“Final Project PSY 610 Structural Equation Modeling”

A Structural Model of The Effects of Preschool Attention on Kindergarten Literacy

Merly Klaas

Contents

Orientation

The paper I use for this project examined the influence of attention in prekindergarten on kindergarten decoding abilities in kindergarten. This study argued that early attention problems in prekindergarten may influence the acquisition of emergent literacy skills including alphabetic knowledge, phonological awareness, and receptive and expressive vocabulary with decoding as an outcome. In addition, Maternal Education was included in the model because it relates to early experiences with prints indicator that accounts for differences in reading at school entry.

Participants were 250 children attending public, lottery-funded prekindergarten in 26 classrooms in 18 schools from three urban and metropolitan counties in northeast Georgia.

For this study, the authors used the following measurements:

1. Attention Problem : An experimental short form of the BASC Teacher Rating Scale- Preschool (TRS-P, Reynolds & Kamphaus, 1992) called the BASC Screener (Yanosky, 2005; Yanosky et al., 2011). There were 6-items with highest loading on attention problem included in the model. They were measuring issues of attention shifting, attention span, concentration, listening attentively, listening to directions, and distraction.

2. Emergent Literacy:

- Alphabetic knowledge was measured through an experimenter-developed alphabetic knowledge test.
- Phonological awareness was measured by a subset of items from the Phonological Awareness Test (PAT, Robertson & Salter, 1997).
- Receptive vocabulary was assessed using the Peabody Picture Vocabulary Test- III (PPVT-III, Dunn & Dunn, 1997).
- The Expressive Vocabulary Test (EVT, Williams, 1997) was administered to evaluate children's expressive vocabulary knowledge.

3. Kindergarten Reading Ability Kindergarten reading ability was assessed using The Early Decoding Test, which is an experimenter-constructed brief assessment designed to identify early readers (Schwanenflugel et al., 2010).

4. Maternal Education Maternal Education was measured by an open-ended question:*'What is the last grade that you completed in school?'* Responses were coded for this study, with 3% completing middle school, 13% high school, 54% completing high school, 12% some college, 12% Bachelors degree, and 6% Masters degree or higher.

Emergent Literacy was the mediator of Attention Problem and Attention Problem to Decoding skill in Kindergarten.

Reproducing the model:

Write a narrative describing the process of reproducing the model. How well was the model described in the paper? Could you reproduce it from the main paper itself? Were there supplemental materials that you needed to use? Were you able to get the same numbers published in the article? Was anything missing? Would you reach the same conclusions that the authors did?

The three nested models were clearly described in the paper. Here are some steps the authors reported on the paper for their three models:

- **The First Model** was a measurement model, which allowed for examination of interrelationships between the items and scores used in the full structural model. Model 1 was used the measurement model to examine how well the emergent literacy dimensions measured the latent factor of emergent literacy and how well the attention items from the BASC Screener measured a latent factor of attention.
- **The Second Model** was the full structural model with all latent variables were considered observed and part of structural relationships. The structural model was recursive, with no feedback loops, which was a sufficient condition for identification.
- **The Third Model** was nested from the full structural model with the path from attention to literacy set to zero.

I was able to reproduce the models described in the paper using correlation table and standard deviations reported for each variable except for Maternity Education. For this, I had to manually calculate the standard deviation for Maternal Education using information reported in the method section about the percentage of each maternal education category for 250 respondents (3% completing middle school, 13% high school, 54% completing high school, 12% some college, 12% Bachelors degree, and 6% Masters degree) to complete the data.

I run all three nested models described in the paper including the measurement model, full structural model, and model 3 in which path from attention to literacy set to zero.

In the measurement model (Model 1), two latent variables were measured by each observed items using the latent variable definition (=~) operator in Lavaan.

