = MLR" ,  
)
```

```

MODEL =
  "mathscr on compstu mealpct int_ab; ",

OUTPUT = "sampstat standardized modindices (ALL)",

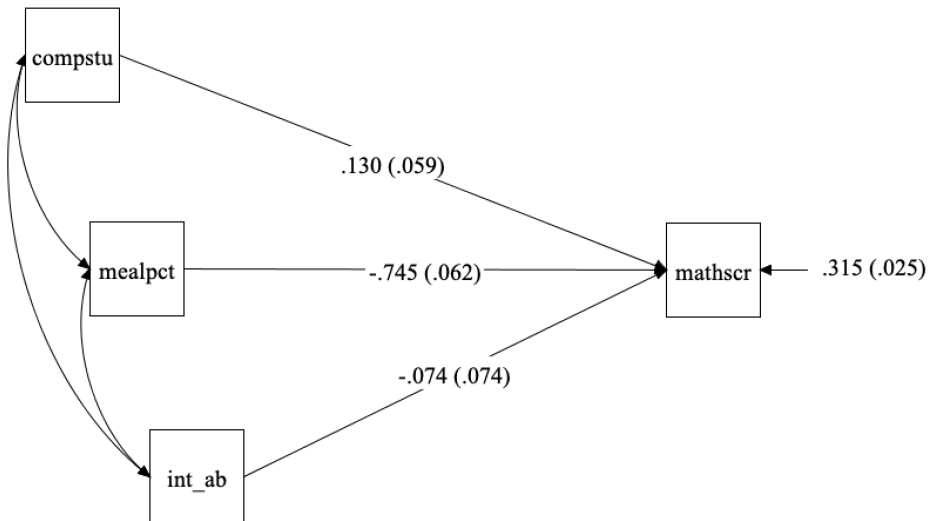
usevariables = colnames(path_vars),
rdata = path_vars)

m3_path_fit <- mplusModeler(m3_path,
  dataout=here("mplus_files", "Lab2.dat"),
  modelout=here("mplus_files", "m3_path_Lab2.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

---

View path diagram for model 3 with standardized estimates (using the Diagrammer in Mplus)




---

#### 4.4 Estimate model 4

---

```

m4_path <- mplusObject(
  TITLE = "m4 model indirect - Lab 1",
  VARIABLE =
    "usevar =
      str          ! covariate
      elpct        ! mediator
      mealpct      ! mediator
      mathscr      ! outcome",

```

```

DEFINE =
  "int_ab = compstu*mealpct;  ! create interaction term" ,

ANALYSIS =
  "estimator = MLR" ,

MODEL =
  "mathscr on str;           ! direct path (c')
  mathscr on elpct mealpct;  ! b paths
  elpct mealpct on str;     ! a paths

  Model indirect:
  mathscr ind str;" ,

OUTPUT = "sampstat standardized modindices (ALL)",

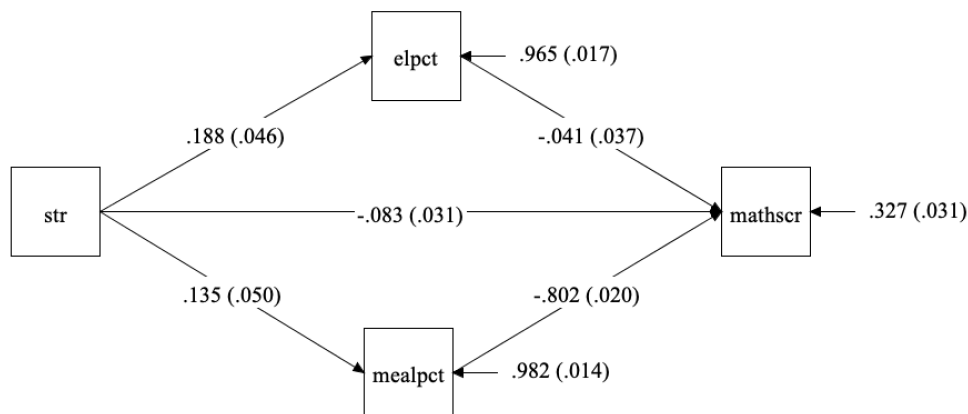
usevariables = colnames(path_vars),
rdata = path_vars)

m4_path_fit <- mplusModeler(m4_path,
  dataout=here("mplus_files", "Lab2.dat"),
  modelout=here("mplus_files", "m4_path_Lab2.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

---

View path diagram for model 4 with standardized estimates (using the Diagrammer in Mplus)




---

## 4.5 Estimate model 5

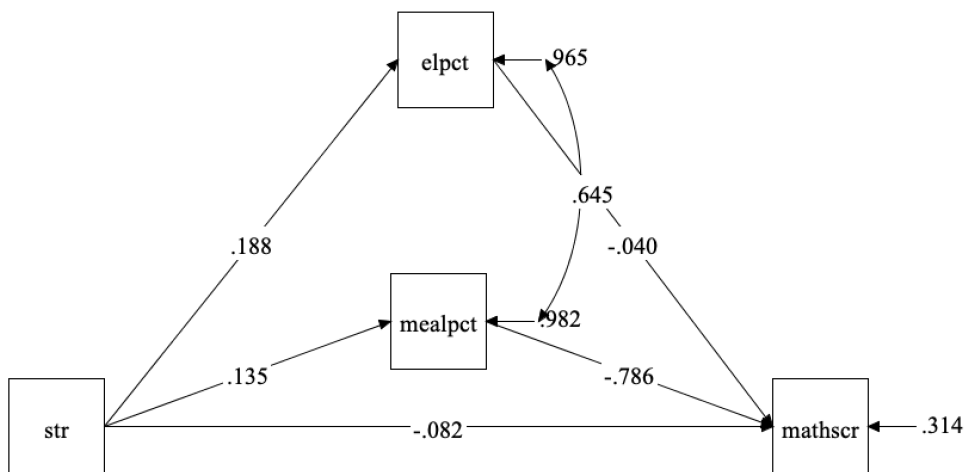
---

add modification statement - correlate mediators mealpct with elpct

