Example 9: Structural Equation Modeling with Latent Variables (or their Observed Variables)
(complete syntax and output available electronically for Mplus 8.8; partial for R Lavaan 0.6-12)

These data were adapted from my dissertation work (see references below) in which 152 adults age 63–87 years were measured on visual impairment (distance acuity and five degrees of contrast sensitivity), processing speed, divided visual attention, and selective visual attention (as measured by the Useful Field of View subtests for each), attentional search efficiency (DriverScan), and simulator driving impairment (as measured by six driving performance indicators).

Hoffman, L., Yang, X., Bovaird, J. A., & Embretson, S. E. (2006). Measuring attention in older adults: Development and psychometric evaluation of DriverScan. *Educational and Psychological Measurement*, *66*, 984-1000.

Hoffman, L., McDowd, J. M., Atchley, P., & Dubinsky R. A. (2005). The role of visual attention in predicting driving impairment in older adults. *Psychology and Aging*, *20*(4), 610-622.

This example will demonstrate how to estimate structural equation models, including models with mediation and latent variable interactions. But because simultaneous estimation of all effects of interest may not always be possible, this example will also show how to generate and use EAP factor score estimates instead. (For a version of this handout that also works with plausible values of factor scores, see Example9c from this previous class.)

Mplus Code to Read in Data:

```
TITLE:       SEM Example for Driverscan
DATA:        FILE = driverscanSEM.csv;     ! FILE is file to be analyzed
             FORMAT = free;                ! Free is default
             TYPE = INDIVIDUAL;            ! Individual data is default

VARIABLE:  ! Every variable in data set
      NAMES = PersonID sex age75 lncs15 lncs3 lncs6 lncs12 lncs18 far lnps
              lnda lnsa Dscan lane da_task crash stop speed time;
      ! Every variable in EACH MODEL
      USEVARIABLES = (to be changed for each model);
      IDVARIABLE = PersonID;        ! To keep ID variable for merging
      MISSING = ALL (-9999);        ! Value to denote missing values

ANALYSIS:    ESTIMATOR = MLR;   ! For continuous items whose residuals may not be normal

OUTPUT:      SAMPSTAT           ! Sample descriptives to verify data
             MODINDICES (3.84)  ! Cheat codes to improve model fit (at p<.05)
             STDYX              ! Requests fully standardized solution
             RESIDUAL           ! Requests standardized and normalized residuals
             SVALUES;           ! Write code with estimated parameters as start values
             TECH4;             ! Latent variable correlation matrix

SAVEDATA:    SAVE = FSCORES; FILE = FactorScores.dat;  ! Change .dat name by model
             MISSFLAG = 99;                            ! Missing data item indicator

MODEL:       ! (model syntax goes here, to be changed for each model)
```

We will begin by fitting single-factor measurement models for each latent factor. This is for two reasons: (1) we need to ensure each unidimensional factor fits its indicators, and (2) we will generate the EAP factor scores to use later to demonstrate how to include reliability-corrected factor scores as a replacement for latent variables.

Given MLR estimation, the EAP (expected a posteriori estimate) is the mean of the expected factor score distribution for each person. So anytime factor score SE>0 (and reliability is <1), this means the factor score still has error with it that we should correct for to avoid bias in the structural model parameters…

Measurement Model 1 for Visual Impairment (including Omega)

```
VARIABLE:  ! Every variable in THIS MODEL
             USEVARIABLES = lncs15 lncs3 lncs6 lncs12 lncs18 far;

MODEL:     ! Measurement model
  Vision BY far@1
        lncs15* lncs3* lncs6* lncs12* lncs18* (L2-L6);          ! 1 marker loading
  [far* lncs15* lncs3* lncs6* lncs12* lncs18*];                 ! All intercepts
   far* lncs15* lncs3* lncs6* lncs12* lncs18* (E1-E6);          ! Residual variances
  [Vision@0]; Vision* (Fvar);                                   ! Factor M=0, Var=?

MODEL CONSTRAINT: ! TO GET OMEGA
NEW(SumLoad2 SumError SumRCov Omega);
SumLoad2 =( 1+L2+L3+L4+L5+L6)**2;
SumError = E1+E2+E3+E4+E5+E6;
SumRCov = 2*(0);
! Omega = true variance / total variance
Omega = SumLoad2*Fvar / (SumLoad2*Fvar+SumError+SumRCov);
```

```
MODEL FIT INFORMATION
Number of Free Parameters                      18

Loglikelihood
          H0 Value                       -747.948
          H0 Scaling Correction Factor     1.1255
            for MLR
          H1 Value                       -739.282
          H1 Scaling Correction Factor     1.1171
            for MLR

Information Criteria
          Akaike (AIC)                   1531.897
          Bayesian (BIC)                 1586.327
          Sample-Size Adjusted BIC       1529.357
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                            15.752*
          Degrees of Freedom                    9
          P-Value                          0.0722
          Scaling Correction Factor        1.1003
            for MLR

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                          0.070
          90 Percent C.I.                   0.000  0.126
          Probability RMSEA <= .05          0.246

CFI/TLI
          CFI                               0.973
          TLI                               0.955

Chi-Square Test of Model Fit for the Baseline Model
          Value                           264.950
          Degrees of Freedom                   15
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)
          Value                             0.041
```

Measurement Model 1 for Vision:

