Project Report

Tim, Xiaofan, Yanwen, Jingning

11/20/2019

Introduction

Client is investigating how foreign language teachers feel about and utilize methods from the Teacher Effectiveness for Language Learning (TELL), and seeking advice about how to improving the survey.

Our purpose for our client in this project: 1. A lot of people don't answer the survey because it's long. Can we reduce the number of questions? 2. Is the survey currently answering the research questions?

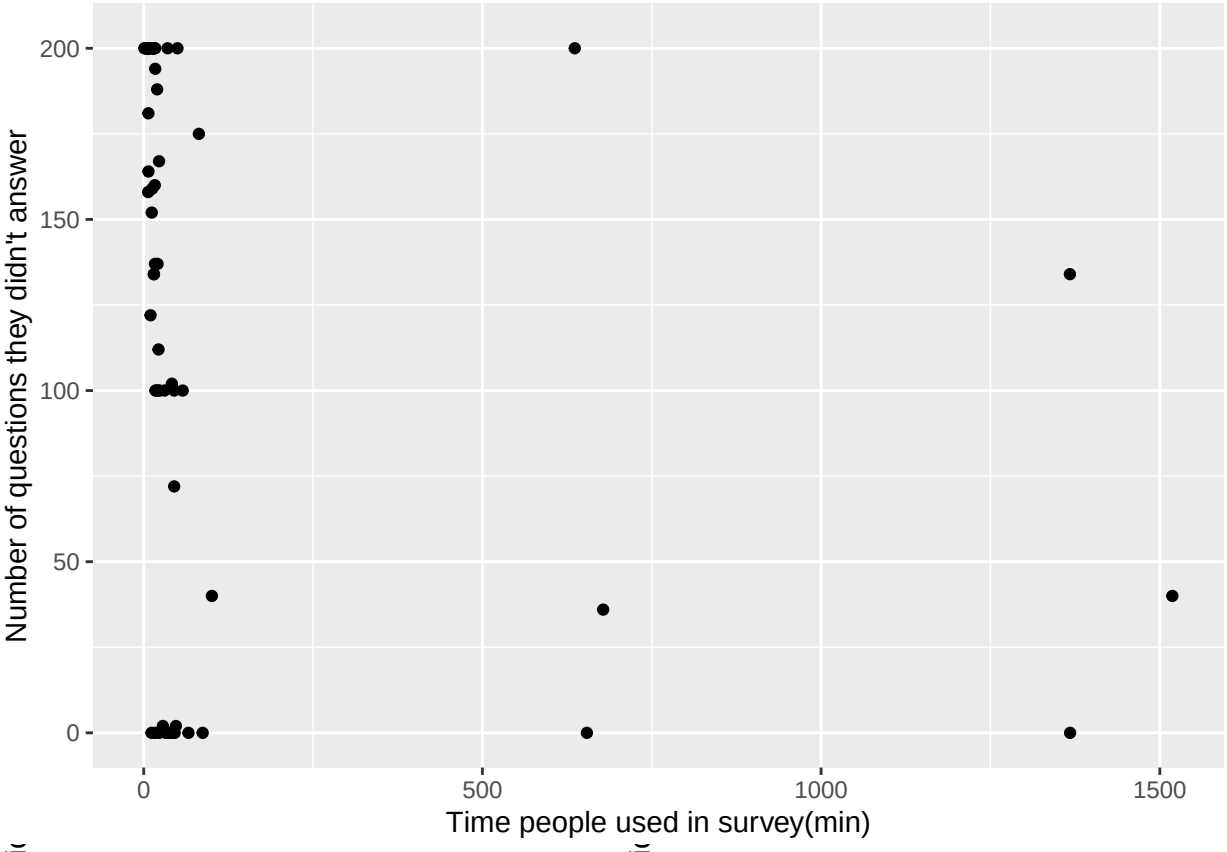
EDA & Conerns

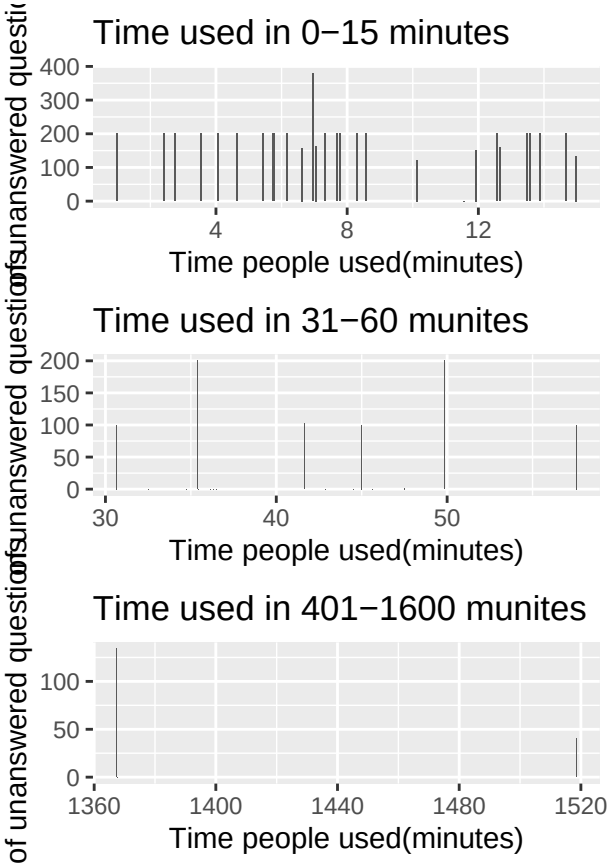
Data Structure

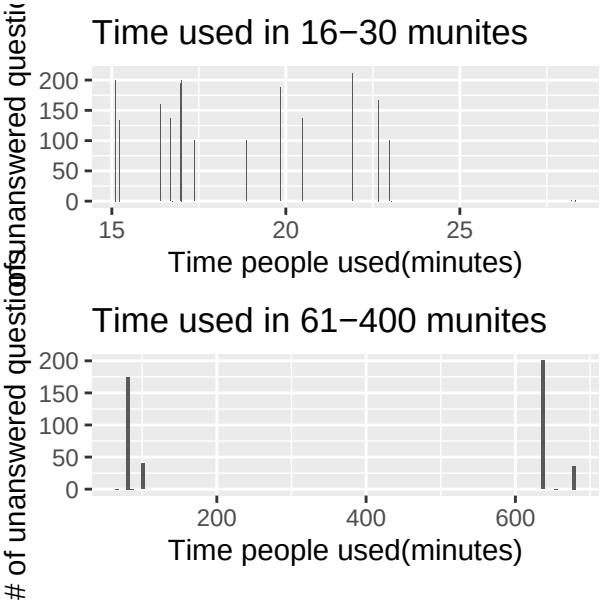
We are provided the data in an excel file with 6 spreadsheets including one sheet of notes, one sheet of personal information and 4 sheets of Teacher Effectiveness for Language Learning (TELL) framework survey questions. The dataset of personal information contains questions regarding respondents' teaching language and education background. The dataset of Teacher Effectiveness for Language Learning (TELL) framework survey contains around 200 questions asking about respondents' attitudes of contribution and confidence towards each practice in the framework. There are 4 domains of Teacher Effectiveness for Language Learning (TELL) survey questions: planning, learning experience, learning tools, and performance & feedback. Each domain contains several subdomains and each subdomain contains a different number of questions. For this project, we focus on reducing the number of questions in the dataset of Teacher Effectiveness for Language Learning (TELL) framework survey.

EDA

We conduct a basic Exploratory Data Analysis (EDA) for this project. Firstly, We focus on the time for respondents to complete this survey.