The Attention Problems manifested in six observed variables:

- Item 1 (a1)
- Item 2 (a2)
- Item 3 (a3)
- Item 4 (a4)
- Item 5 (a5)
- Item 6 (a5)

Emergent Literacy was manifested thorough four observed variables:

- Alphabetic Knowledge (ealpha)
- Phonological Awareness (ephono)
- Receptive Vocabulary (erecep)
- Expressive Vocabulary (eexpr)

I was able to get relatively close numbers published in the article including the fit statistics and standardized estimates for each latent variable. However, I did not get the same degree of freedom for Model 1 (Measurement Model). In the paper, Model 1 reported $df = 50$, while my model 1 $df = 45$. The paper explained that model 1 was used examine how well the emergent literacy dimensions measured the latent factor of emergent literacy and how well the attention items from the BASC Screener measured a latent factor of attention. I suspected that there were some modifications being done but was not specifically reported in the paper. However, since the focus of this analysis is to reproduce the Full Structural Model, further investigation is considered unnecessary.

In the Full Strucural Model (Model 2), I added the regression path nested in Model 1. According to the graph, Emergent Literacy variable was fully mediated all relationships among variables. Maternal Education as one of the exogenous variables and Decoding as the endogenous variable were added to the model. Below are the regression paths in this model:

- Attention Problems to Emergent Literacy: **emlit ~ atprob**
- Maternal Education to Emergent Literacy: **emlit ~ meduc**
- Emergent Literacy to Decoding Skill: **decod ~ emlit**

I also added a covariance between two exogenous variables as reflected in the graph.

- Maternal Education and Attention Problems: **meduc ~~ atprob**

The only graph included in this paper was representing the Full Structural Model (Model 2).

There are some findings worth mentioning regarding this process:

a. The paper used estimates from Std.lv column instead of Std.all column. The std.lv column shows results that are standardized so that the latent variables have a variance of one. It is more common to see std.all estimates being reported instead of std.lv. There is no explanation for this decision in the paper.

b. The estimate coefficient for attention problem latent variable to emergent literacy latent variable is reported as .26. I believe this is a misprint because the relationship between these two variables should be negative as the higher attention problem, the lower the emergent literacy score and vice versa. The estimate I got for this path is -.28

c. The loadings for item5 and item6 in attention problem variables reported were possible swapped. The paper reported loading for item5 = .68 and item6 = .62, while I got factor loadings for item5 = .62 and item6 = .68 on Attention Problem.

d. Item loadings on Emergent Literacy variable were moderately high around 0.69 - 0.73. The weakest loading was item receptive vocabulary: 0.69 and the highest loading was expressive vocabulary: 0.73.

e. Item loadings for attention shifting (a1), attention span (a2), and listening attentively (a4) were relatively high above 0.91 while the other three items concentration (a3), listening to direction (a5) and distraction (a6) were moderately high around 0.72 - 0.77.

Lastly, Model 3 was run to recheck the fit of the model after attention problem path to literacy was set to zero. With only Maternal Education variable, the fit statistics got worse. The author, then, conclude that individual differences in attention contributes more to literacy compared to maternal education.

Considering these results, I would have the same conclusion with some extensions. In the following model 4, I would like to add the covariance between observed variables as suggested by the modification indices to improve the fit statistics and try to add direct path from Attention Problems to Decoding Skill.

Extending the model:

Describe the other model of interest that you ran. What do you conclude based on it?

I decided to test two extended models.

First, in this extended model (Model 4a), I added covariances between items on Attention Problem Latent Variable as suggested by the modification indices analysis to improve the fit statistics of the model.

Covariances between items on Attention Problems:

- Item 5 Listening to Direction to Item 6 Distraction: **a5 ~~ a6**
- Item 2 Attention Span to Concentration: **a2 ~~ a3**
- Item 2 Attention Span to Distraction: **a2 ~~ a6**

With this modification, the fit statistics are improved. The proposed model was tested using the Lavaan Package R. Using both Kline (2005) and Hu and Bentler's (1999) guidelines for evaluating overall model fit, an SRMR<.08, an NNFI > .95, and a CFI > .95 indicated an adequate model fit to the observed data. Model 4a df = 49, SMSR = 0.038, NNFI = 0.967, CFI = 0.975.

In the second extended model (Model 4b), I added direct path from attention problem to decoding. I hypothesized that children with attention problem in preschool will also struggle to read in the kindergarten without any mediator.

- Attention Span to Decoding: **decod ~ atprob**

However, the fit statistics were not improved, instead it got worse. df = 51, Model 4b SMSR = 0.043, NNFI = 0.924, CFI = 0.941.

