REIMS sample analysis

Nathan Alexander

2024-06-03

Set up

Set up `MPlusAutomation`

```
install.packages("devtools")
library(devtools)

install_github("michaelhallquist/MplusAutomation")
```

Load packages

```
## Version:  1.2
## We work hard to write this free software. Please help us get credit by citing:
##
## Hallquist, M. N. & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale La
##
## -- see citation("MplusAutomation").

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract() masks MplusAutomation::extract()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## here() starts at /Users/nathanalexander/Dropbox/Projects/immerse
##
##
## Attaching package: 'data.table'
##
##
## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
##
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
##
## The following object is masked from 'package:purrr':
##
##     transpose
##
##
## here() starts at /Users/nathanalexander/Dropbox/Projects/immerse/reims
```

Data

```
# set data
reims <- read_csv(here("data", "reims_clean.csv"))
```

```
## New names:
## Rows: 103 Columns: 57
## -- Column specification
## -------------------------------------------------------- Delimiter: "," dbl
## (57): groupflag, age, sex, race, mathperson1, mathperson2, mathperson3, ...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `belongrace...19` -> `belongrace`
```

```
# inspect data
reims
```

```
## # A tibble: 103 x 57
##    groupflag   age   sex  race mathperson1 mathperson2 mathperson3 mathperson4
##        <dbl> <dbl> <dbl> <dbl>       <dbl>       <dbl>       <dbl>       <dbl>
##  1         1    18     2     6           0           0           0           0
##  2         1    22     2     1           0           0           0           0
##  3         1    20     1     7           1           1           1           1
##  4         1    20     1     3           1           1           1           1
##  5         1    18     1     4           0           0           0           1
##  6         1    18     2     6           0           0           0           0
##  7         1    19     1     6           0           0           0           0
##  8         1    23     2     4           0           0           0           0
##  9         1    37     2     3           0           0           0           0
## 10         1    27     2     6           0           0           0           0
## # i 93 more rows
## # i 49 more variables: dislikemathclass <dbl>, pursuestem <dbl>,
## #   boysbetter <dbl>, learnrace <dbl>, racegroups <dbl>, knowrace <dbl>,
## #   connectrace1 <dbl>, affectrace <dbl>, proudrace <dbl>, mixrace <dbl>,
## #   unclearrace <dbl>, connectrace2 <dbl>, dontknowrace <dbl>,
## #   belongrace <dbl>, understandrace <dbl>, talkrace <dbl>, priderace <dbl>,
## #   avoidrace <dbl>, practicerace <dbl>, playotherrace <dbl>, ...
```

```
summary(reims)
```

```
##    groupflag           age              sex             race
##  Min.   :0.0000   Min.   :14.00   Min.   :1.00   Min.   :1.00
##  1st Qu.:1.0000   1st Qu.:18.00   1st Qu.:1.00   1st Qu.:3.00
##  Median :1.0000   Median :20.00   Median :2.00   Median :4.00
##  Mean   :0.8835   Mean   :20.82   Mean   :1.65   Mean   :3.99
##  3rd Qu.:1.0000   3rd Qu.:21.00   3rd Qu.:2.00   3rd Qu.:6.00
##  Max.   :1.0000   Max.   :47.00   Max.   :3.00   Max.   :7.00
##   mathperson1      mathperson2      mathperson3      mathperson4
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :1.0000   Median :1.0000   Median :1.0000
##  Mean   :0.5922   Mean   :0.5631   Mean   :0.5146   Mean   :0.5146
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  dislikemathclass   pursuestem       boysbetter        learnrace
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :0.00000   Median :0.0000
##  Mean   :0.2524   Mean   :0.5534   Mean   :0.07767   Mean   :0.4078
##  3rd Qu.:0.5000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##    racegroups        knowrace       connectrace1      affectrace
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.3301   Mean   :0.6311   Mean   :0.8641   Mean   :0.4563
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    proudrace         mixrace         unclearrace      connectrace2
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.00000   Median :0.0000   Median :1.0000
##  Mean   :0.8252   Mean   :0.08738   Mean   :0.2427   Mean   :0.6602
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##   dontknowrace      belongrace     understandrace      talkrace
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.3107   Mean   :0.5146   Mean   :0.5437   Mean   :0.4563
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    priderace        avoidrace        practicerace    playotherrace
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :1.0000   Median :0.00000   Median :1.0000   Median :1.000
##  Mean   :0.5631   Mean   :0.05825   Mean   :0.5049   Mean   :0.767
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.000
##    strongrace     enjoyotherrace      feelrace      racediscrimination
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
```

```
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.5049   Mean   :0.8932   Mean   :0.7184   Mean   :0.4369
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  academicability_peers mathability_peers   activities     deciderules
##  Min.   :0.0000        Min.   :0.0000    Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000        1st Qu.:0.0000    1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000        Median :1.0000    Median :1.000   Median :0.0000
##  Mean   :0.4854        Mean   :0.5146    Mean   :0.534   Mean   :0.2039
##  3rd Qu.:1.0000        3rd Qu.:1.0000    3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :1.0000        Max.   :1.0000    Max.   :1.000   Max.   :1.0000
##    makeadiff        adultcares      adultnotices     adultlistens
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :1.0000   Median :0.0000   Median :1.000
##  Mean   :0.4175   Mean   :0.5437   Mean   :0.4951   Mean   :0.699
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##   adultpraise      adultmybest     adultmysuccess    mtchmematter
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.6699   Mean   :0.7087   Mean   :0.7087   Mean   :0.1456
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  mtchlowerstandards mtchraceexpect     mtchtalkrace    mtchignorerace
##  Min.   :0.00000    Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000    1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000    Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.05825    Mean   :0.05825   Mean   :0.2524   Mean   :0.1748
##  3rd Qu.:0.00000    3rd Qu.:0.00000   3rd Qu.:0.5000   3rd Qu.:0.0000
##  Max.   :1.00000    Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##    mtchvalues     mtchrespect        mtchfair      mtchsuccess
##  Min.   :0.000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:1.0000   1st Qu.:1.000   1st Qu.:1.0000
##  Median :1.000   Median :1.0000   Median :1.000   Median :1.0000
##  Mean   :0.835   Mean   :0.8641   Mean   :0.835   Mean   :0.8155
##  3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##  mtchmistakesok     mtchbiased     mtchmadeinteresting mtchgenderbias
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000      Min.   :0.00000
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000      1st Qu.:0.00000
##  Median :1.0000   Median :0.0000   Median :1.0000      Median :0.00000
##  Mean   :0.8252   Mean   :0.1845   Mean   :0.6602      Mean   :0.06796
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000      3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000      Max.   :1.00000
##   mtchmadeeasy
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :1.0000
##  Mean   :0.5922
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

Make updates/edits to data We observe some of the other tags on our data. We also look at the distribution of some values. Take note that we changed the variable names (thanks Dina!) to less than eight characters so that our model could run.

```
# view distribution of indicator variables
table(reims$groupflag)
```

```
##
##  0  1
## 12 91
```

```
hist(reims$age)
```


```
table(reims$sex)
```

```
##
##  1  2  3
## 37 65  1
```

```
table(reims$race)
```

```
##
##  1  2  3  4  5  6  7
## 23  1 12 29  1 30  7
```

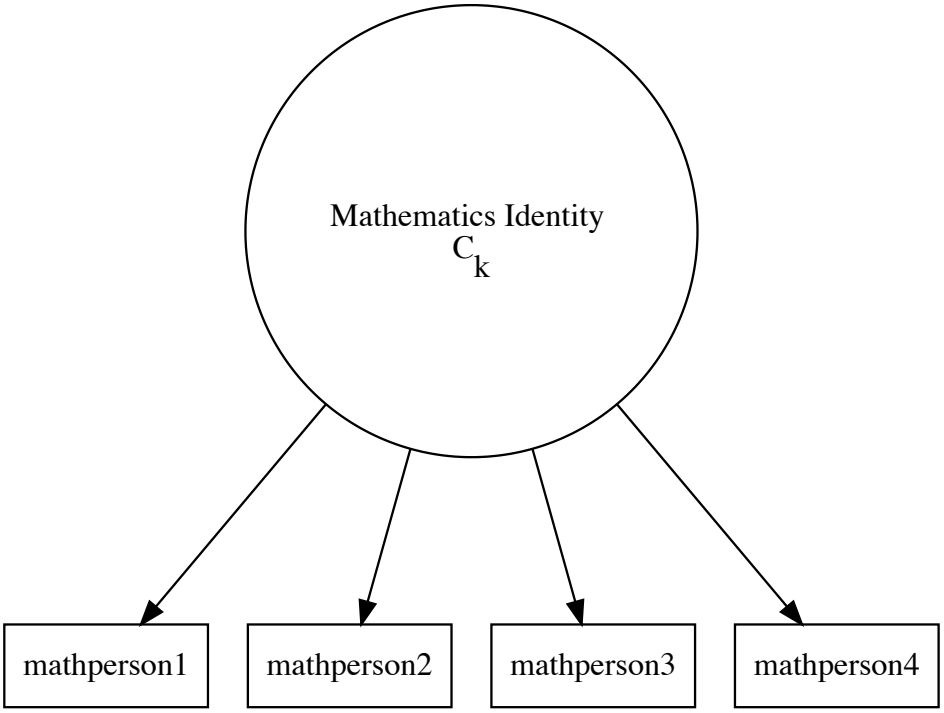
```r
# subset data
reims1 <- reims %>%
  select(mathperson1, mathperson2, mathperson3, mathperson4, dislikemathclass, pursuestem, boysbetter) %
  rename(m1 = mathperson1,
         m2 = mathperson2,
         m3 = mathperson3,
         m4 = mathperson4,
         dislike = dislikemathclass,
         pursue = pursuestem,
         boys = boysbetter)