Number of questions they didn't answer
200
150
100
50
0
0
500
1000
1500
Time people used in survey(min)


Time used in 0−15 minutes
# of unanswered questions
400
300
200
100
0
4
8
12
Time people used(minutes)
Time used in 31−60 munites
200
150
100
50
0
30
40
50
Time people used(minutes)
Time used in 401−1600 munites
100
50
0
1360
1400
1440
1480
1520
Time people used(minutes)


Time used in 16−30 munites
# of unanswered questions
200
150
100
50
0
15
20
25
Time people used(minutes)
Time used in 61−400 munites
200
150
100
50
0
200
400
600
Time people used(minutes)

Data Cleaning

Concerns

We come up with several concerns after the initial EDA. Firstly, the observations we can use in the analysis are very limited since there are many N/A in the dataset. Secondly, some respondents seem like choosing the same answer through the whole survey and if we identify these answers as non-valid, then our sample size would become even smaller. With this limited sample size, the accuracy and reference of results from our subsequent analysis could be affected.

Method we used

We will use Confirmatory Factor Analysis (CFA) to reduce the survey questions number. CFA is a special form of factor analysis and mostly used in social science research. It is used to check whether measures of a construct are consistent with a researcher's understanding of the nature of that construct.
Here we will use CFA to see if there are survey questions equivalent to each other so we can reduce those repeated questions. We will analysis each subdomain separately and will only consider the problems regarding confidence or not. Within each subdomain, there will be several questions and our null hypothesis is that all survey questions are identical to each other. Then our alternative hypothesis is that the questions are not all equal.
We will focus on the p value result we have from CFA and we will take a p value larger than 0.05 to reject our null hypothesis. When we are not able to reject our null hypothesis, we will look at our factor loading to check the correlations between questions. Then we will fit new model by dropping question with lowest factor loading and see if we will reject our null hypothesis now. We will keep doing this until we have a subdomain with an acceptable p value, which gives us a set of survey questions are not identical to each other.

Analysis

Learning Tool Domain Analysis

```
## lavaan 0.6-5 ended normally after 12 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##   Row rank of the constraints matrix                 1
##
##                                                   Used       Total
##   Number of observations                            27          84
##
## Model Test User Model:
##
##   Test statistic                                 0.109
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.741
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard errors                             Standard
##
## Latent Variables:
```

```
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   lt1 =~
##     LT1_Cnfdn (aa)    0.384    0.165    2.331    0.020    0.384    0.488
##     LT1b_Cnfd (aa)    0.384    0.165    2.331    0.020    0.384    0.472
##     LT1c_Cnfd         0.612    0.282    2.167    0.030    0.612    0.722
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .LT1a_Confidenc    0.472    0.168    2.816    0.005    0.472    0.762
##    .LT1b_Confidenc    0.515    0.177    2.914    0.004    0.515    0.777
##    .LT1c_Confidenc    0.344    0.314    1.095    0.273    0.344    0.479
##     lt1               1.000                              1.000    1.000
```

```
##                npar                fmin               chisq
##               5.000               0.002               0.109
##                  df              pvalue      baseline.chisq
##               1.000               0.741               7.404
##         baseline.df     baseline.pvalue                 cfi
##               3.000               0.060               1.000
##                 tli                nnfi                 rfi
##               1.607               1.607               0.956
##                 nfi                pnfi                 ifi
##               0.985               0.328               1.139
##                 rni                logl   unrestricted.logl
##               1.202            -94.804             -94.750
##                 aic                 bic              ntotal
##             199.609             206.088              27.000
##                bic2               rmsea      rmsea.ci.lower
##             190.555               0.000               0.000
##      rmsea.ci.upper        rmsea.pvalue                 rmr
##               0.355               0.749               0.018
##          rmr_nomean                srmr        srmr_bentler
##               0.018               0.028               0.028
## srmr_bentler_nomean                crmr         crmr_nomean
##               0.028               0.028               0.028
##          srmr_mplus   srmr_mplus_nomean                cn_05
##               0.026               0.026             951.942
##               cn_01                 gfi                agfi
##            1643.449               0.997               0.984
##                pgfi                 mfi                ecvi
##               0.166               1.017               0.374
```

```
##                lhs op             rhs    mi    epc sepc.lv sepc.all
## 10 LT1a_Confidence ~~ LT1c_Confidence 0.109  0.053   0.053    0.133
## 11 LT1b_Confidence ~~ LT1c_Confidence 0.109 -0.053  -0.053   -0.127
##    sepc.nox
## 10    0.133
## 11   -0.127
```