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
VISION BY				
FAR	1.000	0.000	999.000	999.000
LNCS15	0.497	0.103	4.815	0.000
LNCS3	0.594	0.118	5.018	0.000
LNCS6	0.764	0.136	5.628	0.000
LNCS12	1.296	0.207	6.277	0.000
LNCS18	1.504	0.237	6.353	0.000
Means				
VISION	0.000	0.000	999.000	999.000
Intercepts				
LNCS15	-3.698	0.035	-105.136	0.000
LNCS3	-3.938	0.035	-113.273	0.000
LNCS6	-3.730	0.043	-87.639	0.000
LNCS12	-2.368	0.066	-36.000	0.000
LNCS18	-1.406	0.081	-17.389	0.000
FAR	3.026	0.067	45.130	0.000
Variances				
VISION	**0.224**	0.067	3.333	0.001
Residual Variances				
LNCS15	0.133	0.018	7.435	0.000
LNCS3	0.105	0.014	7.451	0.000
LNCS6	0.145	0.028	5.231	0.000
LNCS12	0.282	0.047	5.947	0.000
LNCS18	0.488	0.062	7.933	0.000
FAR	0.460	0.055	8.349	0.000
New/Additional Parameters				
SUMLOAD2	31.983	7.564	4.228	0.000
SUMERROR	1.613	0.102	15.822	0.000
SUMRCOV	0.000	0.000	0.000	1.000
OMEGA	0.816	0.024	33.851	0.000

STANDARDIZED MODEL RESULTS
STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
VISION BY				
FAR	0.572	0.062	9.190	0.000
LNCS15	0.541	0.074	7.305	0.000
LNCS3	0.656	0.062	10.605	0.000
LNCS6	0.688	0.057	12.062	0.000
LNCS12	0.756	0.051	14.815	0.000
LNCS18	0.713	0.041	17.293	0.000

For factor score reliability

SAMPLE STATISTICS FOR ESTIMATED FACTOR SCORES

Means

VISION	VISION_SE
0.000	**0.194**

Covariances

VISION	0.186

$$\rho = \frac{.224}{.224 + .194^2} = .856$$

Factor score reliability uses the factor variance as "true" and the SE^2 of the factor scores (given just above) as "error" (because these factor scores have error in them anytime reliability is < 1).

If we were going to sum the indicators, omega would have been used for reliability instead.

Normalized Residuals for Covariances/Correlations/Residual Correlations

	LNCS15	LNCS3	LNCS6	LNCS12	LNCS18
LNCS15	0.000				
LNCS3	1.651	0.000			
LNCS6	-0.045	0.261	0.000		
LNCS12	-0.455	-0.241	0.021	0.000	
LNCS18	-0.629	-0.458	-0.177	0.353	0.000
FAR	-0.471	-0.731	-0.062	0.198	0.558

Local fit looks good as well…

Measurement Model 2 for Driving Impairment (including Omega)

```
VARIABLE:   ! Every variable in THIS MODEL
              USEVARIABLES = lane da_task crash stop speed time;

MODEL:      ! Measurement model
  Driving BY crash@1
             da_task* lane* stop* speed* time* (L2-L6);    ! 1 marker loading
  [lane* da_task* crash* stop* speed* time*];              ! All intercepts
   lane* da_task* crash* stop* speed* time* (E1-E6);       ! Residual variances
  [Driving@0]; Driving* (Fvar);                            ! Factor M=0, Var=?
  speed WITH time* (ResCov);                               ! Residual covariance

MODEL CONSTRAINT: ! TO GET OMEGA
NEW(SumLoad2 SumError SumRCov Omega);
SumLoad2 = ( 1+L2+L3+L4+L5+L6)**2;
SumError =  E1+E2+E3+E4+E5+E6;
SumRCov = 2*(ResCov);
! Omega = true variance / total variance
Omega = SumLoad2*Fvar / (SumLoad2*Fvar+SumError+SumRCov);
```

```
*** WARNING
  Data set contains cases with missing on all variables.
  These cases were not included in the analysis.
  Number of cases with missing on all variables:  20
```

A total of 20 participants were unable to complete the simulator driving task, so they are not included in this model…

```
MODEL FIT INFORMATION
Number of Free Parameters                       19
Loglikelihood
          H0 Value                         -37.119
          H0 Scaling Correction Factor      1.1566
            for MLR
          H1 Value                         -30.710
          H1 Scaling Correction Factor      1.1108
            for MLR

Information Criteria
          Akaike (AIC)                     112.239
          Bayesian (BIC)                   167.012
          Sample-Size Adjusted BIC         106.915
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                             12.791*
          Degrees of Freedom                     8
          P-Value                           0.1192
          Scaling Correction Factor         1.0021
            for MLR

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.067
          90 Percent C.I.                    0.000  0.133
          Probability RMSEA <= .05           0.293

CFI/TLI
          CFI                                0.922
          TLI                                0.854

Chi-Square Test of Model Fit for the Baseline Model
          Value                             76.677
          Degrees of Freedom                    15
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)
          Value                              0.054
```

Measurement Model 2 for Driving:

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DRIVING BY				
CRASH	1.000	0.000	999.000	999.000
LANE	0.150	0.057	2.608	0.009
DA_TASK	0.173	0.074	2.348	0.019
STOP	0.347	0.163	2.124	0.034
SPEED	0.422	0.138	3.054	0.002
TIME	0.048	0.043	1.104	0.270
SPEED WITH				
TIME	-0.023	0.004	-5.393	0.000
Means				
DRIVING	0.000	0.000	999.000	999.000
Intercepts				
LANE	0.815	0.015	53.293	0.000
DA_TASK	0.256	0.013	20.102	0.000
CRASH	0.859	0.053	16.292	0.000
STOP	0.205	0.038	5.349	0.000
SPEED	0.836	0.042	19.687	0.000
TIME	3.146	0.009	349.081	0.000
Variances				
DRIVING	**0.159**	0.062	2.574	0.010
Residual Variances				
LANE	0.027	0.004	6.596	0.000
DA_TASK	0.017	0.004	4.613	0.000
CRASH	0.209	0.055	3.781	0.000
STOP	0.174	0.031	5.575	0.000
SPEED	0.210	0.028	7.391	0.000
TIME	0.010	0.001	8.639	0.000
New/Additional Parameters				
SUMLOAD2	4.578	1.185	3.865	0.000
SUMERROR	0.647	0.067	9.627	0.000
SUMRCOV	-0.046	0.009	-5.393	0.000
OMEGA	0.548	0.076	7.166	0.000