The results of this reproduced analysis are presented in the following table:

Table 1: Fit Indices of the Reproduced Analysis Models

Model	df	SRMR	NNFI	CFI
1. Measurement Model	45	0.041	0.96	0.97
2. Full Structural Model	52	0.044	0.93	0.94
3. Attention path set to 0	53	0.100	0.91	0.93
4. Model 4a	49	0.038	0.97	0.97
5. Model 4b	51	0.043	0.92	0.94

This means the full mediation model with covariances between items on Attention Problems variable (Model 2) fits the data better compared to Model 4b where direct path from attention problems to decoding was added. Furthermore, **Model 4a**, where covariances between items on Attention Problems were added, fits the data better compared to all other models.

Conclusion:

This study sheds light on the importance of attention in children literacy development in addition to the more common foundation pre-literacy skills such as alphabetic knowledge, phonological awareness, receptive and expressive vocabulary. Another study Miller et al. (2014) also found that there is no direct relationship between attention problem and reading comprehension. Rather the relationship was mediated by reading growth. The common rule of thumb suggests that the average of children attention span could be calculated by multiplying child age with 2 or 3. For example, for three-year-old toddler, the average attention span is between 6 to 9 minutes and for five-year-old child, the average attention span is between 10 - 15 minutes. However, a more rigorous measurement should be done to measure children attention span. Children with high attention problems need additional targeted activities to improve their attention. Thus, teachers and caregivers need to be equipped on how to support children attention development that appropriate to their age.

The structural equation modeling analysis from this study about preschool attention and kindergarten literacy Dice and Schwanenflugel (2012) can easily be reproduced. The authors did a great job in providing necessary information including the correlation table, SDs, and detailed explanation about how they run the models presented in this paper. I was able to reproduce the analysis with the statistical results relatively close to match the numbers reported in the paper for the Full Structural Model (Model 2). There were slightly differences in the decimal points, however, this is acceptable because only the correlation matrix instead of raw data was used for this analysis. To improve, the authors could describe more about the modification steps taken especially for Model 1.

A link or copy of the original journal article

This analysis used Rosseel (2012), Wickham et al. (2021), Chan et al. (2021), Müller (2020) packages from R Software.

Appendix

```
# calculating SD for Maternal Education
dat <- import(here("data", "maternal edu.csv"))

meducsd <- dat %>%
  summarize(mean = mean(value),
            sd = sd(value))
```

```
dice_lower_corr <- '
1.00,
.43, 1.00,
.44, .49, 1.00,
.49, .46, .49, 1.00,
.40, .50, .64, .49, 1.00,
-.26, -.31, -.27, -.30, -.31, 1.00,
-.27, -.29, -.23, -.25, -.31, .84, 1.00,
-.26, -.24, -.21, -.23, -.27, .68, .76, 1.00,
-.22, -.27, -.20, -.23, -.29, .85, .85, .67, 1.00,
-.17, -.20, -.12, -.15, -.22, .65, .62, .58, .65, 1.00,
-.16, -.24, -.24, -.24, -.30, .69, .66, .58, .70, .73, 1.00,
.36, .62, .41, .51, .44, -.30, -.27, -.29, -.27, -.24, -.30, 1.00

'

dice_sds <- c(1.134, 14.728, 7.888, 16.671, 12.574, .916, .867, .803, .954, .864, .900, 7.881)
dicedata <- getCov(dice_lower_corr,
                   sds = dice_sds,
                   names = c("meduc", "ealpha", "ephono", "erecep", "eexpr",
                             "a1", "a2", "a3", "a4", "a5", "a6", "decod")
                   )
```