head(reims1, n=10)
```

```
## # A tibble: 10 x 7
##       m1    m2    m3    m4 dislike pursue  boys
##    <dbl> <dbl> <dbl> <dbl>   <dbl>  <dbl> <dbl>
##  1     0     0     0     0       0      0     0
##  2     0     0     0     0       1      0     0
##  3     1     1     1     1       0      0     0
##  4     1     1     1     1       1      1     1
##  5     0     0     0     1       0      0     0
##  6     0     0     0     0       1      0     0
##  7     0     0     0     0       0      1     0
##  8     0     0     0     0       0      0     0
##  9     0     0     0     0       1      0     0
## 10     0     0     0     0       1      1     0
```

```r
## add ids and covariates; tell mplus that what we are

# use the reorder function to get hte variables that you want to model in the output.
```

Models

Model 1



Model 1 MplusAutomation code

```
input <- mplusObject(
  TITLE = "REIMS Mathematics Identity Model 1",
  VARIABLE = "categorical = m1 m2 m3 m4;
  usevar = m1-m4;
  classes = c(3);",

  ANALYSIS =
    "estimator = mlr;
    type = mixture;",

  OUTPUT = "tech11 tech14;",

  PLOT = "type = plot3;
    series = m1-m4(*);",

  usevariables = colnames(reims1),
  rdata = reims1)

output <- mplusModeler(input,
                       dataout = here("mplus", "reims1.dat"),
                       modelout = here("mplus", "reims1.inp"),
                       check = T, run = T, hashfilename = F)
```

```
## When hashfilename = FALSE, writeData cannot be 'ifmissing', setting to 'always'

## The following lines are not empty and do not end in a : or ;.
## 2: REIMS Mathematics Identity Model 1
## 4: FILE = "/Users/nathanalexander/Dropbox/Projects/immerse/reims/mplus/

## Rerun with parseMplus(add = TRUE) to add semicolons to all lines

## The file(s)
##  'reims1.dat'
## currently exist(s) and will be overwritten

## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

Take a look at the item probability plot:

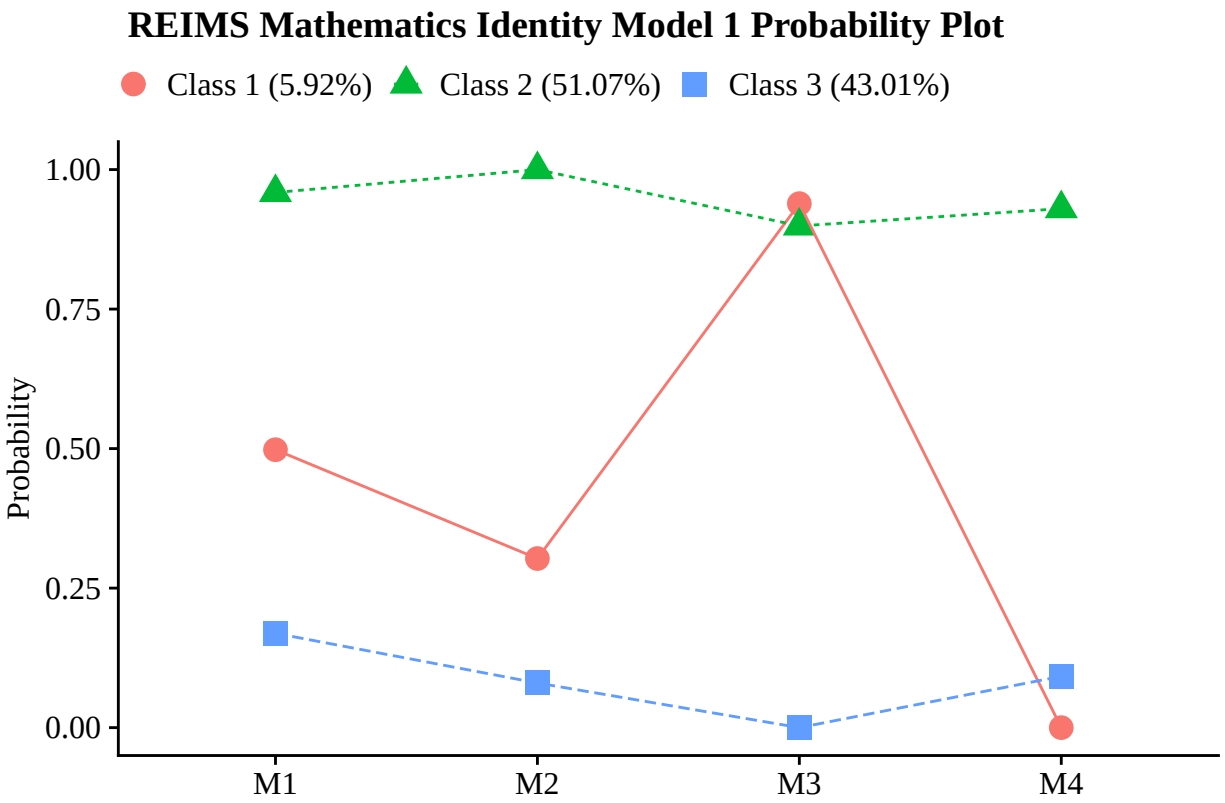
```
source(here("plot_lca.txt")) # custom function created by Dina to plot our lsa output
```

```
##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##     dcast, melt

## The following object is masked from 'package:tidyr':
##
##     smiths

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
model1 <- readModels(here("mplus", "reims1.out")) # read in output
```

```
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

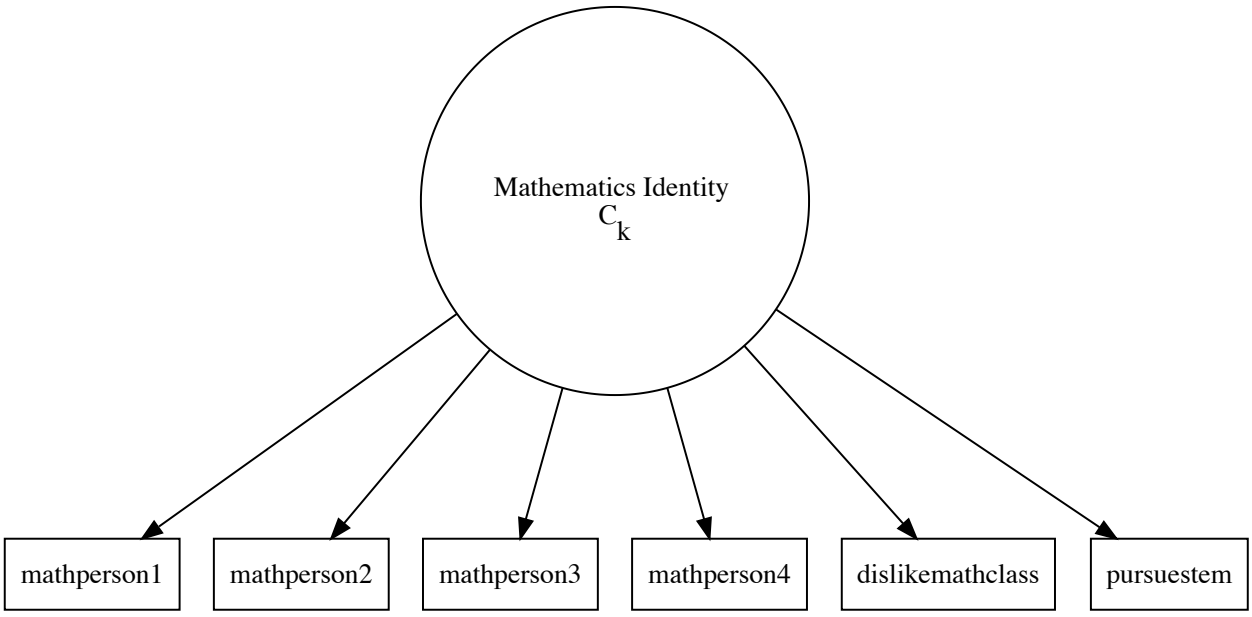
```
plot_lca(model_name = model1) # there is an error with the non atomic measure columns
```



Show probability plot of data and observe the different classes.

Model 2

Let's run a four class model and add three variables.



Model 2 MplusAutomation code

