Appendix C.IV: Replication Method - Confirmatory Latent Class Analysis (CLCA)

Adam Garber

Updated: October 2, 2024

```
library(tidyverse)
library(MplusAutomation)
library(rhdf5)
library(here)
library(glue)
library(gt)
library(janitor)
library(reshape2)
library(cowplot)
```

Read in data files for the initial sample and replication sample:

- Initial sample: `n_3000_lca_rep2.dat` (N=3000; Replication 2)
- Replication sample: `n_3000_lca_rep3.dat` (N=3000; Replication 3)

```
# Read in initial sample
sample_1 <- read.delim2(here("C1-Simulation", "n_3000_lca_rep2.dat"), sep = "", header = FALSE) %>%
    select(-V18) %>%
    setNames(c("primary", "change", "interupt", "initiat", "engage", "approach",
        "response", "expect", "new", "same", "relative", "objects", "sequence", "trans",
        "avoid", "control", "touch")) %>%
    purrr::modify_if(is.character, as.numeric) %>%
    mutate(sample = 1)

sample_2 <- read.delim2(here("C1-Simulation", "n_3000_lca_rep3.dat"), sep = "", header = FALSE) %>%
    select(-V18) %>%
    setNames(c("primary", "change", "interupt", "initiat", "engage", "approach",
        "response", "expect", "new", "same", "relative", "objects", "sequence", "trans",
        "avoid", "control", "touch")) %>%
    purrr::modify_if(is.character, as.numeric) %>%
    mutate(sample = 2)
```

Combine samples 1 and 2 for multigroup analysis

```
combined_data <- rbind(sample_1, sample_2)
```

Save the `combined_data`

```
write_csv(combined_data, here("data", "C4_Combined_Data.csv"))
```

Step 1

Produce start values from the initial sample (`sample_1`) to use in Step 2

```
m_start  <- mplusObject(

  TITLE = "Starts - Sample 1",

  VARIABLE =
   "categorical = primary-touch;
    usevar = primary-touch;
    classes = c(5); ",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    starts = 500 200;
    !STSEED = 887676; !!! USE SEED TO REPLICATE THESIS RESULTS !!!
   ",

  OUTPUT = "svalues;",

  PLOT = "type = plot3;
          series = primary-touch(*);",

  usevariables = colnames(sample_1),
  rdata = sample_1)

m_start_fit <- mplusModeler(m_start,
                 dataout=here("C4-Replication", "Starts_Sample_1.dat"),
                 modelout=here("C4-Replication", "Starts_Sample_1.inp") ,
                 check=TRUE, run = TRUE, hashfilename = FALSE)
```