Table 1: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt1	LT1a_Confidence	0.384	0.165	2.331	0.02	0.488
lt1	LT1b_Confidence	0.384	0.165	2.331	0.02	0.472

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt1	LT1c_Confidence	0.612	0.282	2.167	0.03	0.722

```
## lavaan 0.6-5 ended normally after 12 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##   Row rank of the constraints matrix                 1
##
##                                                   Used       Total
##   Number of observations                            28          84
##
## Model Test User Model:
##
##   Test statistic                                 0.003
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.953
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   lt2 =~
##     LT2_Cnfdn (aa)    0.443    0.126    3.531    0.000    0.443    0.587
##     LT2b_Cnfd (aa)    0.443    0.126    3.531    0.000    0.443    0.603
##     LT2c_Cnfd         0.776    0.222    3.499    0.000    0.776    0.817
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .LT2a_Confidenc    0.373    0.127    2.930    0.003    0.373    0.655
##    .LT2b_Confidenc    0.345    0.122    2.838    0.005    0.345    0.637
##    .LT2c_Confidenc    0.300    0.270    1.110    0.267    0.300    0.333
##     lt2               1.000                              1.000    1.000
```

```
##                npar                fmin               chisq
##               5.000               0.000               0.003
##                  df              pvalue      baseline.chisq
##               1.000               0.953              15.764
##         baseline.df     baseline.pvalue                 cfi
##               3.000               0.001               1.000
##                 tli                nnfi                 rfi
##               1.234               1.234               0.999
##                 nfi                pnfi                 ifi
##               1.000               0.333               1.067
##                 rni                logl   unrestricted.logl
##               1.078            -93.396             -93.394
##                 aic                 bic              ntotal
##             196.792             203.453              28.000
```

```
##                bic2               rmsea      rmsea.ci.lower
##             187.908               0.000               0.000
##      rmsea.ci.upper        rmsea.pvalue                 rmr
##               0.000               0.955               0.003
##          rmr_nomean                srmr        srmr_bentler
##               0.003               0.005               0.005
## srmr_bentler_nomean                crmr         crmr_nomean
##               0.005               0.004               0.004
##          srmr_mplus   srmr_mplus_nomean               cn_05
##               0.005               0.005           31204.347
##               cn_01                 gfi                agfi
##           53894.843               1.000               1.000
##                pgfi                 mfi                ecvi
##               0.167               1.018               0.357
```

```
##                lhs op             rhs    mi    epc sepc.lv sepc.all
## 10 LT2a_Confidence ~~ LT2c_Confidence 0.003 -0.009  -0.009   -0.027
## 11 LT2b_Confidence ~~ LT2c_Confidence 0.003  0.009   0.009    0.028
##    sepc.nox
## 10   -0.027
## 11    0.028
```