For factor score reliability

SAMPLE STATISTICS FOR ESTIMATED FACTOR SCORES

Means

DRIVING	DRIVING_SE
0.000	**0.247**

Covariances

	DRIVING
DRIVING	0.098

$$\rho = \frac{.159}{.159 + .247^2} = .723$$ Uh-oh…

Factor score reliability uses the factor variance as "true" and the SE^2 of the factor scores (given just above) as "error" (because these factor scores have error in them anytime reliability is < 1).

STANDARDIZED MODEL RESULTS
STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DRIVING BY				
CRASH	0.657	0.117	5.596	0.000
LANE	0.340	0.123	2.767	0.006
DA_TASK	0.470	0.132	3.576	0.000
STOP	0.315	0.115	2.748	0.006
SPEED	0.345	0.107	3.226	0.001
TIME	0.185	0.145	1.275	0.202
SPEED WITH				
TIME	-0.494	0.090	-5.478	0.000

Normalized Residuals for Covariances/Correlations/Residual Correlations

	LANE	DA_TASK	CRASH	STOP	SPEED
LANE	0.000				
DA_TASK	-0.487	0.000			
CRASH	0.359	-0.390	0.000		
STOP	0.769	0.503	-0.004	0.000	
SPEED	0.458	-0.836	0.471	-0.482	0.000
TIME	-1.508	**2.067**	-0.346	-0.545	0.000

Local fit looks mostly ok, with one exception…

Measurement Model 3 for Attentional Impairment (including Omega)

```
VARIABLE:   ! Every variable in THIS MODEL
              USEVARIABLES = lnda lnsa Dscan;

MODEL:      ! Measurement model
  Attn BY lnda@1
          lnsa* dscan* (L2-L3); ! 1 marker loading
  [lnda* lnsa* dscan*];         ! All intercepts
   lnda* lnsa* dscan* (E1-E3);  ! Residual variances
  [Attn@0]; Attn* (Fvar);       ! Factor M=0, Var=?

MODEL CONSTRAINT:  ! TO GET OMEGA
NEW(SumLoad2 SumError SumRCov Omega);
SumLoad2 = ( 1+L2+L3)**2;
SumError =  E1+E2+E3;
SumRCov = 2*(0);
! Omega = true variance / total variance
Omega = SumLoad2*Fvar / (SumLoad2*Fvar+SumError+SumRCov);
```

Can you guess why I didn't include the model fit results???

Measurement Model 3 for Attention:

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ATTN BY				
LNDA	1.000	0.000	999.000	999.000
LNSA	0.516	0.071	7.275	0.000
DSCAN	1.107	0.139	7.933	0.000
Means				
ATTN	0.000	0.000	999.000	999.000
Intercepts				
LNDA	4.354	0.079	54.825	0.000
LNSA	5.581	0.036	154.256	0.000
DSCAN	-0.012	0.081	-0.154	0.878
Variances				
ATTN	**0.443**	0.088	5.008	0.000
Residual Variances				
LNDA	0.516	0.068	7.597	0.000
LNSA	0.081	0.017	4.674	0.000
DSCAN	0.449	0.086	5.243	0.000
New/Additional Parameters				
SUMLOAD2	6.876	0.960	7.165	0.000
SUMERROR	1.045	0.102	10.212	0.000
SUMRCOV	0.000	0.000	0.000	1.000
OMEGA	0.745	0.038	19.728	0.000

STANDARDIZED MODEL RESULTS
STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ATTN BY				
LNDA	0.680	0.055	12.275	0.000
LNSA	0.770	0.055	14.087	0.000
DSCAN	0.740	0.056	13.153	0.000