Model 1 Measurement Model

```
mod1_cfa <- '
emlit =~ ealpha + ephono + erecep + eexpr
atprob =~ a1 + a2 + a3 + a4 + a5 + a6

'

mod1_fitted <- cfa(mod1_cfa,
                       sample.cov = dicedata,
                       sample.nobs = 250)
summary(mod1_fitted, standardized = TRUE,
        fit.measures = TRUE)
```

```
## lavaan 0.6-9 ended normally after 75 iterations
##
##   Estimator                                         ML
```

```
##   Optimization method                           NLMINB
##   Number of model parameters                        21
##
##   Number of observations                           250
##
## Model Test User Model:
##
##   Test statistic                               101.366
##   Degrees of freedom                                34
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1698.909
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.959
##   Tucker-Lewis Index (TLI)                       0.946
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -5081.541
##   Loglikelihood unrestricted model (H1)      -5030.858
##
##   Akaike (AIC)                               10205.082
##   Bayesian (BIC)                             10279.033
##   Sample-size adjusted Bayesian (BIC)        10212.461
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.089
##   90 Percent confidence interval - lower         0.069
##   90 Percent confidence interval - upper         0.109
##   P-value RMSEA <= 0.05                          0.001
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.041
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit =~
##     ealpha            1.000                               9.561    0.650
##     ephono            0.637    0.067    9.497    0.000    6.091    0.774
##     erecep            1.114    0.134    8.318    0.000   10.655    0.640
```

```
##     eexpr             1.050    0.109    9.636    0.000   10.034    0.800
##   atprob =~
##     a1                1.000                               0.836    0.915
##     a2                0.951    0.039   24.391    0.000    0.795    0.919
##     a3                0.737    0.045   16.272    0.000    0.617    0.769
##     a4                1.045    0.043   24.329    0.000    0.874    0.918
##     a5                0.743    0.051   14.461    0.000    0.621    0.720
##     a6                0.819    0.051   15.989    0.000    0.685    0.762
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit ~~
##     atprob           -3.226    0.661   -4.878    0.000   -0.404   -0.404
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ealpha          124.635   13.210    9.435    0.000  124.635    0.577
##    .ephono           24.868    3.321    7.489    0.000   24.868    0.401
##    .erecep          163.271   17.134    9.529    0.000  163.271    0.590
##    .eexpr            56.786    8.301    6.840    0.000   56.786    0.361
##    .a1                0.137    0.017    8.189    0.000    0.137    0.163
##    .a2                0.117    0.015    8.013    0.000    0.117    0.156
##    .a3                0.262    0.025   10.357    0.000    0.262    0.408
##    .a4                0.143    0.018    8.050    0.000    0.143    0.157
##    .a5                0.358    0.034   10.572    0.000    0.358    0.482
##    .a6                0.338    0.033   10.394    0.000    0.338    0.419
##     emlit            91.411   17.302    5.283    0.000    1.000    1.000
##     atprob            0.699    0.075    9.370    0.000    1.000    1.000
```

```
modindices(mod1_fitted, sort = TRUE, min = 10 )
```

```
##    lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
## 78  a5 ~~  a6 46.942  0.162   0.162    0.466    0.466
## 69  a2 ~~  a3 18.456  0.061   0.061    0.351    0.351
## 72  a2 ~~  a6 10.297 -0.052  -0.052   -0.261   -0.261
```