```
m5_path <- mplusObject(  
  TITLE = "m5 model indirect - Lab 1",  
  VARIABLE =  
    "usevar =  
      str          ! covariate  
      elpct        ! mediator  
      mealpct      ! mediator  
      mathscr      ! outcome",  
  
  DEFINE =  
    "int_ab = compstu*mealpct; ! create interaction term" ,  
  
  ANALYSIS =  
    "estimator = MLR" ,  
  
  MODEL =  
    "mathscr on str;          ! direct path (c')  
    mathscr on elpct mealpct; ! b paths  
    elpct mealpct on str;    ! a paths  
  
    mealpct with elpct      ! modification statement  
  
    Model indirect:  
    mathscr ind str; " ,  
  
  OUTPUT = "sampstat standardized modindices (ALL)",  
  
  usevariables = colnames(path_vars),  
  rdata = path_vars)  
  
m5_path_fit <- mplusModeler(m5_path,  
  dataout=here("mplus_files", "Lab2.dat"),  
  modelout=here("mplus_files", "m5_path_Lab2.inp"),  
  check=TRUE, run = TRUE, hashfilename = FALSE)
```

---

View path diagram for model 5 with standardized estimates (using the Diagrammer in Mplus)



## 5 Compare model fit

Read into R summary of all models

```
all_models <- readModels(here("mplus_files"))
```

Extract fit indice data from output files

```
summary_fit <- LatexSummaryTable(all_models,
  keepCols=c("Filename", "Parameters", "ChiSqM_Value", "CFI", "TLI",
    "SRMR", "RMSEA_Estimate", "RMSEA_90CI_LB", "RMSEA_90CI_UB"),
  sortBy = "Filename")
```

Create a customizable table using the {gt} package

```
model_table <- summary_fit %>%
  gt() %>%
  tab_header(
    title = "Fit Indices", # Add a title
    subtitle = "" # And a subtitle
  ) %>%
  tab_options(
    table.width = pct(80)
  ) %>%
  tab_footnote(
    footnote = "California Test Score Data Set 1998-1999",
    location = cells_title()
  ) %>%
```



```
cols_label(  
  Filename = "Model",  
  Parameters = "Par",  
  ChiSqM_Value = "ChiSq",  
  RMSEA_Estimate = "RMSEA",  
  RMSEA_90CI_LB = "Lower CI",  
  RMSEA_90CI_UB = "Upper CI")  
  
model_table
```

---

## 6 End of Lab 2

---

## 7 References

- Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural equation modeling: a multidisciplinary journal*, 25(4), 621-638.
- Horst, A. (2020). Course & Workshop Materials. GitHub Repositories, <https://allisonhorst.github.io/>
- Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., . . . & Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSL: 09): Base-Year Data File Documentation. NCES 2011-328. National Center for Education Statistics.
- Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

---

# UC SANTA BARBARA