For factor score reliability

SAMPLE STATISTICS FOR ESTIMATED FACTOR SCORES

Means

ATTN	ATTN_SE
0.000	**0.313**

Covariances

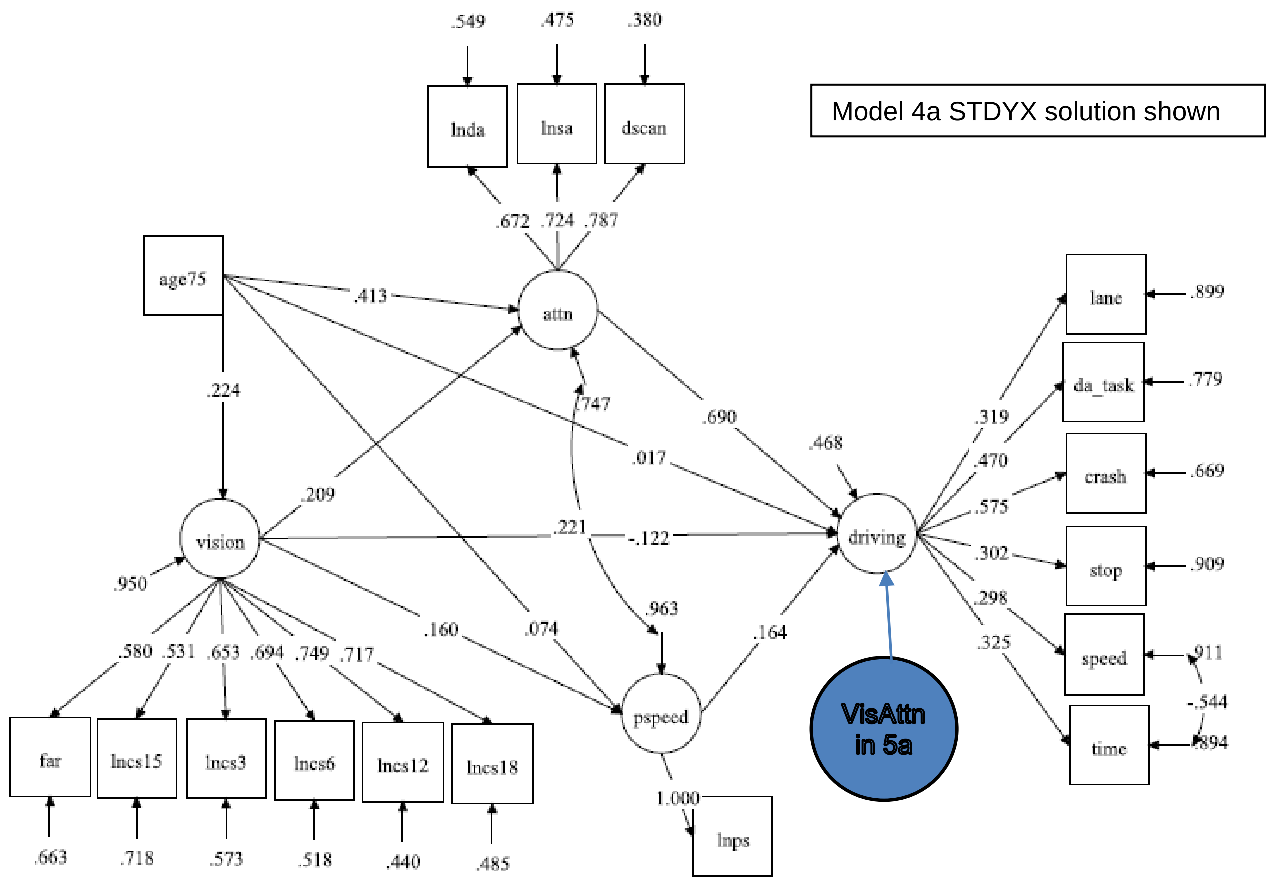
ATTN	0.345

$$\rho = \frac{.443}{.443 + .313^2} = .819$$

Factor score reliability uses the factor variance as "true" and the SE^2 of the factor scores (given just above) as "error" (because these factor scores have error in them anytime reliability is < 1).

Now we are ready to test the model of interest, **Model 4a** as shown below (drawn by Mplus, made prettier by me). We'll begin with a **saturated structural model** that has main effects of the latent variables only. This model uses directed arrows and covariances among the latent variables (but bivariate relations instead of unique relations will be provided by the model-estimated latent variable covariance matrix in the output).

```
VARIABLE:  ! Every variable in THIS MODEL
             USEVARIABLES = lncs15 lncs3 lncs6 lncs12 lncs18 far
                            lane da_task crash stop speed time
                            lnda lnsa Dscan age75 lnps;
```




```
MODEL:        ! Measurement models
  Vision BY far@1 lncs15* lncs3* lncs6* lncs12* lncs18*;   ! 1 marker loading
  [far* lncs15* lncs3* lncs6* lncs12* lncs18*];            ! All intercepts
   far* lncs15* lncs3* lncs6* lncs12* lncs18*;             ! Residual variances
  [Vision@0]; Vision*;                                     ! Factor M=0, Var=?

  Driving BY crash@1 da_task* lane* stop* speed* time*;    ! 1 marker loading
  [lane* da_task* crash* stop* speed* time*];              ! All intercepts
   lane* da_task* crash* stop* speed* time*;               ! Residual variances
  [Driving@0]; Driving*;                                   ! Factor M=0, Var=?
  speed WITH time* (ResCov);                               ! Residual covariance

 Attn BY lnda@1 lnsa* dscan*;                              ! 1 marker loading
  [lnda* lnsa* dscan*];                                    ! All intercepts
   lnda* lnsa* dscan*;                                     ! Residual variances
  [Attn@0]; Attn*;                                         ! Factor M=0, Var=?

 Pspeed BY lnps@1; lnps@0;          ! Bring processing speed into likelihood
  [lnps* Pspeed@0]; Pspeed*;        ! Move its variance to a factor, factor mean=0
```

```
! Structural model with all possible main effects
  Vision Attn Pspeed Driving ON Age75*  (Age1-Age4); ! Age --> outcomes
         Attn Pspeed Driving ON Vision* (Vis1-Vis3); ! Vision --> outcomes
  Attn WITH Pspeed*;                                  ! Res cov for Attn and Pspeed
  Driving ON Pspeed* Attn* (Speed1 Attn1);            ! Pspeed, Attn --> Driving

MODEL CONSTRAINT:  ! Example of how to request indirect effects
NEW(AgeVis AgeSpeed AgeAttn);
AgeVis   = Age1*Vis3;    ! Indirect effect of age to vision to driving
AgeSpeed = Age3*Speed1;  ! Indirect effect of age to processing speed to driving
AgeAttn  = Age2*Attn1;   ! Indirect effect of age to attention to driving
```

```
MODEL FIT INFORMATION
Number of Free Parameters                       58
Loglikelihood
          H0 Value                       -1310.811
          H0 Scaling Correction Factor      1.1063
            for MLR
          H1 Value                       -1238.221
          H1 Scaling Correction Factor      1.0405
            for MLR

Information Criteria
          Akaike (AIC)                    2737.622
          Bayesian (BIC)                  2913.007
          Sample-Size Adjusted BIC        2729.438
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                            144.331*
          Degrees of Freedom                   110
          P-Value                           0.0156
          Scaling Correction Factor         1.0059
            for MLR

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.045
          90 Percent C.I.                    0.021  0.064
          Probability RMSEA <= .05           0.635

CFI/TLI
          CFI                                0.936
          TLI                                0.921

SRMR (Standardized Root Mean Square Residual)
          Value                              0.063
```