Model 2 Full Structural Model

```
mod2_cfa <- '
emlit =~ ealpha + ephono + erecep + eexpr
atprob =~ a1 + a2 + a3 + a4 + a5 + a6

emlit ~ atprob
emlit ~ meduc
decod ~ emlit

meduc ~~ atprob

'
```

```r
mod2_fitted <- cfa(mod2_cfa,
                   sample.cov = dicedata,
                   sample.nobs = 250)
summary(mod2_fitted, standardized = TRUE,
        fit.measures = TRUE)
```

```
## lavaan 0.6-9 ended normally after 80 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        26
##
##   Number of observations                           250
##
## Model Test User Model:
##
##   Test statistic                               162.731
##   Degrees of freedom                                52
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1964.741
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.942
##   Tucker-Lewis Index (TLI)                       0.926
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6235.327
##   Loglikelihood unrestricted model (H1)      -6153.961
##
##   Akaike (AIC)                               12522.653
##   Bayesian (BIC)                             12614.211
##   Sample-size adjusted Bayesian (BIC)        12531.789
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.092
##   90 Percent confidence interval - lower         0.076
##   90 Percent confidence interval - upper         0.109
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.044
##
## Parameter Estimates:
##
```

```
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit =~
##     ealpha            1.000                               10.682    0.727
##     ephono            0.530    0.051   10.383    0.000    5.665    0.720
##     erecep            1.081    0.108   10.039    0.000   11.550    0.694
##     eexpr             0.860    0.082   10.546    0.000    9.185    0.732
##   atprob =~
##     a1                1.000                                0.836    0.915
##     a2                0.951    0.039   24.401    0.000    0.795    0.919
##     a3                0.738    0.045   16.295    0.000    0.617    0.770
##     a4                1.045    0.043   24.316    0.000    0.874    0.918
##     a5                0.743    0.051   14.467    0.000    0.621    0.720
##     a6                0.818    0.051   15.981    0.000    0.684    0.762
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit ~
##     atprob           -3.602    0.792   -4.550    0.000   -0.282   -0.282
##     meduc             4.900    0.619    7.921    0.000    0.459    0.519
##   decod ~
##     emlit             0.502    0.051    9.867    0.000    5.362    0.682
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   atprob ~~
##     meduc            -0.256    0.064   -4.021    0.000   -0.307   -0.271
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ealpha          101.948   11.314    9.011    0.000  101.948    0.472
##    .ephono           29.882    3.284    9.100    0.000   29.882    0.482
##    .erecep          143.417   15.286    9.382    0.000  143.417    0.518
##    .eexpr            73.114    8.175    8.943    0.000   73.114    0.464
##    .a1                0.136    0.017    8.190    0.000    0.136    0.163
##    .a2                0.117    0.015    8.014    0.000    0.117    0.156
##    .a3                0.261    0.025   10.355    0.000    0.261    0.407
##    .a4                0.143    0.018    8.065    0.000    0.143    0.158
##    .a5                0.358    0.034   10.572    0.000    0.358    0.481
##    .a6                0.338    0.033   10.397    0.000    0.338    0.419
##    .decod            33.112    3.484    9.503    0.000   33.112    0.535
##     meduc             1.281    0.115   11.180    0.000    1.281    1.000
##    .emlit            65.220   11.112    5.869    0.000    0.572    0.572
##     atprob            0.699    0.075    9.371    0.000    1.000    1.000
```

```
model_performance(mod2_fitted)
```

```
## # Indices of model performance
##
## Chi2(52) | p (Chi2) | Baseline(66) | p (Baseline) |   GFI |  AGFI |   NFI |  NNFI |   CFI | RMSEA |
```

```
## -----------------------------------------------------------------------------------------------
## 162.731  |   < .001 |     1964.741 |       < .001 | 0.906 | 0.859 | 0.917 | 0.926 | 0.942 | 0.092 |
```

```
modindices(mod2_fitted, sort = TRUE,
           min = 10)
```

```
##       lhs op   rhs     mi    epc sepc.lv sepc.all sepc.nox
## 99     a5 ~~    a6 46.953  0.162   0.162    0.466    0.466
## 51 ephono ~~ eexpr 26.663 20.815  20.815    0.445    0.445
## 48 ealpha ~~ decod 25.857 24.241  24.241    0.417    0.417
## 84     a2 ~~    a3 18.144  0.061   0.061    0.348    0.348
## 58 ephono ~~ decod 10.403 -8.255  -8.255   -0.262   -0.262
## 87     a2 ~~    a6 10.182 -0.052  -0.052   -0.259   -0.259
```