```
input <- mplusObject(
  TITLE = "REIMS Mathematics Identity Model 2",
  VARIABLE = "categorical = m1 m2 m3 m4 dislike pursue;
  usevar = m1-pursue;
  classes = c(4);",

  ANALYSIS =
    "estimator = mlr;
    type = mixture;",

  OUTPUT = "tech11 tech14;",

  PLOT = "type = plot3;
    series = m1-pursue(*);",

  usevariables = colnames(reims1),
  rdata = reims1)

output <- mplusModeler(input,
                       dataout = here("mplus", "reims1.dat"),
                       modelout = here("mplus", "reims1.inp"),
                       check = T, run = T, hashfilename = F)
```

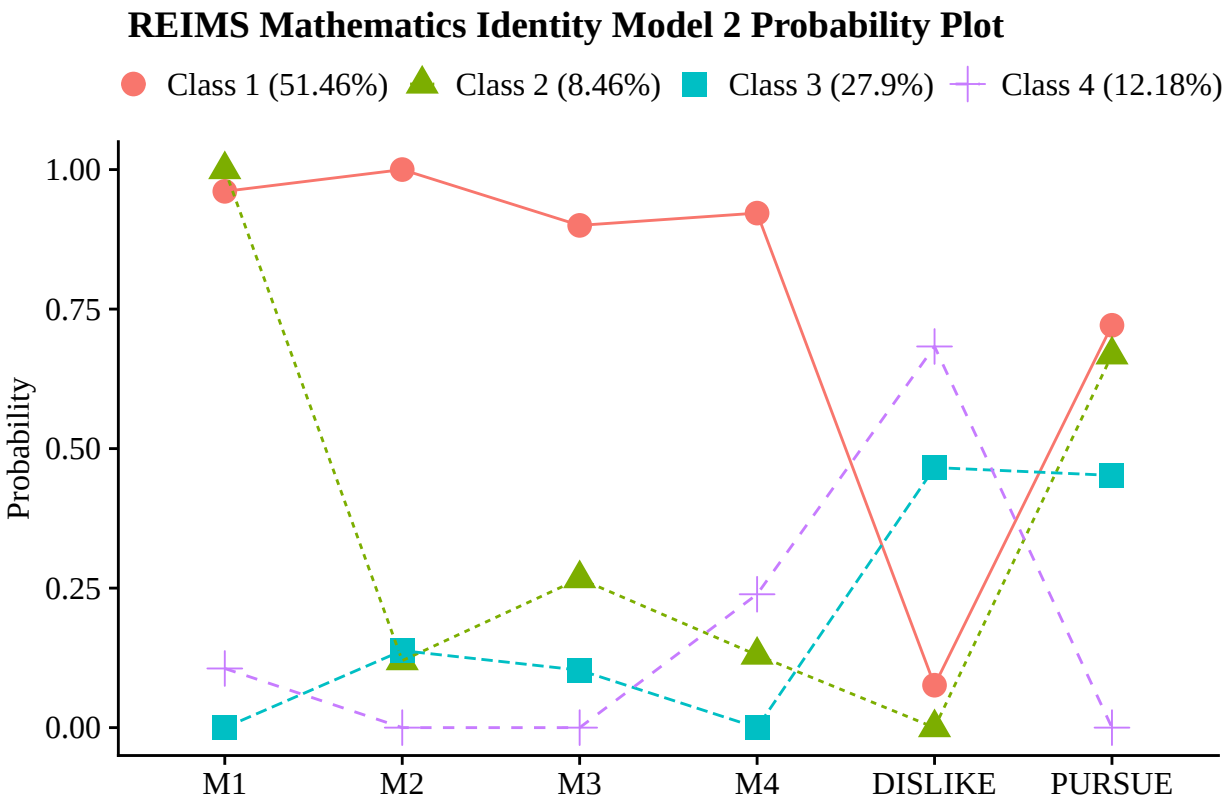
```
## When hashfilename = FALSE, writeData cannot be 'ifmissing', setting to 'always'

## The following lines are not empty and do not end in a : or ;.
## 2: REIMS Mathematics Identity Model 2
## 4: FILE = "/Users/nathanalexander/Dropbox/Projects/immerse/reims/mplus/

## Rerun with parseMplus(add = TRUE) to add semicolons to all lines

## The file(s)
##  'reims1.dat'
## currently exist(s) and will be overwritten

## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

Take a look at the item probability plot:

```
source(here("plot_lca.txt")) # custom function created by Dina to plot our lsa output

model2 <- readModels(here("mplus", "reims1.out")) # read in output
```

```
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

```
plot_lca(model_name = model2) # there is an error with the non atomic measure columns
```



Show probability plot of data and observe the different classes.

Enumeration

We use the mplusObject function in the MPlusAutomation package and saves all models run.

```
# add new libraries
library(cowplot)
library(glue)
```

Proportion of indicators using R:

```
# set up data to find proportions of binary indicators
df <- reims1 %>%
  pivot_longer(c(m1, m2, m3, m4, dislike, pursue),
  names_to = "Variable")

# create table of variables and counts
t1 <- table(df$Variable, df$value)

# find proportions and round to 3 decimal places
prop <- prop.table(t1, margin =1) %>%
  round(3)
```

```r
# combine everything to one table
dframe <- data.frame(Variables=rownames(t1), Proportion=prop[,2], Count=t1[,2])

# remove row names
row.names(dframe) <- NULL

# Make it a gt() table
prop_table <- dframe %>%
  gt()
prop_table
```

Variables	Proportion	Count
dislike	0.252	26
m1	0.592	61
m2	0.563	58
m3	0.515	53
m4	0.515	53
pursue	0.553	57

```r
# save as a word doc
gtsave(prop_table, here("figures", "prop_table.docx"))
```

Use an enumeration function

```r
lca_4 <- lapply(1:4, function(k) {
  lca_enum <- mplusObject(

    TITLE = glue("{k}-Class"),

    VARIABLE = glue(
      "categorical m1-pursue;
      usevar = m1-pursue;
      classes = c({k});"),

    ANALYSIS =
      "estimator = mlr;
      type = mixture;
      starts = 500 100;",

    OUTPUT = "tech11 tech14 svalues;",

    usevariables = colnames(reims1),
    rdata = reims1)

  lca_enum_fit <- mplusModeler(lca_enum,
                               dataout = glue(here("enum", "reims1.dat")),
                               modelout = glue(here("enum", "c{k}_reims1.inp")),
                               check = T, run = T, hashfilename = F)})
```

table of fit

We want to begin by extracting the data:

```
output_reims1 <- readModels(here("enum"))
```

```
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

```
enum_extract <- LatexSummaryTable(
  output_reims1,
  keepCols = c(
    "Title",
    "Parameters",
    "LL",
    "BIC",
    "aBIC",
    "BLRT_PValue",
    "T11_VLMR_PValue",
    "Observations"
  ),
  sortBy = "Title"
) # select first set of models (Class 1 through 4)

allFit <- enum_extract %>%
  mutate(CAIC = -2 * LL + Parameters * (log(Observations) + 1)) %>%
  mutate(AWE = -2 * LL + 2 * Parameters * (log(Observations) + 1.5)) %>%
  mutate(SIC = -.5 * BIC) %>%
  mutate(expSIC = exp(SIC - max(SIC))) %>%
  mutate(BF = exp(SIC - lead(SIC))) %>%
  mutate(cmPk = expSIC / sum(expSIC)) %>%
  dplyr::select(1:5, 9:10, 6:7, 13, 14) %>%
  arrange(Parameters)
```