Step 2

Estimate baseline multigroup model with thresholds unconstrained across groups

```
m_step2  <- mplusObject(

  TITLE = "Step 2 - Replication",
```

```
VARIABLE =
 "categorical = primary-touch;
  usevar = primary-touch;

  knownclass = CG (sample=1 sample=2); !!! LABEL SAMPLES !!!
  classes = CG(2) C(5);                !!! CG is KNOWN CLASS !!!
 ",

ANALYSIS =
 "estimator = mlr;
  type = mixture;
  starts = 500 200;
  !STSEED = ; !!! USE SEED TO REPLICATE THESIS RESULTS !!!
 ",

MODEL =
  "%OVERALL%

   !!! INITIAL SAMPLE (SAMPLE 1) !!! INITIAL SAMPLE (SAMPLE 1) !!!

   %CG#1.C#1%

   [ primary$1*-0.21760  ];
   [ change$1*-0.45914   ];
   [ interupt$1*-0.64034 ];
   [ initiat$1*-0.46809  ];
   [ engage$1*-0.33510   ];
   [ approach$1*-0.42163 ];
   [ response$1*-0.58459 ];
   [ expect$1*-0.60368   ];
   [ new$1*1.17714       ];
   [ same$1*2.92160      ];
   [ relative$1*1.86782  ];
   [ objects$1*-1.97843  ];
   [ sequence$1*-1.40431 ];
   [ trans$1*-1.31741    ];
   [ avoid$1*0.99669     ];
   [ control$1*1.69313   ];
   [ touch$1*1.24603     ];

   %CG#1.C#2%

   [ primary$1*3.45112   ];
   [ change$1*2.10148    ];
   [ interupt$1*2.33173  ];
   [ initiat$1*4.67053   ];
   [ engage$1*2.61317    ];
   [ approach$1*3.00689  ];
   [ response$1*3.01867  ];
   [ expect$1*2.47608    ];
   [ new$1*2.13464       ];
   [ same$1*2.61928      ];
   [ relative$1*2.46710  ];
```

```
    [ objects$1*3.49010   ];
    [ sequence$1*2.16238  ];
    [ trans$1*1.76331     ];
    [ avoid$1*1.62916     ];
    [ control$1*2.77768   ];
    [ touch$1*2.98188     ];

    %CG#1.C#3%

    [ primary$1*-1.15805  ];
    [ change$1*-1.05640   ];
    [ interupt$1*-1.47052 ];
    [ initiat$1*-1.88494  ];
    [ engage$1*-0.85925   ];
    [ approach$1*-1.29676 ];
    [ response$1*-1.60551 ];
    [ expect$1*-1.84879   ];
    [ new$1*1.21384       ];
    [ same$1*1.25810      ];
    [ relative$1*1.00299  ];
    [ objects$1*2.02420   ];
    [ sequence$1*2.49353  ];
    [ trans$1*1.91476     ];
    [ avoid$1*1.22229     ];
    [ control$1*3.67528   ];
    [ touch$1*2.02070     ];

    %CG#1.C#4%

    [ primary$1*-2.45913  ];
    [ change$1*-2.07089   ];
    [ interupt$1*-2.29483 ];
    [ initiat$1*-2.28241  ];
    [ engage$1*-1.87192   ];
    [ approach$1*-2.25327 ];
    [ response$1*-2.27708 ];
    [ expect$1*-2.29726   ];
    [ new$1*-2.86563      ];
    [ same$1*-2.02751     ];
    [ relative$1*-2.09303 ];
    [ objects$1*-3.25774  ];
    [ sequence$1*-2.18406 ];
    [ trans$1*-2.90318    ];
    [ avoid$1*-2.16937    ];
    [ control$1*-1.95077  ];
    [ touch$1*-1.64226    ];

    %CG#1.C#5%

    [ primary$1*3.92177   ];
    [ change$1*1.18687    ];
    [ interupt$1*2.44270  ];
    [ initiat$1*1.17768   ];
```

```
    [ engage$1*4.46855    ];
    [ approach$1*1.51338  ];
    [ response$1*2.11385  ];
    [ expect$1*2.56064    ];
    [ new$1*-2.43669      ];
    [ same$1*-1.04154     ];
    [ relative$1*-1.78324 ];
    [ objects$1*-1.14240  ];
    [ sequence$1*-0.84974 ];
    [ trans$1*-0.89213    ];
    [ avoid$1*-1.11250    ];
    [ control$1*-1.78770  ];
    [ touch$1*-0.87239    ];

    !!! REPLICATION SAMPLE (SAMPLE 2) !!! REPLICATION SAMPLE (SAMPLE 2) !!!

    %CG#2.C#1%

    [ primary$1*-0.21760  ];
    [ change$1*-0.45914   ];
    [ interupt$1*-0.64034 ];
    [ initiat$1*-0.46809  ];
    [ engage$1*-0.33510   ];
    [ approach$1*-0.42163 ];
    [ response$1*-0.58459 ];
    [ expect$1*-0.60368   ];
    [ new$1*1.17714       ];
    [ same$1*2.92160      ];
    [ relative$1*1.86782  ];
    [ objects$1*-1.97843  ];
    [ sequence$1*-1.40431 ];
    [ trans$1*-1.31741    ];
    [ avoid$1*0.99669     ];
    [ control$1*1.69313   ];
    [ touch$1*1.24603     ];

    %CG#2.C#2%

    [ primary$1*3.45112   ];
    [ change$1*2.10148    ];
    [ interupt$1*2.33173  ];
    [ initiat$1*4.67053   ];
    [ engage$1*2.61317    ];
    [ approach$1*3.00689  ];
    [ response$1*3.01867  ];
    [ expect$1*2.47608    ];
    [ new$1*2.13464       ];
    [ same$1*2.61928      ];
    [ relative$1*2.46710  ];
    [ objects$1*3.49010   ];
    [ sequence$1*2.16238  ];
    [ trans$1*1.76331     ];
    [ avoid$1*1.62916     ];
```

```
[ control$1*2.77768   ];
[ touch$1*2.98188     ];

%CG#2.C#3%

[ primary$1*-1.15805  ];
[ change$1*-1.05640   ];
[ interupt$1*-1.47052 ];
[ initiat$1*-1.88494  ];
[ engage$1*-0.85925   ];
[ approach$1*-1.29676 ];
[ response$1*-1.60551 ];
[ expect$1*-1.84879   ];
[ new$1*1.21384       ];
[ same$1*1.25810      ];
[ relative$1*1.00299  ];
[ objects$1*2.02420   ];
[ sequence$1*2.49353  ];
[ trans$1*1.91476     ];
[ avoid$1*1.22229     ];
[ control$1*3.67528   ];
[ touch$1*2.02070     ];

%CG#2.C#4%

[ primary$1*-2.45913  ];
[ change$1*-2.07089   ];
[ interupt$1*-2.29483 ];
[ initiat$1*-2.28241  ];
[ engage$1*-1.87192   ];
[ approach$1*-2.25327 ];
[ response$1*-2.27708 ];
[ expect$1*-2.29726   ];
[ new$1*-2.86563      ];
[ same$1*-2.02751     ];
[ relative$1*-2.09303 ];
[ objects$1*-3.25774  ];
[ sequence$1*-2.18406 ];
[ trans$1*-2.90318    ];
[ avoid$1*-2.16937    ];
[ control$1*-1.95077  ];
[ touch$1*-1.64226    ];

%CG#2.C#5%

[ primary$1*3.92177   ];
[ change$1*1.18687    ];
[ interupt$1*2.44270  ];
[ initiat$1*1.17768   ];
[ engage$1*4.46855    ];
[ approach$1*1.51338  ];
[ response$1*2.11385  ];
[ expect$1*2.56064    ];
```

```
     [ new$1*-2.43669      ];
     [ same$1*-1.04154     ];
     [ relative$1*-1.78324 ];
     [ objects$1*-1.14240  ];
     [ sequence$1*-0.84974 ];
     [ trans$1*-0.89213    ];
     [ avoid$1*-1.11250    ];
     [ control$1*-1.78770  ];
     [ touch$1*-0.87239    ];   ",


  OUTPUT = "",

  PLOT = "type = plot3;
          series = primary-touch(*);",

  usevariables = colnames(combined_data),
  rdata = combined_data)

m_step2_fit <- mplusModeler(m_step2,
                 dataout=here("C4-Replication", "Step2_Replication.dat"),
                 modelout=here("C4-Replication", "Step2_Replication.inp") ,
                 check=TRUE, run = TRUE, hashfilename = FALSE)
```