Model 3 Attention to emergent literacy path set to 0

```
mod3_cfa <- '
emlit =~ ealpha + ephono + erecep + eexpr
atprob =~ a1 + a2 + a3 + a4 + a5 + a6

emlit ~ 0*atprob
emlit ~ meduc
decod ~ emlit

meduc ~~ atprob

'
mod3_fitted <- cfa(mod3_cfa,
                        sample.cov = dicedata,
                        sample.nobs = 250)
summary(mod3_fitted, standardized = TRUE,
        fit.measures = TRUE)
```

```
## lavaan 0.6-9 ended normally after 73 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        25
##
##   Number of observations                           250
##
## Model Test User Model:
##
##   Test statistic                               184.363
##   Degrees of freedom                                53
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1964.741
```

```
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.931
##   Tucker-Lewis Index (TLI)                       0.914
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6246.143
##   Loglikelihood unrestricted model (H1)      -6153.961
##
##   Akaike (AIC)                               12542.285
##   Bayesian (BIC)                             12630.322
##   Sample-size adjusted Bayesian (BIC)        12551.069
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.100
##   90 Percent confidence interval - lower         0.084
##   90 Percent confidence interval - upper         0.115
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.107
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit =~
##     ealpha            1.000                               10.648    0.724
##     ephono            0.537    0.052   10.384    0.000    5.718    0.726
##     erecep            1.087    0.109    9.980    0.000   11.574    0.696
##     eexpr             0.860    0.082   10.432    0.000    9.163    0.730
##   atprob =~
##     a1                1.000                                0.835    0.914
##     a2                0.953    0.039   24.376    0.000    0.796    0.920
##     a3                0.739    0.045   16.280    0.000    0.617    0.770
##     a4                1.047    0.043   24.278    0.000    0.874    0.918
##     a5                0.744    0.051   14.459    0.000    0.621    0.721
##     a6                0.818    0.051   15.892    0.000    0.683    0.760
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit ~
##     atprob            0.000                                0.000    0.000
##     meduc             5.609    0.642    8.738    0.000    0.527    0.596
```

```
##   decod ~
##     emlit             0.500    0.051    9.726    0.000    5.323    0.677
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   atprob ~~
##     meduc            -0.256    0.064   -4.020    0.000   -0.307   -0.271
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ealpha          102.662   11.449    8.967    0.000  102.662    0.475
##    .ephono           29.271    3.274    8.941    0.000   29.271    0.472
##    .erecep          142.860   15.351    9.306    0.000  142.860    0.516
##    .eexpr            73.521    8.269    8.891    0.000   73.521    0.467
##    .a1                0.138    0.017    8.216    0.000    0.138    0.165
##    .a2                0.115    0.015    7.963    0.000    0.115    0.154
##    .a3                0.261    0.025   10.351    0.000    0.261    0.407
##    .a4                0.142    0.018    8.023    0.000    0.142    0.157
##    .a5                0.358    0.034   10.568    0.000    0.358    0.481
##    .a6                0.340    0.033   10.402    0.000    0.340    0.422
##    .decod            33.523    3.531    9.493    0.000   33.523    0.542
##     meduc             1.281    0.115   11.180    0.000    1.281    1.000
##    .emlit            73.093   12.314    5.936    0.000    0.645    0.645
##     atprob            0.698    0.075    9.353    0.000    1.000    1.000
```

```r
modindices(mod3_fitted,
           sort = TRUE,
           min = 20)
```

```
##        lhs op    rhs     mi    epc sepc.lv sepc.all sepc.nox
## 99      a5 ~~     a6 47.206  0.163   0.163    0.467    0.467
## 48  ealpha ~~  decod 27.797 25.525  25.525    0.435    0.435
## 51  ephono ~~  eexpr 26.586 21.135  21.135    0.456    0.456
## 112  meduc  ~  emlit 20.644 -0.158  -1.686   -1.489   -1.489
## 11   emlit  ~ atprob 20.644 -3.579  -0.281   -0.281   -0.281
## 109 atprob  ~  emlit 20.644 -0.032  -0.404   -0.404   -0.404
## 105  emlit ~~ atprob 20.644 -2.314  -0.324   -0.324   -0.324
```