Then we create a table:

```
fit_table <- allFit %>%
  gt() %>%
  tab_header(title = md("**Model Fit Summary Table**")) %>%
  cols_label(
    Title = "Classes",
    Parameters = md("Par"),
    LL = md("*LL*"),
    T11_VLMR_PValue = "VLMR",
    BLRT_PValue = "BLRT",
    BF = md("BF"),
    cmPk = md("*cmPk*")
  ) %>%
  tab_footnote(
    footnote = md(
      "*Note.* Par = Parameters; *LL* = model log likelihood;
BIC = Bayesian information criterion;
aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion;
AWE = approximate weight of evidence criterion;
```

```
BLRT = bootstrapped likelihood ratio test p-value;
VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value;
*cmPk* = approximate correct model probability."
    ),
locations = cells_title()
  ) %>%
  tab_options(column_labels.font.weight = "bold") %>%
  fmt_number(c(3:7),
             decimals = 2) %>%
  sub_missing(1:11,
              missing_text = "--") %>%
  fmt(
    c(8:9, 11),
    fns = function(x)
      ifelse(x < 0.001, "<.001",
             scales::number(x, accuracy = .01))
  ) %>%
  fmt(
    10,
    fns = function (x)
      ifelse(x > 100, ">100",
             scales::number(x, accuracy = .01))
  ) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold")

            ),
    locations = list(cells_body(
     columns = BIC,
     row = BIC == min(BIC[1:6]) # Change this to the number of classes you estimated
    ),
    cells_body(
     columns = aBIC,
     row = aBIC == min(aBIC[1:6])
    ),
    cells_body(
     columns = CAIC,
     row = CAIC == min(CAIC[1:6])
    ),
    cells_body(
     columns = AWE,
     row = AWE == min(AWE[1:6])
    ),
    cells_body(
     columns = cmPk,
     row =  cmPk == max(cmPk[1:6])
     ),
    cells_body(
     columns = BF,
     row =  BF > 10),
    cells_body(
     columns =  T11_VLMR_PValue,
```

```
      row =  ifelse(T11_VLMR_PValue < .001 & lead(T11_VLMR_PValue) > .05, T11_VLMR_PValue < .001, NA)),
    cells_body(
      columns =  BLRT_PValue,
      row =  ifelse(BLRT_PValue < .001 & lead(BLRT_PValue) > .05, BLRT_PValue < .001, NA))
  )
)

fit_table
```

Model Fit Summary Table[1]

Classes	Par	*LL*	BIC	aBIC	CAIC	AWE	BLRT	VLMR	BF	*cmPk*
1-Class	6	−411.90	851.62	832.66	857.61	897.42	–	–	0.00	<.001
2-Class	13	−294.07	648.40	607.34	661.40	747.65	**<.001**	**<.001**	**>100**	1.00
3-Class	20	−288.38	669.46	606.28	689.46	822.15	0.18	0.09	**>100**	<.001
4-Class	27	−286.09	697.31	612.03	724.31	903.45	1.00	0.08	–	<.001

[1] *Note.* Par = Parameters; *LL* = model log likelihood; BIC = Bayesian information criterion; aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; BLRT = bootstrapped likelihood ratio test p-value; VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value; *cmPk* = approximate correct model probability.

save the table:

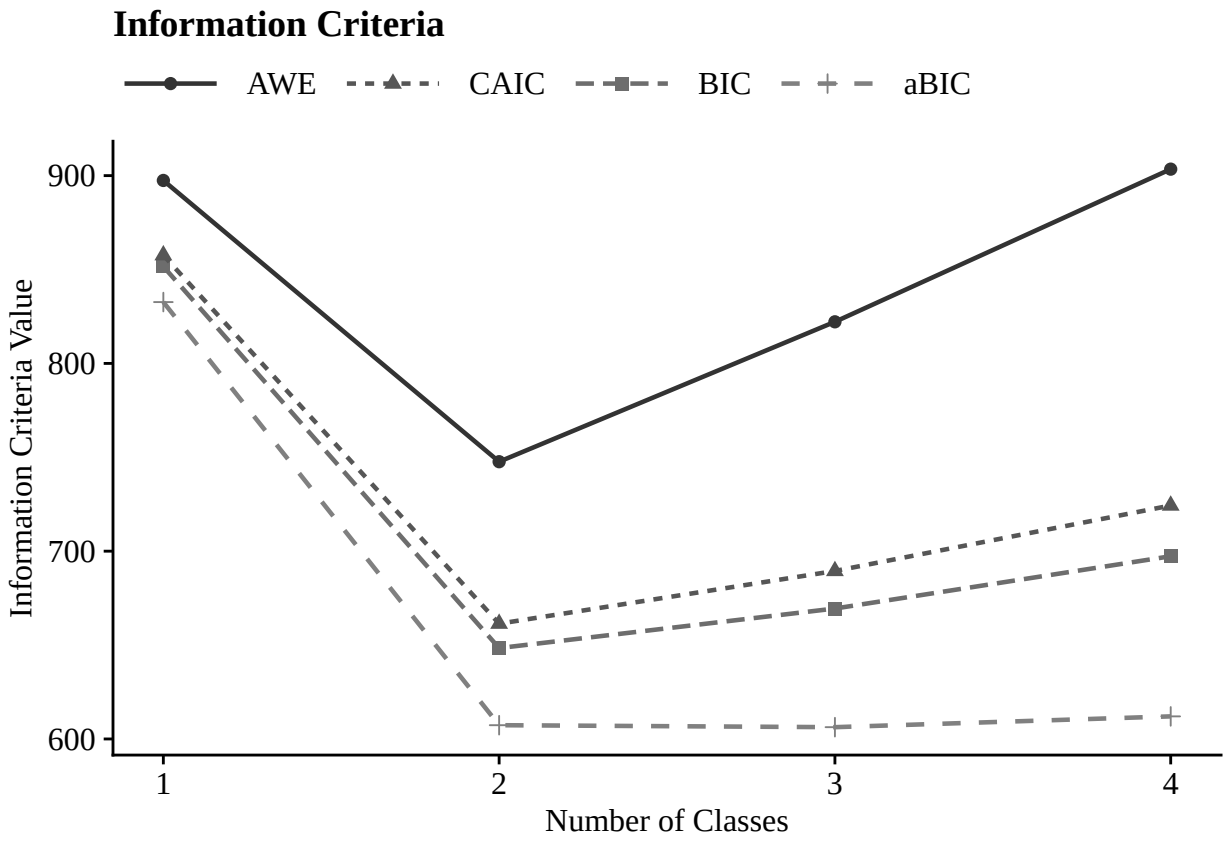
```
gtsave(fit_table, here("figures", "fit_table.png"))
```

Information Criterion Plot

```
allFit %>%
  dplyr::select(2:7) %>%
  rowid_to_column() %>%
  pivot_longer(`BIC`:`AWE`,
               names_to = "Index",
               values_to = "ic_value") %>%
  mutate(Index = factor(Index,
                        levels = c ("AWE", "CAIC", "BIC", "aBIC"))) %>%
  ggplot(aes(
    x = rowid,
    y = ic_value,
    color = Index,
    shape = Index,
    group = Index,
    lty = Index
  )) +
  geom_point(size = 2.0) + geom_line(size = .8) +
  scale_x_continuous(breaks = 1:nrow(allFit)) +
  scale_colour_grey(end = .5) +
  theme_cowplot() +
  labs(x = "Number of Classes", y = "Information Criteria Value", title = "Information Criteria") +
  theme(
```

```
    text = element_text(family = "serif", size = 12),
    legend.text = element_text(family="serif", size=12),
    legend.key.width = unit(3, "line"),
    legend.title = element_blank(),
    legend.position = "top"
  )
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



save the figure:

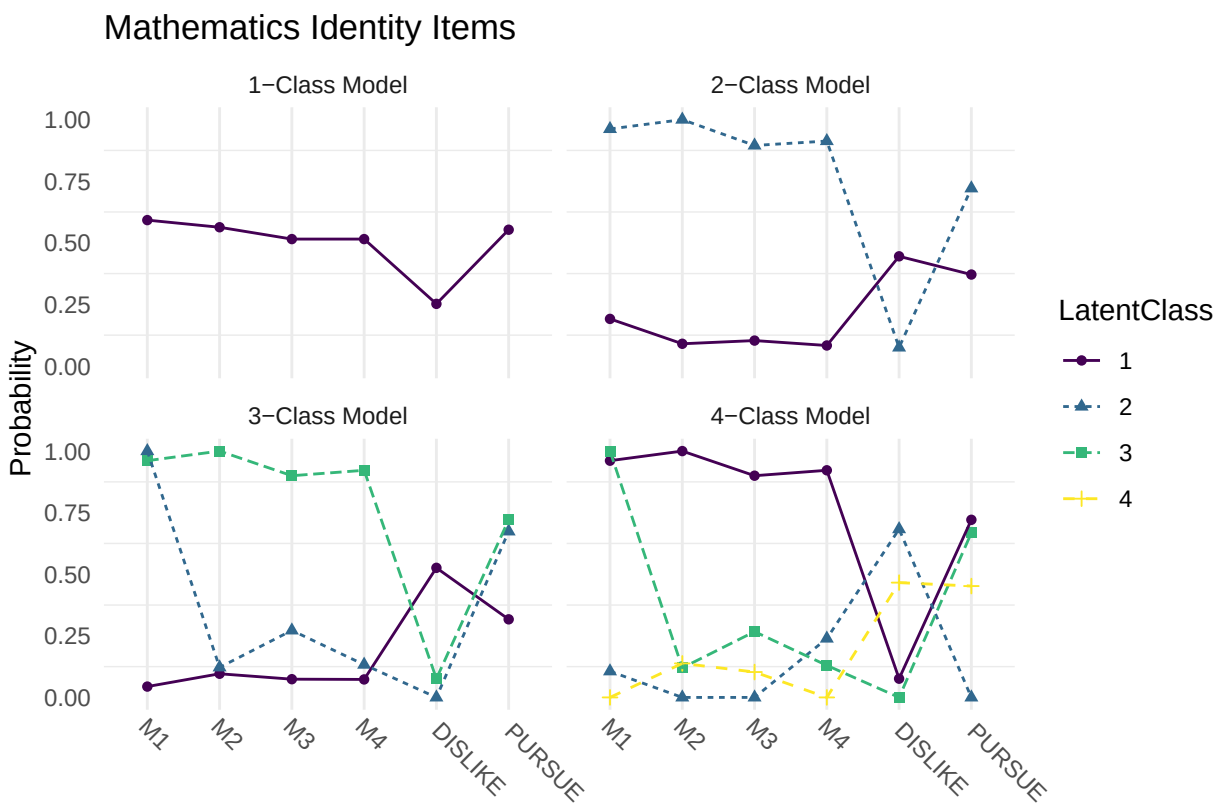
```
ggsave(here("figures", "info_criteria.png"), dpi=300, height=5, width=7, units="in")
```

Compare class solutions

Compare probability plots for $K = 1:4$ class solutions