Overall model fit is good enough according to RMSEA and SRMR (how much worse is our H_0 model than the perfect saturated H_1 model), but maybe a little lacking according to CFI and TLI (how much better is our H_0 model against the worst possible null model of no covariances).

But any misfit must be due to the cross-factor measurement model (i.e., covariances of indicators from different factors not predicted accurately) **because our structural model is saturated**—every possible direct relationship among the latent variables has been included.

```
UNSTANDARDIZED MODEL RESULTS (TRUNCATED FOR SPACE)
                                                 Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
MEASUREMENT MODEL RESULTS GIVEN FIRST (BY STATEMENTS)
 VISION   BY
    FAR                1.000      0.000    999.000    999.000
    LNCS15             0.481      0.099      4.837      0.000
    LNCS3              0.584      0.115      5.076      0.000
    LNCS6              0.759      0.136      5.583      0.000
    LNCS12             1.265      0.203      6.248      0.000
    LNCS18             1.491      0.232      6.416      0.000

DRIVING  BY
    CRASH              1.000      0.000    999.000    999.000
    LANE               0.161      0.066      2.444      0.015
    DA_TASK            0.197      0.065      3.022      0.003
    STOP               0.381      0.164      2.330      0.020
    SPEED              0.418      0.164      2.540      0.011
    TIME               0.097      0.053      1.819      0.069

ATTN     BY
    LNDA               1.000      0.000    999.000    999.000
    LNSA               0.491      0.061      8.000      0.000
    DSCAN              1.192      0.170      7.022      0.000

PSPEED   BY
    LNPS               1.000      0.000    999.000    999.000
```

```
REGRESSION PATHS GIVEN NEXT (ON STATEMENTS)
ATTN     ON
    VISION             0.287      0.137      2.095      0.036
PSPEED   ON
    VISION             0.167      0.100      1.658      0.097

DRIVING  ON
    VISION            -0.089      0.109     -0.814      0.415
    PSPEED             0.114      0.083      1.387      0.165
    ATTN               0.365      0.127      2.884      0.004

VISION   ON
    AGE75              0.024      0.011      2.187      0.029
ATTN     ON
    AGE75              0.059      0.014      4.393      0.000
PSPEED   ON
    AGE75              0.008      0.008      0.988      0.323
DRIVING  ON
    AGE75              0.001      0.011      0.119      0.905

COVARIANCES GIVEN LAST (WITH STATEMENTS)
ATTN     WITH
    PSPEED             0.061      0.027      2.292      0.022
SPEED    WITH
    TIME              -0.025      0.004     -5.512      0.000

INDIRECT EFFECTS REQUESTED USING MODEL CONSTRAINT
 New/Additional Parameters
    AGEVIS            -0.002      0.003     -0.830      0.406
    AGESPEED           0.001      0.001      0.764      0.445
    AGEATTN            0.022      0.009      2.507      0.012

STANDARDIZED MODEL RESULTS (TRUNCATED FOR SPACE)
STDYX Standardization
                                                        Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
VISION   BY
    FAR                0.580      0.062      9.424      0.000
    LNCS15             0.531      0.076      6.999      0.000
    LNCS3              0.653      0.061     10.646      0.000
    LNCS6              0.694      0.059     11.851      0.000
    LNCS12             0.749      0.051     14.647      0.000
    LNCS18             0.717      0.042     17.024      0.000
DRIVING  BY
    CRASH              0.575      0.107      5.378      0.000
    LANE               0.319      0.130      2.446      0.014
    DA_TASK            0.470      0.100      4.694      0.000
    STOP               0.302      0.115      2.630      0.009
    SPEED              0.298      0.102      2.911      0.004
    TIME               0.325      0.132      2.470      0.014
ATTN     BY
    LNDA               0.672      0.058     11.501      0.000
    LNSA               0.724      0.053     13.543      0.000
    DSCAN              0.787      0.045     17.608      0.000
PSPEED   BY
    LNPS               1.000      0.000    999.000    999.000
DRIVING  ON
    VISION            -0.122      0.148     -0.826      0.409
    PSPEED             0.164      0.120      1.368      0.171
    ATTN               0.690      0.149      4.617      0.000
PSPEED   ON
    VISION             0.160      0.094      1.715      0.086
ATTN     ON
    VISION             0.209      0.096      2.191      0.028
DRIVING  ON
    AGE75              0.017      0.148      0.118      0.906
VISION   ON
    AGE75              0.224      0.087      2.582      0.010
ATTN     ON
    AGE75              0.413      0.081      5.085      0.000
PSPEED   ON
    AGE75              0.074      0.075      0.986      0.324
ATTN     WITH
    PSPEED             0.221      0.088      2.523      0.012
 SPEED    WITH
    TIME              -0.544      0.090     -6.061      0.000
```

Left: The ON statements among the latent variables describe the standardized (correlation metric) unique relations of each latent predictor for the same latent outcome.