Model 4a Modification based on Modindices

```r
mod4a_cfa <- '
emlit =~ ealpha + ephono + erecep + eexpr
atprob =~ a1 + a2 + a3 + a4 + a5 + a6

emlit ~ atprob
emlit ~ meduc
decod ~ emlit

meduc ~~ atprob

#covariances between items on Attention Problem
```

```
a5 ~~ a6
a2 ~~ a3
a2 ~~ a6

'


mod4a_fitted <- cfa(mod4a_cfa,
                        sample.cov = dicedata,
                        sample.nobs = 250)
summary(mod4a_fitted, standardized = TRUE,
        fit.measures = TRUE)
```

```
## lavaan 0.6-9 ended normally after 85 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        29
##
##   Number of observations                           250
##
## Model Test User Model:
##
##   Test statistic                                95.854
##   Degrees of freedom                                49
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1964.741
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.975
##   Tucker-Lewis Index (TLI)                       0.967
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6201.888
##   Loglikelihood unrestricted model (H1)      -6153.961
##
##   Akaike (AIC)                               12461.776
##   Bayesian (BIC)                             12563.898
##   Sample-size adjusted Bayesian (BIC)        12471.966
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.062
##   90 Percent confidence interval - lower         0.043
##   90 Percent confidence interval - upper         0.080
##   P-value RMSEA <= 0.05                          0.139
```

```
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.038
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit =~
##     ealpha            1.000                               10.682    0.727
##     ephono            0.530    0.051   10.386    0.000    5.666    0.720
##     erecep            1.081    0.108   10.042    0.000   11.553    0.694
##     eexpr             0.860    0.082   10.547    0.000    9.184    0.732
##   atprob =~
##     a1                1.000                                0.841    0.920
##     a2                0.940    0.039   24.124    0.000    0.791    0.914
##     a3                0.705    0.047   15.061    0.000    0.593    0.740
##     a4                1.047    0.042   25.095    0.000    0.881    0.925
##     a5                0.716    0.052   13.768    0.000    0.602    0.699
##     a6                0.804    0.052   15.607    0.000    0.676    0.754
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit ~
##     atprob           -3.597    0.787   -4.571    0.000   -0.283   -0.283
##     meduc             4.907    0.618    7.937    0.000    0.459    0.520
##   decod ~
##     emlit             0.502    0.051    9.862    0.000    5.358    0.681
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   atprob ~~
##     meduc            -0.255    0.064   -3.971    0.000   -0.303   -0.268
##  .a5 ~~
##    .a6                0.155    0.027    5.629    0.000    0.155    0.426
##  .a2 ~~
##    .a3                0.058    0.016    3.565    0.000    0.058    0.306
##    .a6               -0.016    0.013   -1.201    0.230   -0.016   -0.078
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ealpha          101.934   11.312    9.011    0.000  101.934    0.472
##    .ephono           29.865    3.282    9.098    0.000   29.865    0.482
##    .erecep          143.345   15.281    9.381    0.000  143.345    0.518
##    .eexpr            73.121    8.175    8.944    0.000   73.121    0.464
##    .a1                0.128    0.017    7.652    0.000    0.128    0.153
##    .a2                0.123    0.016    7.736    0.000    0.123    0.165
##    .a3                0.290    0.028   10.218    0.000    0.290    0.452
##    .a4                0.131    0.018    7.369    0.000    0.131    0.144
```

```
##    .a5               0.381    0.036   10.576    0.000    0.381    0.512
##    .a6               0.347    0.034   10.127    0.000    0.347    0.431
##    .decod           33.150    3.486    9.508    0.000   33.150    0.536
##     meduc            1.281    0.115   11.180    0.000    1.281    1.000
##    .emlit           65.129   11.100    5.868    0.000    0.571    0.571
##     atprob           0.707    0.075    9.449    0.000    1.000    1.000
```

```
model_performance(mod4a_fitted)
```

```
## # Indices of model performance
##
## Chi2(49) | p (Chi2) | Baseline(66) | p (Baseline) |   GFI |  AGFI |   NFI |  NNFI |   CFI | RMSEA |
## ----------------------------------------------------------------------------------------------------
## 95.854   |   < .001 |     1964.741 |       < .001 | 0.942 | 0.907 | 0.951 | 0.967 | 0.975 | 0.062 |
```

```
anova(mod2_fitted, mod4a_fitted)
```

```
## Chi-Squared Difference Test
##
##              Df   AIC   BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## mod4a_fitted 49 12462 12564  95.854
## mod2_fitted  52 12523 12614 162.731     66.877       3  1.989e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 4b Adding direct Path from Attention Problems to Decoding Skill

```
mod4b_cfa <- '
emlit =~ ealpha + ephono + erecep + eexpr
atprob =~ a1 + a2 + a3 + a4 + a5 + a6

emlit ~ atprob
emlit ~ meduc
decod ~ emlit

meduc ~~ atprob

#direct path from attention problems to decoding
decod ~ atprob

'


mod4b_fitted <- cfa(mod4b_cfa,
                    sample.cov = dicedata,
                    sample.nobs = 250)
summary(mod4b_fitted, standardized = TRUE,
        fit.measures = TRUE)
```

```
## lavaan 0.6-9 ended normally after 87 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        27
##
##   Number of observations                           250
##
## Model Test User Model:
##
##   Test statistic                               162.125
##   Degrees of freedom                                51
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1964.741
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.941
##   Tucker-Lewis Index (TLI)                       0.924
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6235.023
##   Loglikelihood unrestricted model (H1)      -6153.961
##
##   Akaike (AIC)                               12524.047
##   Bayesian (BIC)                             12619.126
##   Sample-size adjusted Bayesian (BIC)        12533.534
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.093
##   90 Percent confidence interval - lower         0.077
##   90 Percent confidence interval - upper         0.110
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.043
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit =~
```