Table 2: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt2	LT2a_Confidence	0.443	0.126	3.531	0	0.587
lt2	LT2b_Confidence	0.443	0.126	3.531	0	0.603
lt2	LT2c_Confidence	0.776	0.222	3.499	0	0.817

```
## lavaan 0.6-5 ended normally after 14 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##   Row rank of the constraints matrix                 1
##
##                                                   Used       Total
##   Number of observations                            27          84
##
## Model Test User Model:
##
##   Test statistic                                 0.017
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.897
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard errors                             Standard
##
## Latent Variables:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
```

```
##   lt3 =~
##     LT3_Cnfdn         0.931    0.164    5.660    0.000    0.931    0.959
##     LT3b_Cnfd (aa)    0.706    0.144    4.892    0.000    0.706    0.802
##     LT3d_Cnfd (aa)    0.706    0.144    4.892    0.000    0.706    0.629
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .LT3a_Confidenc    0.075    0.170    0.442    0.659    0.075    0.080
##    .LT3b_Confidenc    0.276    0.122    2.259    0.024    0.276    0.357
##    .LT3d_Confidenc    0.761    0.228    3.330    0.001    0.761    0.604
##     lt3               1.000                              1.000    1.000
```

```
##                npar                fmin               chisq
##               5.000               0.000               0.017
##                  df              pvalue      baseline.chisq
##               1.000               0.897              36.819
##         baseline.df     baseline.pvalue                 cfi
##               3.000               0.000               1.000
##                 tli                nnfi                 rfi
##               1.087               1.087               0.999
##                 nfi                pnfi                 ifi
##               1.000               0.333               1.027
##                 rni                logl   unrestricted.logl
##               1.029            -95.478             -95.469
##                 aic                 bic              ntotal
##             200.955             207.434              27.000
##                bic2               rmsea      rmsea.ci.lower
##             191.902               0.000               0.000
##      rmsea.ci.upper        rmsea.pvalue                 rmr
##               0.233               0.900               0.014
##          rmr_nomean                srmr        srmr_bentler
##               0.014               0.012               0.012
## srmr_bentler_nomean                crmr         crmr_nomean
##               0.012               0.007               0.007
##          srmr_mplus   srmr_mplus_nomean                cn_05
##               0.011               0.011            6147.974
##               cn_01                 gfi                agfi
##           10617.940               1.000               0.998
##                pgfi                 mfi                ecvi
##               0.167               1.018               0.371
```

```
##                lhs op             rhs    mi    epc sepc.lv sepc.all
## 9  LT3a_Confidence ~~ LT3b_Confidence 0.017 -0.025  -0.025   -0.171
## 10 LT3a_Confidence ~~ LT3d_Confidence 0.017  0.025   0.025    0.103
##    sepc.nox
## 9    -0.171
## 10    0.103
```

Table 3: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt3	LT3a_Confidence	0.931	0.164	5.660	0	0.959
lt3	LT3b_Confidence	0.706	0.144	4.892	0	0.802
lt3	LT3d_Confidence	0.706	0.144	4.892	0	0.629

```
## lavaan 0.6-5 ended normally after 19 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##   Row rank of the constraints matrix                 1
##
##                                                   Used       Total
##   Number of observations                            28          84
##
## Model Test User Model:
##
##   Test statistic                                 0.016
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.899
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   lt4 =~
##     LT4_Cnfdn         0.370    0.140    2.640    0.008    0.370    0.469
##     LT4b_Cnfd (aa)    1.077    0.150    7.191    0.000    1.077    1.012
##     LT4c_Cnfd (aa)    1.077    0.150    7.191    0.000    1.077    0.886
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .LT4a_Confidenc    0.485    0.129    3.749    0.000    0.485    0.780
##    .LT4b_Confidenc   -0.027    0.095   -0.282    0.778   -0.027   -0.024
##    .LT4c_Confidenc    0.316    0.127    2.486    0.013    0.316    0.214
##     lt4               1.000                              1.000    1.000

##                npar                fmin               chisq
##               5.000               0.000               0.016
##                  df              pvalue      baseline.chisq
##               1.000               0.899              52.579
##         baseline.df     baseline.pvalue                 cfi
##               3.000               0.000               1.000
##                tli                nnfi                 rfi
##               1.060               1.060               0.999
##                nfi                pnfi                 ifi
##               1.000               0.333               1.019
##                rni                logl   unrestricted.logl
##               1.020            -93.343            -93.335
##                aic                 bic              ntotal
##           196.686             203.347              28.000
##                bic2               rmsea      rmsea.ci.lower
##           187.802               0.000               0.000
##      rmsea.ci.upper        rmsea.pvalue                 rmr
```

```
##               0.225                0.902                0.008
##          rmr_nomean                 srmr         srmr_bentler
##               0.008                0.007                0.007
## srmr_bentler_nomean                 crmr          crmr_nomean
##               0.007                0.006                0.006
##          srmr_mplus   srmr_mplus_nomean                cn_05
##               0.006                0.006             6648.164
##               cn_01                  gfi                 agfi
##           11481.859                1.000                0.998
##                pgfi                  mfi                 ecvi
##               0.167                1.018                0.358
```

```
##                lhs op             rhs    mi   epc sepc.lv sepc.all
## 9  LT4a_Confidence ~~ LT4b_Confidence 0.016  0.01    0.01    0.089
## 10 LT4a_Confidence ~~ LT4c_Confidence 0.016 -0.01   -0.01   -0.026
##    sepc.nox
## 9     0.089
## 10   -0.026
```