```
model_results <- data.frame()

for (i in 1:length(output_reims1)) {
  temp <- output_reims1[[i]]$parameters$probability.scale %>%
```

```
    mutate(model = paste0(i, "-Class Model"))

  model_results <- rbind(model_results, temp)
}

compare_plot <-
  model_results %>%
  filter(category ==2) %>%
  dplyr::select(est, model, LatentClass, param)

compare_plot$param <- fct_inorder(compare_plot$param)

ggplot(
  compare_plot,
  aes(
    x=param,
    y=est,
    color = LatentClass,
    shape = LatentClass,
    group = LatentClass,
    lty = LatentClass
  )
) +
  geom_point() +
  geom_line() +
  scale_color_viridis_d() +
  facet_wrap(~ model, ncol = 2) +
  labs(title = "Mathematics Identity Items", x = " ", y = "Probability") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank(),
        axis.text.x = element_text(angle = -45, hjust = -.1))
```



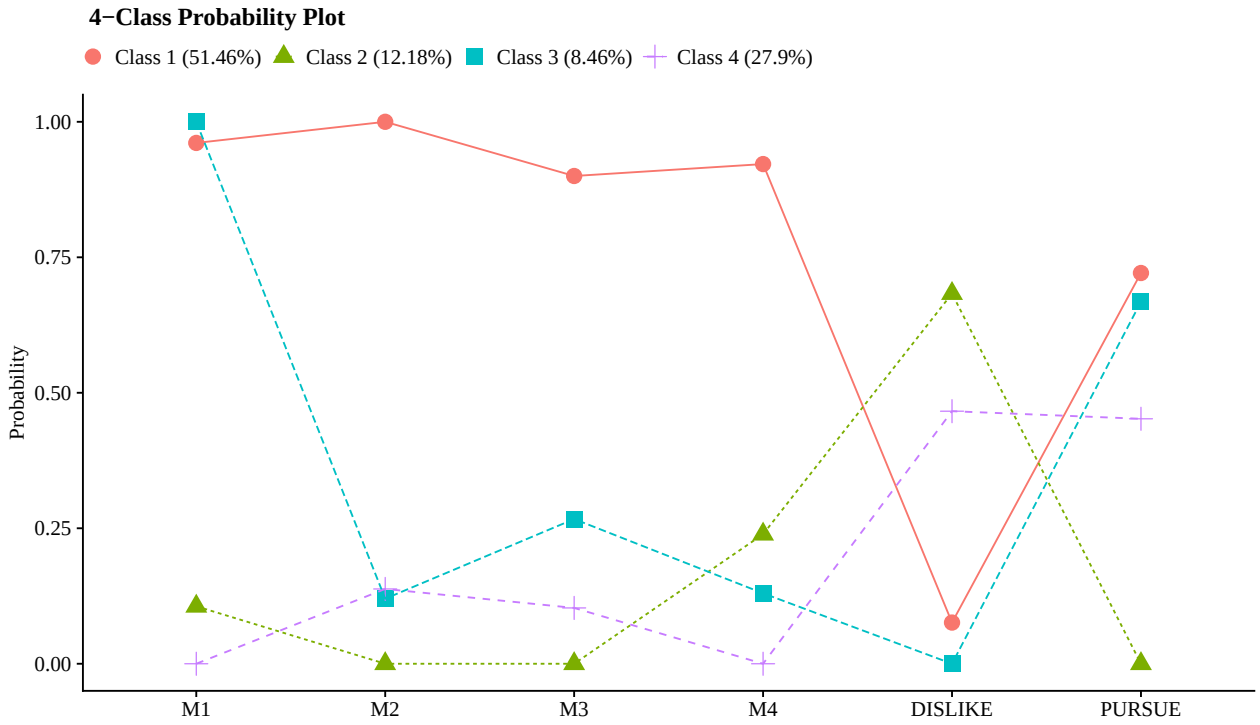
save the figure:

```
ggsave(here("figures", "compare_kclass_plot.png"), dpi=300, height=5, width=7, units="in")
```

4-Class Probability Plot

Use the `plot_lca` function provided in the folder to plot the item probability plot. This function requires one argument: - `model_name`: The name of the Mplus `readModels` object (e.g., `output_lsal$c4_lsal.out`) - this was updated for reims.

```
source("plot_lca.txt")

plot_lca(model_name = output_reims1$c4_reims1.out)
```

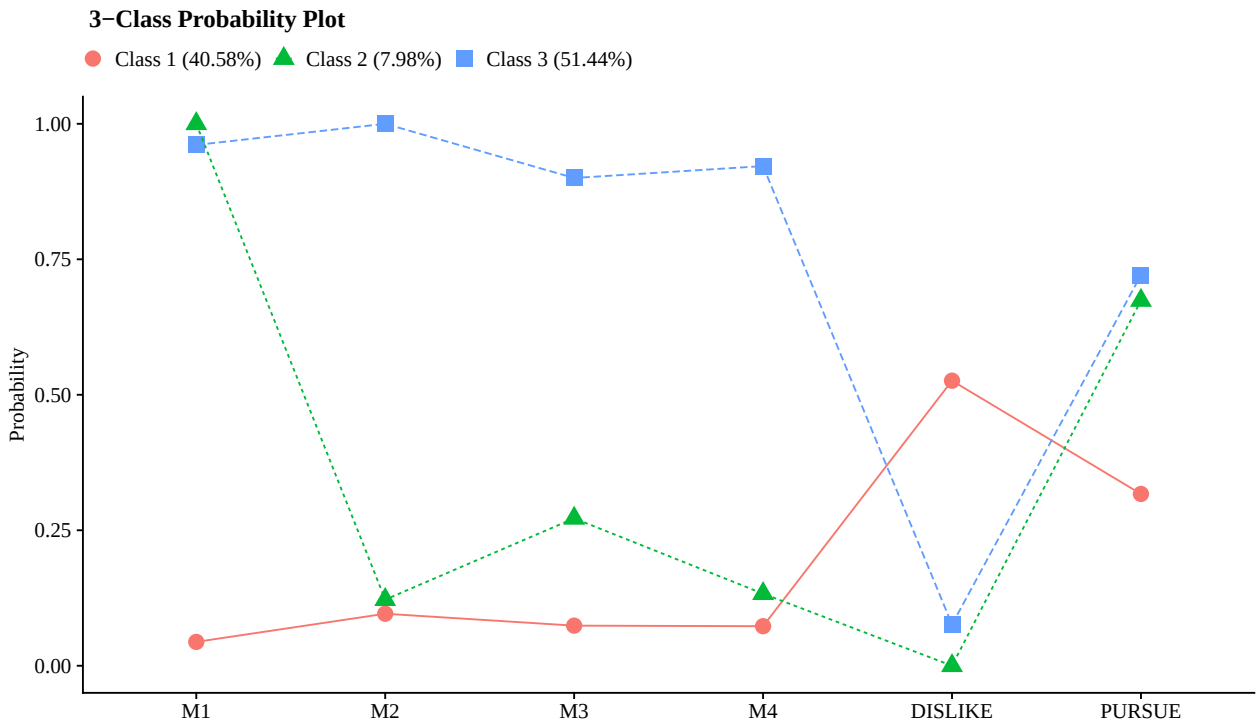


save the figure:

```
ggsave(here("figures", "probability_plot_4class.png"), dpi="retina", height=5, width=7, units="in")
```

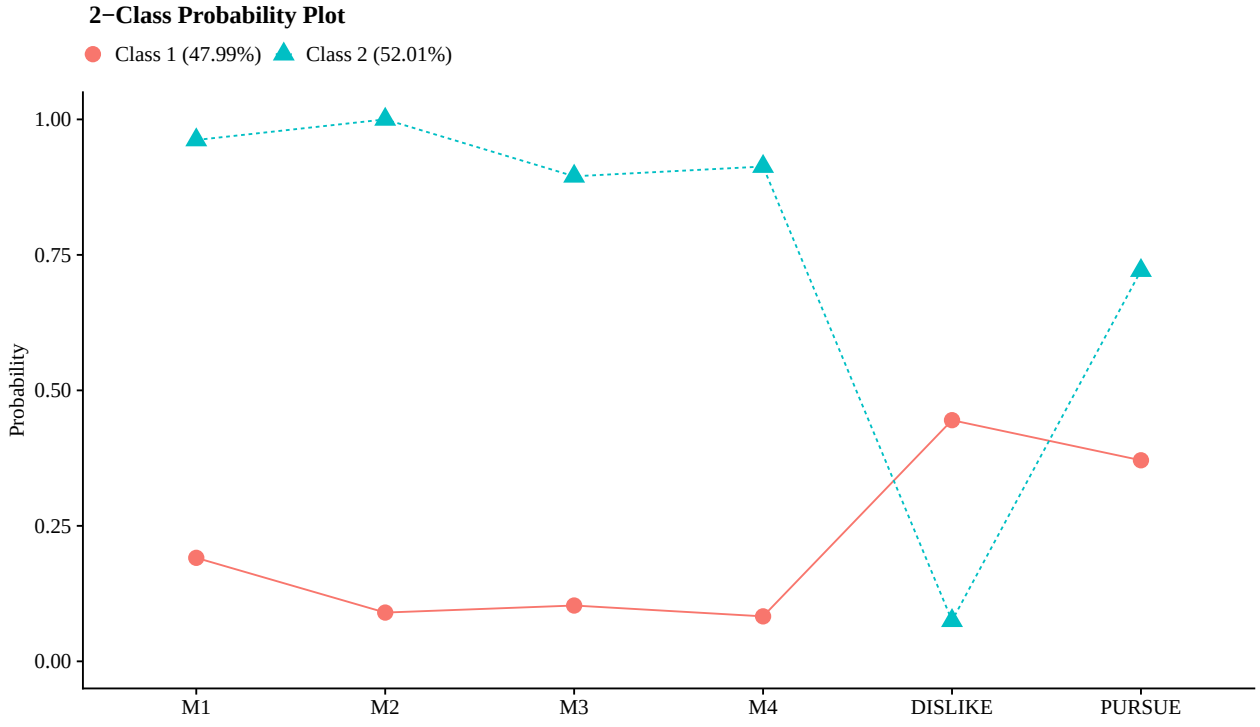
3-Class Probability Plot

```
source("plot_lca.txt")

plot_lca(model_name = output_reims1$c3_reims1.out)
```



2-Class Probability Plot

```
source("plot_lca.txt")

plot_lca(model_name = output_reims1$c2_reims1.out)
```



save the figure:

```
ggsave(here("figures", "probability_plot_2class.png"), dpi="retina", height=5, width=7, units="in")
```

Observed response patterns

Save response frequencies for the 2-class model with `response is _____.dat` under `SAVEDATA`.

```
patterns  <- mplusObject(

  TITLE = "LCA - Save response patterns",

  VARIABLE =
  "categorical = m1-pursue;
   usevar =  m1-pursue;
   classes = c(4);",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    starts = 0;
    processors = 10;
    optseed = 830529;",

  SAVEDATA =
   "File=savedata.dat;
    Save=cprob;

    ! Code to save response frequency data

    response is resp_patterns.dat;",

  OUTPUT = "patterns tech10 tech11 tech14",

  usevariables = colnames(reims1),
  rdata = reims1)

patterns_fit <- mplusModeler(patterns,
                dataout=here("mplus", "patterns.dat"),
                modelout=here("mplus", "patterns.inp") ,
                check=TRUE, run = TRUE, hashfilename = FALSE)
```

read in observed response pattern data and relabel the columns:

```
# read in response frequency data that we just created:
patterns <- read_table(here("mplus", "resp_patterns.dat"),
                       col_names=FALSE, na = "*")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_double(),
```

```
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double(),
##   X7 = col_double(),
##   X8 = col_double(),
##   X9 = col_double(),
##   X10 = col_double(),
##   X11 = col_double(),
##   X12 = col_double()
## )
```

```
# extract the column names
names <- names(readModels(here("mplus", "patterns.out"))[['savedata']])
```

```
## Warning in (function (..., deparse.level = 1) : number of columns of result is
## not a multiple of vector length (arg 1)

## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

```
# add the names back to the dataset
colnames(patterns) <- c("Frequency", names)
```

we then create a table with the response patterns, then top of conditional response pattern for each modal class assignment

```
# Order responses by highest frequency
order_highest <- patterns %>%
  arrange(desc(Frequency))

# Loop `patterns` data to list top 5 conditional response patterns for each class
loop_cond  <- lapply(1:max(patterns$C), function(k) {
order_cond <- patterns %>%
  filter(C == k) %>%
  arrange(desc(Frequency)) %>%
  head(5)
  })

# Convert loop into data frame
table_data <- as.data.frame(bind_rows(loop_cond))

# Combine unconditional and conditional responses patterns
response_patterns <-  rbind(order_highest[1:5,], table_data)
```

we then use `{gt}` to make a nicely formatted table.

```r
resp_table <- response_patterns %>%
  gt() %>%
    tab_header(
    title = "Observed Response Patterns",
    subtitle = html("Response patterns, estimated frequencies, estimated posterior class probabilities a
    tab_source_note(
    source_note = md("Data Source: **Racial Ethnic Identity in Mathematics Survey (REIMS)**")) %>%
    cols_label(
      Frequency = html("<i>f</i><sub>r</sub>"),
    M1 = "Math Person 1",
    M2 = "Math Person 2",
    M3 = "Math Person 3",
    M4 = "Math Person 4",
    DISLIKE = "Dislike Math",
    PURSUE = "Pursue Math",
    CPROB1 = html("P<sub><i>k</i></sub>=1"),
    CPROB2 = html("P<sub><i>k</i></sub>=2"),
    CPROB3 = html("P<sub><i>k</i></sub>=3"),
    CPROB4 = html("P<sub><i>k</i></sub>=4"),    # Change based on number of classes
    C = md("*k*")) %>%
  tab_row_group(
    label = "Unconditional response patterns",
    rows = 1:5) %>%
  tab_row_group(
    label = md("*k* = 1 Conditional response patterns"),
    rows = 6) %>% #EDIT THESE VALUES BASED ON THE LAST COLUMN
  tab_row_group(
    label = md("*k* = 2 Conditional response patterns"),
    rows = 7:11)  %>% #EDIT THESE VALUES BASED ON THE LAST COLUMN
  tab_row_group(
    label = md("*k* = 3 Conditional response patterns"),
    rows = 12:16)  %>% #EDIT THESE VALUES BASED ON THE LAST COLUMN
  tab_row_group(
    label = md("*k* = 4 Conditional response patterns"),
    rows = 17:21)  %>% #EDIT THESE VALUES BASED ON THE LAST COLUMN
    row_group_order(
      groups = c("Unconditional response patterns",
                 md("*k* = 1 Conditional response patterns"),
                 md("*k* = 2 Conditional response patterns"),
                 md("*k* = 3 Conditional response patterns"),
                 md("*k* = 4 Conditional response patterns"))) %>%
    tab_footnote(
    footnote = html(
      "<i>Note.</i> <i>f</i><sub>r</sub> = response pattern frequency; P<sub><i>k</i></sub> = posterior
    )
  ) %>%
  cols_align(align = "center") %>%
  opt_align_table_header(align = "left") %>%
  gt::tab_options(table.font.names = "Times New Roman")

resp_table
```