Below: The estimated latent variable correlation matrix describes the bivariate relations among the latent predictors and outcomes instead. It's useful to understand both types of relations in describing the results (that way you can differentiate what is not related *bivariately* from what is not related *any more* after controlling for something else).

```
ESTIMATED CORRELATION MATRIX
FOR LATENT VARIABLES
          VISION   DRIVING   ATTN   PSPEED
VISION     1.000
DRIVING    0.119    1.000
ATTN       0.302    0.705    1.000
PSPEED     0.177    0.331    0.270   1.000
AGE75      0.224    0.325    0.459   0.110
```

```
R-SQUARE
    Latent                                          Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value
    VISION             0.050      0.039      1.291      0.197
    DRIVING            0.532      0.151      3.526      0.000
    ATTN               0.253      0.077      3.264      0.001
    PSPEED             0.037      0.032      1.129      0.259
```

```
! Reduced structural model 4b (no age or vision --> driving)
  Vision Attn Pspeed ON Age75*  (Age2-Age4)      ! Age --> outcomes, not driving
         Attn Pspeed ON Vision* (Vis2-Vis3);     ! Vision --> outcomes, not driving
  Attn WITH Pspeed*;                             ! Res cov for Attn and Pspeed
  Driving ON Pspeed* Attn*  (Speed1 Attn1);      ! Pspeed, Attn --> Driving
```

```
MODEL FIT INFORMATION
Number of Free Parameters                       56
Loglikelihood
          H0 Value                       -1311.286
          H0 Scaling Correction Factor      1.0933
            for MLR
          H1 Value                       -1238.221
          H1 Scaling Correction Factor      1.0405
            for MLR
Information Criteria
          Akaike (AIC)                    2734.572
          Bayesian (BIC)                  2903.909
          Sample-Size Adjusted BIC        2726.670
            (n* = (n + 2) / 24)
Chi-Square Test of Model Fit
          Value                            144.090*
          Degrees of Freedom                   112
          P-Value                           0.0221
          Scaling Correction Factor         1.0142
            for MLR
RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.043
          90 Percent C.I.                    0.018  0.063
          Probability RMSEA <= .05           0.691
CFI/TLI
          CFI                                0.940
          TLI                                0.927
SRMR (Standardized Root Mean Square Residual)
          Value                              0.063
```

Did constraining these two structural paths to 0 make the model worse?
Rescaled $-2\Delta LL(2) = 0.646$, $p = .72$, so no

This model comparison is the appropriate way to test changes to the structural model, whose job is to reproduce the covariance among the latent factors and any observed predictors (but not among any observed predictors themselves).

Relying on good global model fit (which will mostly reflect the measurement models) is not sufficient to say a structural model fits. Instead, one should compare any overidentified structural model (with paths missing) to the saturated structural model to see if the fit is "not worse". One might compute a new version of the H1 model that reflects a saturated structural model (and a new null model that reflects an independent structural model) to be used in computing structural model fit indices…

We will continue with a saturated structural model in the model variants that follow…

What if we wanted to test a latent variable interaction? Model 5a (same measurement model as in Model 4a, including a full structural model with additions shown below)
Note that latent variable interactions can only be model predictors (and they cannot have covariances)
Latent variable interactions do not appear to be possible within R lavaan (or I couldn't find it if so)

```
ANALYSIS:   ESTIMATOR = MLR; ! For continuous items whose residuals may not be normal
            TYPE = RANDOM; ALGORITHM = INTEGRATION;   ! New estimation options needed
! Full structural model
  Vision Attn Pspeed Driving ON Age75*   (Age1-Age4);  ! Age --> outcomes
         Attn Pspeed Driving ON Vision* (Vis1-Vis3);   ! Vision --> outcomes
  Attn WITH Pspeed*;                                   ! Res cov for Attn and Pspeed
  Driving ON Pspeed* Attn*  (Speed1 Attn1);            ! Pspeed, Attn --> Driving

! Interaction between two latent variables (would be same if one variable was observed)
  VisAttn | Vision XWITH Attn;      ! VisAttn = new latent variable interaction
  Driving ON VisAttn* (VxA);        ! Latent variable interaction --> Driving

MODEL CONSTRAINT: ! Original latent factor variance of attn = .443, of vision = .224
NEW (V4low V4high A4low A4high);
  V4low =  Vis3  - VxA*SQRT(.443);  ! Vision slope for -1SD attn
  V4high = Vis3  + VxA*SQRT(.443);  ! Vision slope for +1SD attn
  A4low =  Attn1 - VxA*SQRT(.224);  ! Attn slope for -1SD vision
  A4high = Attn1 + VxA*SQRT(.224);  ! Attn slope for -+1SD vision
```

```
MODEL FIT INFORMATION
Number of Free Parameters                       59
Loglikelihood
          H0 Value                       -1310.261
          H0 Scaling Correction Factor      1.1066
            for MLR

Information Criteria
          Akaike (AIC)                    2738.522
          Bayesian (BIC)                  2916.931
          Sample-Size Adjusted BIC        2730.197
            (n* = (n + 2) / 24)
```

The absolute model fit indices have disappeared once we've used numeric integration (no H_1 saturated covariance matrix to come back to anymore). STDYX disappears for the same reason.