Step 3

Estimate multigroup model with thresholds equal across groups

```
m_step3  <- mplusObject(

  TITLE = "Step 3 - Replication",

  VARIABLE =
   "categorical = primary-touch;
    usevar = primary-touch;

    knownclass = CG (sample=1 sample=2); !!! LABEL SAMPLES !!!
    classes = CG(2) C(5);                !!! CG is KNOWN CLASS !!!
   ",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    starts = 500 200;
    !STSEED = ; !!! USE SEED TO REPLICATE THESIS RESULTS !!!
   ",

  MODEL =
```

```
"%OVERALL%

 !!! INITIAL SAMPLE (SAMPLE 1) !!! INITIAL SAMPLE (SAMPLE 1) !!!

 %CG#1.C#1%

 [ primary$1*-0.21760  ](1);
 [ change$1*-0.45914   ](2);
 [ interupt$1*-0.64034 ](3);
 [ initiat$1*-0.46809  ](4);
 [ engage$1*-0.33510   ](5);
 [ approach$1*-0.42163 ](6);
 [ response$1*-0.58459 ](7);
 [ expect$1*-0.60368   ](8);
 [ new$1*1.17714       ](9);
 [ same$1*2.92160      ](10);
 [ relative$1*1.86782  ](11);
 [ objects$1*-1.97843  ](12);
 [ sequence$1*-1.40431 ](13);
 [ trans$1*-1.31741    ](14);
 [ avoid$1*0.99669     ](15);
 [ control$1*1.69313   ](16);
 [ touch$1*1.24603     ](17);

 %CG#1.C#2%

 [ primary$1*3.45112   ](18);
 [ change$1*2.10148    ](19);
 [ interupt$1*2.33173  ](20);
 [ initiat$1*4.67053   ](21);
 [ engage$1*2.61317    ](22);
 [ approach$1*3.00689  ](23);
 [ response$1*3.01867  ](24);
 [ expect$1*2.47608    ](25);
 [ new$1*2.13464       ](26);
 [ same$1*2.61928      ](27);
 [ relative$1*2.46710  ](28);
 [ objects$1*3.49010   ](29);
 [ sequence$1*2.16238  ](30);
 [ trans$1*1.76331     ](31);
 [ avoid$1*1.62916     ](32);
 [ control$1*2.77768   ](33);
 [ touch$1*2.98188     ](34);

 %CG#1.C#3%

 [ primary$1*-1.15805  ](35);
 [ change$1*-1.05640   ](36);
 [ interupt$1*-1.47052 ](37);
 [ initiat$1*-1.88494  ](38);
 [ engage$1*-0.85925   ](39);
 [ approach$1*-1.29676 ](40);
 [ response$1*-1.60551 ](41);
```

```
[ expect$1*-1.84879   ](42);
[ new$1*1.21384       ](43);
[ same$1*1.25810      ](44);
[ relative$1*1.00299  ](45);
[ objects$1*2.02420   ](46);
[ sequence$1*2.49353  ](47);
[ trans$1*1.91476     ](48);
[ avoid$1*1.22229     ](49);
[ control$1*3.67528   ](50);
[ touch$1*2.02070     ](51);

%CG#1.C#4%

[ primary$1*-2.45913  ](52);
[ change$1*-2.07089   ](53);
[ interupt$1*-2.29483 ](54);
[ initiat$1*-2.28241  ](55);
[ engage$1*-1.87192   ](56);
[ approach$1*-2.25327 ](57);
[ response$1*-2.27708 ](58);
[ expect$1*-2.29726   ](59);
[ new$1*-2.86563      ](60);
[ same$1*-2.02751     ](61);
[ relative$1*-2.09303 ](62);
[ objects$1*-3.25774  ](63);
[ sequence$1*-2.18406 ](64);
[ trans$1*-2.90318    ](65);
[ avoid$1*-2.16937    ](66);
[ control$1*-1.95077  ](67);
[ touch$1*-1.64226    ](68);

%CG#1.C#5%

[ primary$1*3.92177   ](69);
[ change$1*1.18687    ](70);
[ interupt$1*2.44270  ](71);
[ initiat$1*1.17768   ](72);
[ engage$1*4.46855    ](73);
[ approach$1*1.51338  ](74);
[ response$1*2.11385  ](75);
[ expect$1*2.56064    ](76);
[ new$1*-2.43669      ](77);
[ same$1*-1.04154     ](78);
[ relative$1*-1.78324 ](79);
[ objects$1*-1.14240  ](80);
[ sequence$1*-0.84974 ](81);
[ trans$1*-0.89213    ](82);
[ avoid$1*-1.11250    ](83);
[ control$1*-1.78770  ](84);
[ touch$1*-0.87239    ](85);