Observed Response Patterns

Response patterns, estimated frequencies, estima

f_r	Math Person 1	Math Person 2	Math Person 3	Math Person 4	Dislike Math	Pursue
Unconditional response patterns						
30	1	1	1	1	0	1
11	0	0	0	0	1	0
9	1	1	1	1	0	0
9	0	0	0	0	0	0
6	0	0	0	0	1	1
$k = 1$ Conditional response patterns						
1	0	1	1	0	1	0
$k = 2$ Conditional response patterns						
3	1	0	0	0	0	1
2	1	1	0	1	0	1
2	1	1	0	1	0	0
2	1	1	1	0	0	0
1	1	1	0	1	1	1
$k = 3$ Conditional response patterns						
30	1	1	1	1	0	1
9	1	1	1	1	0	0
2	1	1	1	1	1	1
2	1	1	1	0	0	1
1	0	1	1	1	0	0
$k = 4$ Conditional response patterns						
11	0	0	0	0	1	0
9	0	0	0	0	0	0
6	0	0	0	0	1	1
5	0	0	0	0	0	1
2	0	0	0	1	1	0

Note. f_r = response pattern frequency; P_k = posterior class probabilities
Data Source: **Racial Ethnic Identity in Mathematics Survey (REIMS)**

save the table

```
gtsave(resp_table, here("figures", "resp_table.png"))
```

Classification Diagnostics

We will use Mplus to calculate k-class confidence intervals, using the 4-class model.

```
classification <- mplusObject(

  TITLE = "LCA - Calculate k-Class 95% CI",

  VARIABLE =
    "categorical = m1-pursue;
```

```
    usevar = m1-pursue;
    classes = c(4);",

  ANALYSIS =
    "estimator = ml;
    type = mixture;
    starts = 0;
    processors = 10;
    optseed = 945065;
    bootstrap = 100;",

  MODEL =
    "
  !CHANGE THIS SECTION TO YOUR CHOSEN k-CLASS MODEL

  %OVERALL%
  [C#1](c1);
  [C#2](c2);
  [C#3](c3);


  Model Constraint:
  New(p1 p2 p3 p4);

  p1 = exp(c1)/(1+exp(c1)+exp(c2)+exp(c3));
  p2 = exp(c2)/(1+exp(c1)+exp(c2)+exp(c3));
  p3 = exp(c3)/(1+exp(c1)+exp(c2)+exp(c3));
  p4 = 1/(1+exp(c1)+exp(c2)+exp(c3));",

  OUTPUT = "cinterval(bcbootstrap)",

  usevariables = colnames(reims1),
  rdata = reims1)

classification_fit <- mplusModeler(classification,
                dataout=here("mplus", "reims-1.dat"),
                modelout=here("mplus", "class.inp") ,
                check=TRUE, run = TRUE, hashfilename = FALSE)
```

Note from IMMERSE team: Ensure that the classes did not shift during this step (i.e., Class 1 in the enumeration run is now Class 4). Evaluate output and compare the class counts and proportions for the latent classes. Using the OPTSEED function ensures replication of the best loglikelihood value run.

Read in the 4-class model:

```
# Read in the 4-class model and extract information needed
class_output <- readModels(here("mplus", "class.out"))
```

```
## Warning in (function (..., deparse.level = 1) : number of columns of result is
## not a multiple of vector length (arg 1)
```

```
## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## Warning in extractSampstat(outfiletext, curfile): NAs introduced by coercion

## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

```
# Entropy
entropy <- c(class_output$summaries$Entropy, rep(NA, class_output$summaries$NLatentClasses-1))

# 95% k-Class and k-class 95% Confidence Intervals
k_ci <- class_output$parameters$ci.unstandardized %>%
  filter(paramHeader == "New.Additional.Parameters") %>%
  unite(CI, c(low2.5,up2.5), sep=", ", remove = TRUE) %>%
  mutate(CI = paste0("[", CI, "]")) %>%
  rename(kclass=est) %>%
  dplyr::select(kclass, CI)

# AvePPk = Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent C
avePPk <- tibble(avePPk = diag(class_output$class_counts$avgProbs.mostLikely))

# mcaPk = modal class assignment proportion
mcaPk <- round(class_output$class_counts$mostLikely,3) %>%
  mutate(model = paste0("Class ", class)) %>%
  add_column(avePPk, k_ci) %>%
  rename(mcaPk = proportion) %>%
  dplyr::select(model, kclass, CI, mcaPk, avePPk)

# OCCk = odds of correct classification
OCCk <- mcaPk %>%
  mutate(OCCk = round((avePPk/(1-avePPk))/(kclass/(1-kclass)),3))

# Put everything together
class_df <- data.frame(OCCk, entropy)
```

now we use {gt} to make a nicely formatted table

```
class_table <- class_df %>%
  gt() %>%
    tab_header(
    title = "Model Classification Diagnostics for the 4-Class Solution") %>%
    cols_label(
      model = md("*k*-Class"),
      kclass = md("*k*-Class Proportions"),
      CI = "95% CI",
      mcaPk = md("*mcaPk*"),
      avePPk = md("*AvePPk*"),
      OCCk = md("*OCCk*"),
      entropy = "Entropy") %>%
    sub_missing(7,
                missing_text = "") %>%
  cols_align(align = "center") %>%
```

```
  opt_align_table_header(align = "left") %>%
  gt::tab_options(table.font.names = "Times New Roman")

class_table
```

Model Classification Diagnostics for the 4-Class Solution

k-Class	*k*-Class Proportions	95% CI	*mcaPk*	*AvePPk*	*OCCk*	Entropy
Class 1	0.51	[0.395, 0.591]	0.515	0.986	67.667	0.894
Class 2	0.30	[0.175, 0.399]	0.301	0.897	20.320	
Class 3	0.01	[0, 0.029]	0.010	0.997	32901.000	
Class 4	0.18	[0.08, 0.299]	0.175	0.814	19.937	

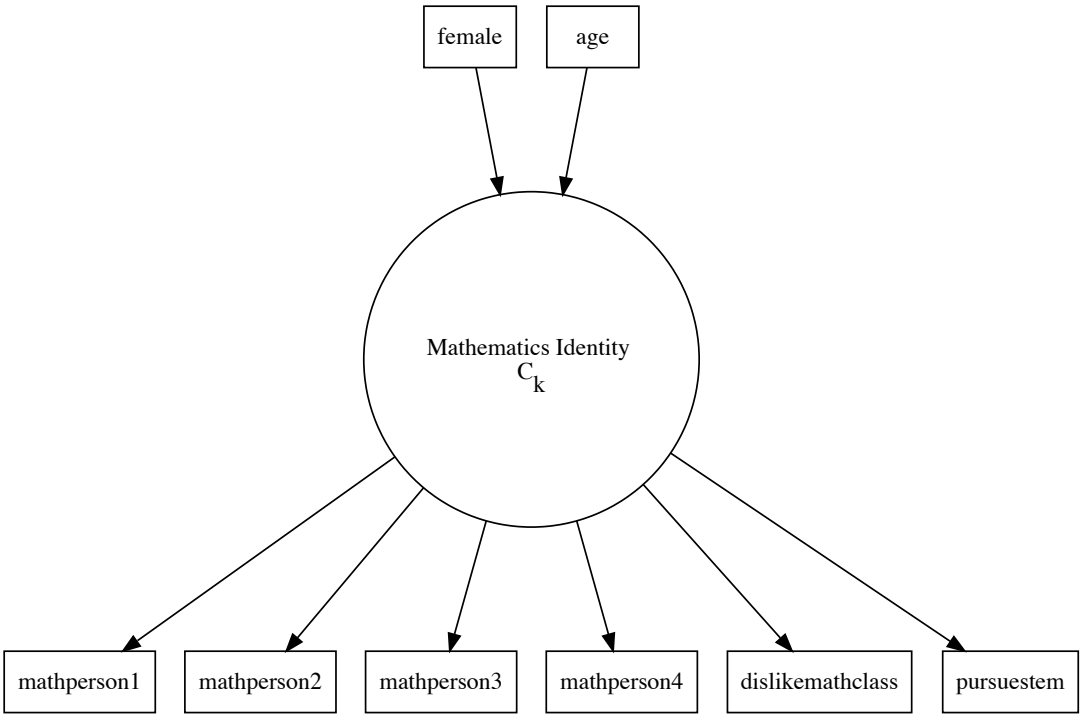
save the table:

```
gtsave(class_table, here("figures", "class_table.png"))
```

Auxillary variables

Model 3

We will now add auxiliary variables to our model and name it model 3.



Manual 3-step latent class regression with 3 covariates

Integrate covariates and with the mixture model