```
##     ealpha           1.000                               10.658    0.725
##     ephono           0.534    0.051   10.390    0.000     5.691    0.723
##     erecep           1.084    0.108   10.009    0.000    11.554    0.694
##     eexpr            0.864    0.082   10.532    0.000     9.209    0.734
##   atprob =~
##     a1               1.000                                0.836    0.915
##     a2               0.951    0.039   24.385    0.000     0.795    0.919
##     a3               0.738    0.045   16.298    0.000     0.617    0.770
##     a4               1.045    0.043   24.310    0.000     0.874    0.918
##     a5               0.743    0.051   14.473    0.000     0.621    0.720
##     a6               0.819    0.051   15.986    0.000     0.685    0.762
##
## Regressions:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   emlit ~
##     atprob          -3.463    0.805   -4.299    0.000   -0.272   -0.272
##     meduc            4.918    0.622    7.909    0.000    0.461    0.522
##   decod ~
##     emlit            0.486    0.055    8.831    0.000    5.175    0.658
##     atprob          -0.434    0.552   -0.786    0.432   -0.363   -0.046
##
## Covariances:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   atprob ~~
##     meduc           -0.256    0.064   -4.018    0.000   -0.306   -0.271
##
## Variances:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ealpha         102.458   11.375    9.007    0.000  102.458    0.474
##    .ephono          29.582    3.274    9.035    0.000   29.582    0.477
##    .erecep         143.317   15.312    9.360    0.000  143.317    0.518
##    .eexpr           72.675    8.173    8.892    0.000   72.675    0.462
##    .a1               0.137    0.017    8.193    0.000    0.137    0.163
##    .a2               0.117    0.015    8.019    0.000    0.117    0.156
##    .a3               0.261    0.025   10.354    0.000    0.261    0.407
##    .a4               0.143    0.018    8.064    0.000    0.143    0.158
##    .a5               0.358    0.034   10.571    0.000    0.358    0.481
##    .a6               0.338    0.033   10.395    0.000    0.338    0.419
##    .decod           33.400    3.493    9.561    0.000   33.400    0.540
##     meduc            1.281    0.115   11.180    0.000    1.281    1.000
##    .emlit           65.501   11.230    5.833    0.000    0.577    0.577
##     atprob           0.699    0.075    9.370    0.000    1.000    1.000
```

```
model_performance(mod4b_fitted)
```

```
## # Indices of model performance
##
## Chi2(51) | p (Chi2) | Baseline(66) | p (Baseline) |   GFI |  AGFI |   NFI |  NNFI |   CFI | RMSEA |
## -----------------------------------------------------------------------------------------------------
## 162.125  |   < .001 |     1964.741 |       < .001 | 0.906 | 0.857 | 0.917 | 0.924 | 0.941 | 0.093 |
```

```
anova(mod2_fitted, mod4b_fitted)
```

```
## Chi-Squared Difference Test
##
##              Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## mod4b_fitted 51 12524 12619 162.13
## mod2_fitted  52 12523 12614 162.73    0.60626       1     0.4362
```

“‘

References

Chan, Chung-hong, Geoffrey CH Chan, Thomas J. Leeper, and Jason Becker. 2021. *Rio: A Swiss-Army Knife for Data File i/o*.

Dice, Jaime L, and Paula Schwanenflugel. 2012. "A Structural Model of the Effects of Preschool Attention on Kindergarten Literacy." *Reading and Writing* 25 (9): 2205–22.

Miller, Amanda C, Douglas Fuchs, Lynn S Fuchs, Donald Compton, Devin Kearns, Wenjuan Zhang, Loulee Yen, Samuel Patton, and Danielle Peterson Kirchner. 2014. "Behavioral Attention: A Longitudinal Study of Whether and How It Influences the Development of Word Reading and Reading Comprehension Among at-Risk Readers." *Journal of Research on Educational Effectiveness* 7 (3): 232–49.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Rosseel, Yves. 2012. "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 1–36. https://www.jstatsoft.org/v48/i02/.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.