New structural model output only—note that the VisAttn interaction is related only to driving:

```
UNSTANDARDIZED MODEL RESULTS
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
ATTN       ON
    VISION             0.305      0.142      2.140      0.032
PSPEED     ON
    VISION             0.168      0.101      1.662      0.096
DRIVING    ON
    VISION            -0.106      0.114     -0.924      0.355  simple vision slope at attn=0
    PSPEED             0.118      0.083      1.423      0.155
    ATTN               0.363      0.130      2.785      0.005  simple attn slope at vision=0
    VISATTN            0.139      0.142      0.978      0.328  n.s. interaction

VISION     ON
    AGE75              0.024      0.011      2.188      0.029
ATTN       ON
    AGE75              0.059      0.014      4.399      0.000
PSPEED     ON
    AGE75              0.008      0.008      0.982      0.326
DRIVING    ON
    AGE75              0.002      0.011      0.135      0.892

ATTN     WITH
    PSPEED             0.060      0.027      2.222      0.026

New/Additional Parameters
    V4LOW             -0.198      0.167     -1.181      0.237  simple vision slope at attn=-1SD
    V4HIGH             0.013      0.126     -0.105      0.916  simple vision slope at attn=+1SD
    A4LOW              0.297      0.139      2.134      0.033  simple attn slope at vision=-1SD
    A4HIGH             0.428      0.153      2.793      0.005  simple attn slope at vision=+1SD

STDYX Standardization
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 ATTN       ON
    VISION             0.220      0.099      2.233      0.026

 PSPEED     ON
    VISION             0.160      0.093      1.720      0.085

 DRIVING    ON
    VISION            -0.145      0.155     -0.939      0.348
    PSPEED             0.170      0.120      1.417      0.157
    ATTN               0.692      0.152      4.564      0.000
    VISATTN            0.125      0.126      0.999      0.318

 VISION     ON
    AGE75              0.227      0.088      2.594      0.009

 ATTN       ON
    AGE75              0.413      0.081      5.071      0.000

 PSPEED     ON
    AGE75              0.074      0.075      0.981      0.327

 DRIVING    ON
    AGE75              0.020      0.151      0.133      0.894

 ATTN     WITH
    PSPEED             0.217      0.088      2.448      0.014
```

What would have happened if we used the mean of each person's factor score distribution from the single-factor models as observed constructs instead (i.e., replaced the latent circles with observed boxes)? Let's compare two possible ways of doing this—with or without reliability correction.

```
TITLE:  SEM Example for Driverscan using Single Factor Scores;
DATA:
  FILE = SEMfactorscores.csv;         ! EAP factor scores merged into original data
  TYPE = INDIVIDUAL; FORMAT = FREE;   ! Defaults
VARIABLE:
! List of ALL variables in data file
  NAMES = PersonID sex age75 lncs15 lncs3 lncs6 lncs12 lncs18 far lnps
          lnda lnsa Dscan lane da_task crash stop speed time
          VisFact DrivFact AttnFact; ! New factor scores
! Variables to be analyzed in this model
  USEVARIABLE = age75 lnps VisFact DrivFact AttnFact;
! Missing data identifier
  MISSING = ALL (-9999);
! ID variable;
  IDVARIABLE = PersonID;

ANALYSIS:    ESTIMATOR = MLR;
             TYPE = RANDOM; ALGORITHM = INTEGRATION;   ! New estimation options for latent interaction
OUTPUT:      STDYX RESIDUAL;     ! Standardized model, local fit
             SAMPSTAT;           ! Get descriptive stats for variables
```

Model 5b: Using Reliability-Corrected Single Factor Scores (and Latent Interaction)

```
MODEL:
! Measurement models for "factors" (factor mean=0 used to do centering)
! "Res" labels used to Incorporate factor score unreliability
  Vision  BY VisFact@1;  Vision*;  VisFact*  (ResVis);  [Vision@0  VisFact*];
  Attn    BY AttnFact@1; Attn*;    AttnFact* (ResAttn); [Attn@0    AttnFact*];
  Pspeed  BY lnps@1;     Pspeed*;  lnps*     (ResPspd); [Pspeed@0  lnps*];
  Driving BY DrivFact@1; Driving*; DrivFact* (ResDriv); [Driving@0 DrivFact*];
  VisAttn | Vision XWITH Attn; ! Latent interaction term (to address unreliability)

! Structural model among "factors"
  Vision Attn Pspeed Driving ON Age75*  (Age1-Age4);  ! Age --> outcomes
         Attn Pspeed Driving ON Vision* (Vis1-Vis3);  ! Vision --> outcomes
  Attn WITH Pspeed*;                                  ! Res cov for Attn and Pspeed
  Driving ON Pspeed* Attn* (Speed1 Attn1);            ! Pspeed, Attn --> Driving
  Driving ON VisAttn* (VxA);                          ! Interaction --> Driving

MODEL CONSTRAINT: ! Factor score variance of attn = .345, of vision = .186
NEW (V4low V4high A4low A4high);
  V4low =  Vis3  - VxA*SQRT(.345);  ! Vision slope for -1SD attn
  V4high = Vis3  + VxA*SQRT(.345);  ! Vision slope for +1SD attn
  A4low =  Attn1 - VxA*SQRT(.186);  ! Attn slope for -1SD vision
  A4high = Attn1 + VxA*SQRT(.186);  ! Attn slope for -+1SD vision

! (1-Reliability)*(factorvar+(SE*SE)) to fix residual variances to "error" variance
  ResVis =(1-.856)*(0.224+(.194*.194));
  ResAttn=(1-.819)*(0.443+(.313*.313));
  ResPspd=0; ! Processing speed assumed perfectly reliable
  ResDriv=(1-.723)*(0.159+(.247*.247));
! Processing speed assumed perfectly reliable
```