Table 4: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt4	LT4a_Confidence	0.370	0.14	2.640	0.008	0.469
lt4	LT4b_Confidence	1.077	0.15	7.191	0.000	1.012
lt4	LT4c_Confidence	1.077	0.15	7.191	0.000	0.886

```
## lavaan 0.6-5 ended normally after 13 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##   Row rank of the constraints matrix                 1
##
##                                                   Used       Total
##   Number of observations                            26          84
##
## Model Test User Model:
##
##   Test statistic                                 0.774
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.379
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   lt5 =~
##     LT5_Cnfdn (aa)    0.398    0.124    3.205    0.001    0.398    0.618
##     LT5b_Cnfd         0.620    0.186    3.331    0.001    0.620    0.844
```

```
##     LT5c_Cnfd (aa)    0.398    0.124    3.205    0.001    0.398    0.532
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .LT5a_Confidenc    0.257    0.101    2.544    0.011    0.257    0.618
##    .LT5b_Confidenc    0.155    0.186    0.832    0.405    0.155    0.287
##    .LT5c_Confidenc    0.402    0.132    3.034    0.002    0.402    0.717
##     lt5               1.000                              1.000    1.000
```

```
##               npar               fmin              chisq
##              5.000              0.015              0.774
##                 df             pvalue      baseline.chisq
##              1.000              0.379             15.415
##         baseline.df     baseline.pvalue                cfi
##              3.000              0.001              1.000
##                tli               nnfi                rfi
##              1.055              1.055              0.849
##                nfi               pnfi                ifi
##              0.950              0.317              1.016
##                rni               logl   unrestricted.logl
##              1.018            -76.441            -76.054
##                aic                bic             ntotal
##            162.882            169.172             26.000
##               bic2              rmsea      rmsea.ci.lower
##            153.653              0.000              0.000
##      rmsea.ci.upper        rmsea.pvalue                rmr
##              0.494              0.394              0.041
##          rmr_nomean               srmr        srmr_bentler
##              0.041              0.079              0.079
## srmr_bentler_nomean               crmr         crmr_nomean
##              0.079              0.063              0.063
##          srmr_mplus   srmr_mplus_nomean               cn_05
##              0.070              0.070            130.124
##              cn_01                gfi               agfi
##            224.020              0.981              0.885
##               pgfi                mfi               ecvi
##              0.163              1.004              0.414
```

```
##                lhs op             rhs    mi    epc sepc.lv sepc.all
## 9  LT5a_Confidence ~~ LT5b_Confidence 0.762 -0.102  -0.102   -0.512
## 11 LT5b_Confidence ~~ LT5c_Confidence 0.762  0.102   0.102    0.409
##    sepc.nox
## 9    -0.512
## 11    0.409
```

Table 5: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	loading
lt5	LT5a_Confidence	0.398	0.124	3.205	0.001	0.618
lt5	LT5b_Confidence	0.620	0.186	3.331	0.001	0.844
lt5	LT5c_Confidence	0.398	0.124	3.205	0.001	0.532

Conclusion / Discussion

Appendix