```
# indicators and variables for full model build

tribble(
  ~"Name", ~"Variable Description",
  #----------|----------|,
  "mathperson1","I am a math person.",
  "mathperson2","My peers think I am a math person.",
  "mathperson3","My teachers think I am a math person.",
  "mathperson4","There are a lot of people who think I am a math person.",
  "dislikemathclass","In general, I dislike math classes.",
  "pursuestem","If given the chance, I will pursue a STEM career.",
  "female","Self-reported sex (0=male, 1=female)",
  "age", "Self-reporeted age") %>%
  gt() %>%
  tab_header(title = md("**LCA Indicators & Auxilary Variables: Mathematics Identity Example**"), subti
  tab_row_group(group = "", rows = 1:6) %>%
  tab_row_group(group = "Auxilary Variables", rows = 7:8) %>%
  row_group_order(groups = c("", "Auxilary Variables")) %>%
  tab_options(column_labels.font.weight = "bold",
              row_group.font.weight = "bold")
```

```
## Warning: Since gt v0.3.0 the `group` argument has been deprecated.
## * Use the `label` argument to specify the group label.
## This warning is displayed once every 8 hours.
```

LCA Indicators & Auxilary Variables: Mathematics Identity Example

Name	Variable Description
mathperson1	I am a math person.
mathperson2	My peers think I am a math person.
mathperson3	My teachers think I am a math person.
mathperson4	There are a lot of people who think I am a math person.
dislikemathclass	In general, I dislike math classes.
pursuestem	If given the chance, I will pursue a STEM career.
Auxilary Variables	
female	Self-reported sex (0=male, 1=female)
age	Self-reporeted age

```
# add female variable from original data set
# the variables were not included in the subset data
reims$sex
```

```
##   [1] 2 2 1 1 1 2 1 2 2 2 2 1 1 2 1 1 2 2 1 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 3 2
##  [38] 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 2 2 1 1 2 1 1 2 1 2 2 2 2 2 1 2
##  [75] 2 2 1 2 1 1 2 1 2 2 2 1 2 2 2 2 2 1 2 1 2 1 1 2 2 1 2 1 1
```

```
reims1$female <- reims$sex
reims1$age <- reims$age
```

Step 1: Class enumeration with auxilary specification

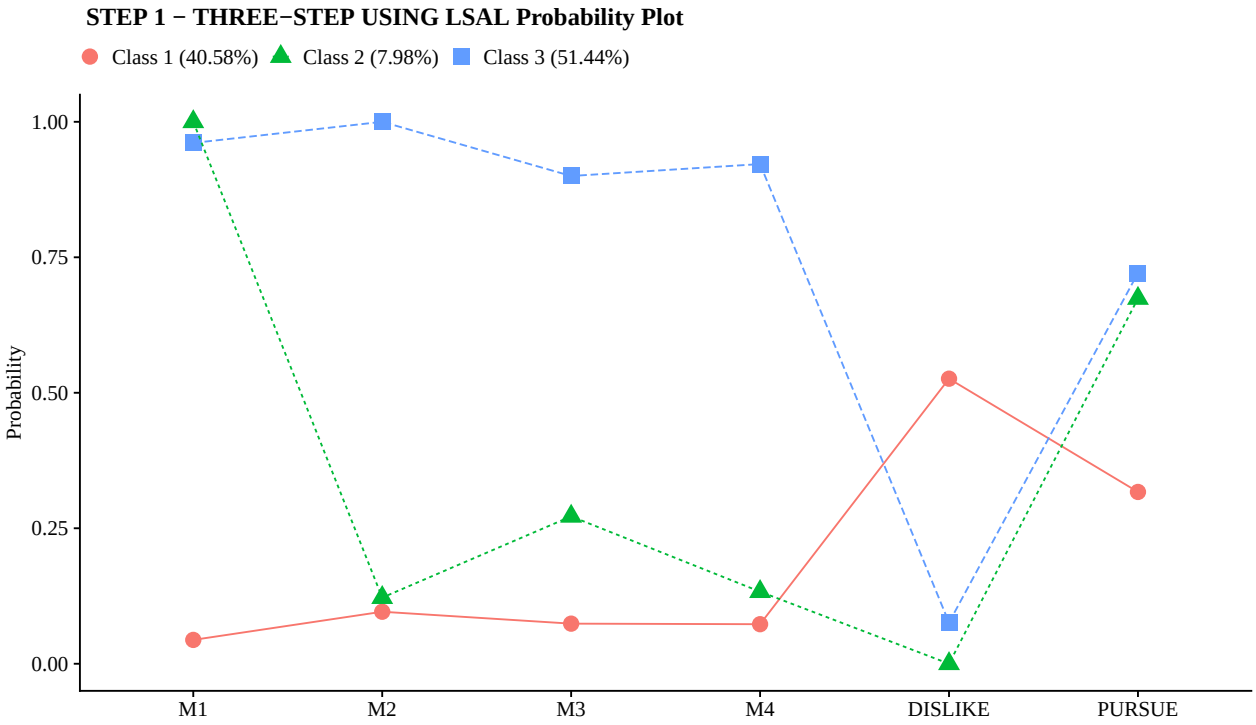
```
step1 <- mplusObject(
  TITLE = "STEP 1 - THREE-STEP USING LSAL",
  VARIABLE =
    "categorical = m1-pursue;
    usevar = m1-pursue;
    classes = c(3);

    auxiliary = ! list all potential covariates and distals here

    female  ! covariate",

  ANALYSIS =
    "estimator = mlr;
    type = mixture;
    starts = 500 100;",

  SAVEDATA =
    "File=3step_savedata.dat;
    Save=cprob;",

  OUTPUT = "residual tech11 tech14",

  PLOT =
    "type = plot3;
    series = m1-pursue(*);",

  usevariables = colnames(reims1),
  rdata = reims1)

step1_fit <- mplusModeler(step1,
                          dataout=here("manual", "Step1.dat"),
                          modelout=here("manual","one.inp"),
                          check=T, run=T, hashfilename = F)
```

save the plot

```
source("plot_lca.txt")

output_reims1 <- readModels(here("manual","one.out"))
```

```
## No PROPORTION OF DATA PRESENT sections found within COVARIANCE COVERAGE OF DATA output.
```

```
plot_lca(model_name = output_reims1)
```



Step 2: Determine measurement error

we want to extract the logits for the classification probabilities for the most likely latent class.

```
logic_cprobs <- as.data.frame(output_reims1[["class_counts"]]

                              [["logitProbs.mostLikely"]])
```

we now extract the saved data set.

```
savedata <- as.data.frame(output_reims1[["savedata"]])
```

we then rename the column in savedata named “c” and change to “N”

```
colnames(savedata)[colnames(savedata)=="C"] <- "N"
```

Step 3: Add auxiliary variables to analysis

Model 1 with female as covariate (to see if there are differences by sex). There are no distal outcomes included in this sample model.

I will model the auxilary variable with 3 classes first.

```
step3 <- mplusObject(

  TITLE = "STEP 3 - 3STEP LSAY",

  VARIABLE =
```

```
  "nominal=N;
  usevar = n;
  classes = c(3);

  usevar = female;",

  ANALYSIS =
  "estimator = mlr;
  type = mixture;
  starts = 0;",

  MODEL =
    glue(
      " %OVERALL%

      C on female (f1-f2); ! covariate as predictor of C

      %C#1%
      [n#1@{logit_cprobs[1,1]}];
      [n#2@{logit_cprobs[1,2]}];

      %C#2%
      [n#1@{logit_cprobs[2,1]}];
      [n#2@{logit_cprobs[2,2]}];

      %C#3%
      [n#1@{logit_cprobs[3,1]}];
      [n#2@{logit_cprobs[3,2]}];",

      MODELTEST = "
      m1=m2;
      m2=m3;
      !f1=f2;
      !f2=f3;
      ",

    MODELCONSTRAINT =
      "New (diff12 diff13 diff23);

      diff12 = m1-m2; ! test pairwise distal mean differences
      diff13 = m1-m3;
      diff23 = m2-m3;",

    usevariables = colnames(savedata),
    rdata = savedata)


step3_fit <- mplusModeler(step3,
                dataout=here("manual", "Step3.dat"),
                modelout=here("manual", "three.inp"),
                check=TRUE, run = TRUE, hashfilename = FALSE)
```

This file is based on resources provided by the Institute of Mixture Modeling for Equity-Oriented Researchers, Scholars, and Educators (2024). IMMERSE In-Person Training Workshop (IES No. 305B220021). Institute of Education Sciences. https://immerse-ucsb.github.io/pre-training