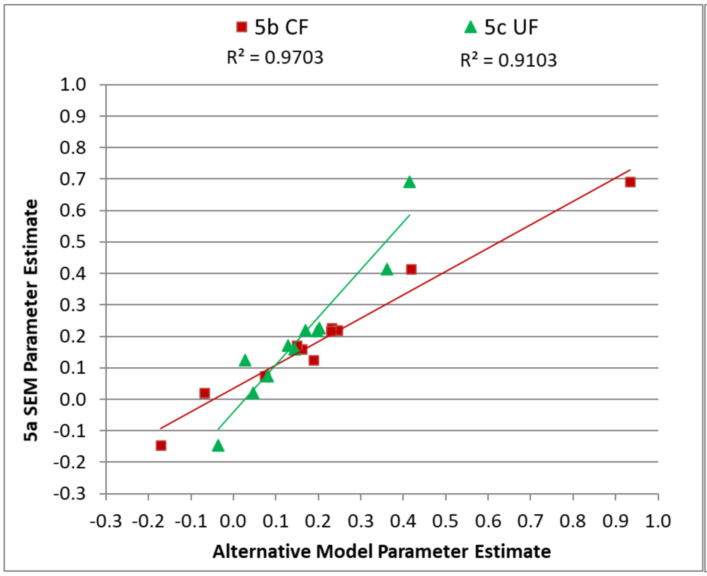
Model 5c: Using Uncorrected Single Factor Scores (Reliability=1 for all; changes to code below)

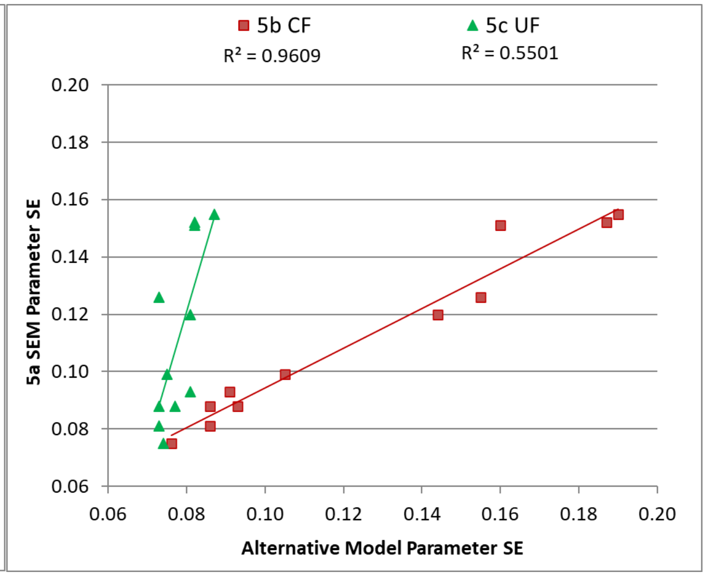
```
VARIABLE: ! Variables to be analyzed in this model
  USEVARIABLE = age75 lnps VisFact DrivFact AttnFact VisAttn;

DEFINE:   VisAttn = VisFact * AttnFact; ! Interaction is now an observed variable instead of latent
ANALYSIS: ESTIMATOR = MLR; ! Integration no longer needed

MODEL: ! All measurement and structural model code is the same as 5b after removing latent interaction

!VisAttn | Vision XWITH Attn; ! Latent interaction term removed (is now observed)

MODEL CONSTRAINT:
! Residual variances as "error" variances now ALL fixed to 0
  ResVis=0;
  ResAttn=0;
  ResPspd=0;
  ResDriv=0;
```

Model fit is acceptable for Model 5c (DF=3), but not available for Model 5b (given latent interaction)

What about the results? Let's compare the standardized solution across our 3 options:

	Estimates			Standard Errors			P-Values		
MODEL	5a SEM	5b CF	5c UF	5a SEM	5b CF	5c UF	5a SEM	5b CF	5c UF
Age -->									
VISION	.227	.232	.203	.088	.086	.077	**.009**	**.007**	**.008**
ATTN	.413	.418	.362	.081	.086	.073	**.000**	**.000**	**.000**
PSPEED	.074	.073	.081	.075	.076	.074	.327	.337	.275
DRIVING	.020	-.069	.046	.151	.160	.082	.894	.665	.576
Vision -->									
PSPEED	.160	.162	.144	.093	.091	.081	.085	.076	.077
ATTN	.220	.246	.170	.099	.105	.075	**.026**	**.019**	**.022**
ATTN<-->PSPEED	.217	.230	.198	.088	.093	.073	**.014**	**.014**	**.007**
DRIVING <--									
PSPEED	.170	.150	.129	.120	.144	.081	.157	.299	.110
VISION	-.145	-.172	-.035	.155	.190	.087	.348	.364	.686
ATTN	.692	.934	.415	.152	.187	.082	**.000**	**.000**	**.000**
VISATTN	.125	.189	.028	.126	.155	.073	.318	.223	.705
R2 Latent Variable									
VISION	.052	.054	.041	.040	.040	.031	.195	.179	.186
ATTN	.260	.283	.185	.081	.091	.060	**.001**	**.002**	**.002**
PSPEED	.037	.037	.032	.032	.032	.027	.258	.245	.237
DRIVING	.551	.872	.226	.147	.248	.061	**.000**	**.000**	**.000**





From our informal comparison of methods, it looks like reliability-corrected version (model 5b) of the full SEM model 5a appears to do a better job of reproducing parameter estimates (left figure) and standard errors (right figure) than the uncorrected version (model 5c). Note that a single estimate of reliability cannot be used as demonstrated here when factors are created using IRT/IFA, in which reliability is trait-specific instead (although it may be possible to trick Mplus into doing so, I'm not aware of any work on this).

For an example SEM results section, see Hoffman et al. (2005) reference given on page 1.