!!! REPLICATION SAMPLE (SAMPLE 2) !!! REPLICATION SAMPLE (SAMPLE 2) !!!
```

```
%CG#2.C#1%

[ primary$1*-0.21760  ](1);
[ change$1*-0.45914   ](2);
[ interupt$1*-0.64034 ](3);
[ initiat$1*-0.46809  ](4);
[ engage$1*-0.33510   ](5);
[ approach$1*-0.42163 ](6);
[ response$1*-0.58459 ](7);
[ expect$1*-0.60368   ](8);
[ new$1*1.17714       ](9);
[ same$1*2.92160      ](10);
[ relative$1*1.86782  ](11);
[ objects$1*-1.97843  ](12);
[ sequence$1*-1.40431 ](13);
[ trans$1*-1.31741    ](14);
[ avoid$1*0.99669     ](15);
[ control$1*1.69313   ](16);
[ touch$1*1.24603     ](17);

%CG#2.C#2%

[ primary$1*3.45112   ](18);
[ change$1*2.10148    ](19);
[ interupt$1*2.33173  ](20);
[ initiat$1*4.67053   ](21);
[ engage$1*2.61317    ](22);
[ approach$1*3.00689  ](23);
[ response$1*3.01867  ](24);
[ expect$1*2.47608    ](25);
[ new$1*2.13464       ](26);
[ same$1*2.61928      ](27);
[ relative$1*2.46710  ](28);
[ objects$1*3.49010   ](29);
[ sequence$1*2.16238  ](30);
[ trans$1*1.76331     ](31);
[ avoid$1*1.62916     ](32);
[ control$1*2.77768   ](33);
[ touch$1*2.98188     ](34);

%CG#2.C#3%

[ primary$1*-1.15805  ](35);
[ change$1*-1.05640   ](36);
[ interupt$1*-1.47052 ](37);
[ initiat$1*-1.88494  ](38);
[ engage$1*-0.85925   ](39);
[ approach$1*-1.29676 ](40);
[ response$1*-1.60551 ](41);
[ expect$1*-1.84879   ](42);
[ new$1*1.21384       ](43);
[ same$1*1.25810      ](44);
[ relative$1*1.00299  ](45);
```

```
    [ objects$1*2.02420   ](46);
    [ sequence$1*2.49353  ](47);
    [ trans$1*1.91476     ](48);
    [ avoid$1*1.22229     ](49);
    [ control$1*3.67528   ](50);
    [ touch$1*2.02070     ](51);

    %CG#2.C#4%

    [ primary$1*-2.45913  ](52);
    [ change$1*-2.07089   ](53);
    [ interupt$1*-2.29483 ](54);
    [ initiat$1*-2.28241  ](55);
    [ engage$1*-1.87192   ](56);
    [ approach$1*-2.25327 ](57);
    [ response$1*-2.27708 ](58);
    [ expect$1*-2.29726   ](59);
    [ new$1*-2.86563      ](60);
    [ same$1*-2.02751     ](61);
    [ relative$1*-2.09303 ](62);
    [ objects$1*-3.25774  ](63);
    [ sequence$1*-2.18406 ](64);
    [ trans$1*-2.90318    ](65);
    [ avoid$1*-2.16937    ](66);
    [ control$1*-1.95077  ](67);
    [ touch$1*-1.64226    ](68);

    %CG#2.C#5%

    [ primary$1*3.92177   ](69);
    [ change$1*1.18687    ](70);
    [ interupt$1*2.44270  ](71);
    [ initiat$1*1.17768   ](72);
    [ engage$1*4.46855    ](73);
    [ approach$1*1.51338  ](74);
    [ response$1*2.11385  ](75);
    [ expect$1*2.56064    ](76);
    [ new$1*-2.43669      ](77);
    [ same$1*-1.04154     ](78);
    [ relative$1*-1.78324 ](79);
    [ objects$1*-1.14240  ](80);
    [ sequence$1*-0.84974 ](81);
    [ trans$1*-0.89213    ](82);
    [ avoid$1*-1.11250    ](83);
    [ control$1*-1.78770  ](84);
    [ touch$1*-0.87239    ](85);   ",


OUTPUT = "",

PLOT = "type = plot3;
        series = primary-touch(*);",
```

```
  usevariables = colnames(combined_data),
  rdata = combined_data)

m_step3_fit <- mplusModeler(m_step3,
                  dataout=here("C4-Replication", "Step3_Replication.dat"),
                  modelout=here("C4-Replication", "Step3_Replication.inp") ,
                  check=TRUE, run = TRUE, hashfilename = FALSE)
```

Step 4: Compare model fit

- Unconstrained model: `Step2_Replication.out` (no threshold constraints)
- Constrained model: `Step3_Replication.out` (thresholds constrained to equality across groups)

```
replication_models <- readModels(here("C4-Replication"), quiet = TRUE)
```

Compare model fit summary statistics: Exploratory & Confirmatory Models

```
enum_extract1 <- LatexSummaryTable(replication_models$Step2_Replication.out, keepCols = c("Title",
    "Parameters", "LL", "BIC", "aBIC", "Observations"))
enum_extract2 <- LatexSummaryTable(replication_models$Step3_Replication.out, keepCols = c("Title",
    "Parameters", "LL", "BIC", "aBIC", "Observations"))
```

Calculate indices derived from the Log Likelihood (LL)

```
allFit <- rbind(enum_extract1, enum_extract2) %>%
    mutate(aBIC = -2 * LL + Parameters * log((Observations + 2)/24)) %>%
    select(1:5) %>%
    mutate(Title = case_when(Title == " Step 2 - Replication" ~ "Unconstrained model",
        Title == " Step 3 - Replication" ~ "Constrained model"))
```

Format fit table

```
allFit %>%
    gt() %>%
    tab_header(title = md("**Model Fit Comparision Table**"), subtitle = md(" ")) %>%
    cols_label(Title = "Model", Parameters = md("Par"), LL = md("*LL*"), BIC = md("BIC"),
        aBIC = md("aBIC")) %>%
    tab_footnote(footnote = md("*Note.* Par = Parameters; *LL* = model log likelihood;
     BIC = Bayesian information criterion; aBIC = sample size adjusted BIC."),
        locations = cells_title()) %>%
    tab_options(column_labels.font.weight = "bold")
```

Model Fit Comparision Table[1]

[1]

Model	Par	*LL*	BIC	aBIC
Unconstrained model	175	-56657.49	114837.4	114281.3
Constrained model	90	-56700.41	114183.8	113897.8

[1] *Note.* Par = Parameters; *LL* = model log likelihood; BIC = Bayesian information criterion; aBIC = sample size adjusted BIC.