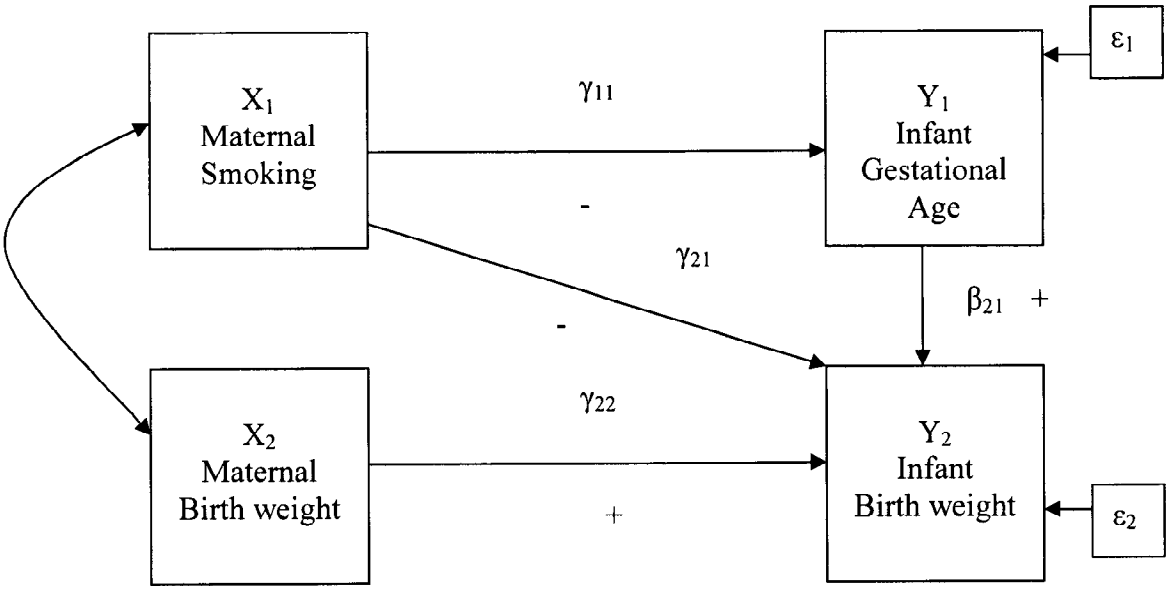
Problem Set 1
Psychosocial Stats II

Part A: Causality

1. $Y_1 = \gamma_{11}X_1 + \varepsilon_1$
   $Y_2 = \gamma_{22}X_2 + \gamma_{21}X_1 + \beta_{21}Y_1 + \varepsilon_2$



2. a. Y_1, Y_2 are endogenous
   b. X_1, X_2 are exogenous
3. For this example, I chose outcomes of infant gestational age and weight at birth. Maternal smoking (X_1) has both direct and indirect negative effects on birthweight because it is associated both with earlier delivery and intrauterine growth restriction. Maternal birth weight also impacts birth weight because anthropometric features are heritable in part, they will be positively related.
4. Removing Y_1 would violate the assumption of pseudo-isolation because it is a correctly specified intervening variable. Removing this indirect effect mispecifies the direct effect (γ_{21}) as a total effect, which underestimates the absolute effect of X_1 in this case since both paths from X_1 to Y_2 are in the same direction.
5. a. If Y_1 is omitted from the model, the covariance between X_1 and ε_2 becomes non-zero because Y_2 now has some unexplained error variance due to X_1.

   Assumed model: $Y_2 = \gamma_{22}X_2 + \gamma_{21}X_1 + \varepsilon_2$
   Which implies: $\varepsilon_2^* = \beta_{21}Y_1 + \varepsilon_2$
   $\text{Cov}(X_1, \varepsilon_2^*) = \text{Cov}(X_1, \beta_{21}Y_1 + \varepsilon_2)$
   $= \text{Cov}(X_1, \beta_{21}(\gamma_{11}X_1 + \varepsilon_1) + \varepsilon_2)$
   $= \beta_{21}\gamma_{11}\text{Var}(X_1) + \beta_{21}\text{Cov}(X_1, \varepsilon_1) + \text{Cov}(X_1, \varepsilon_2)$
   $= \beta_{21}\gamma_{11}\text{Var}(X_1)$

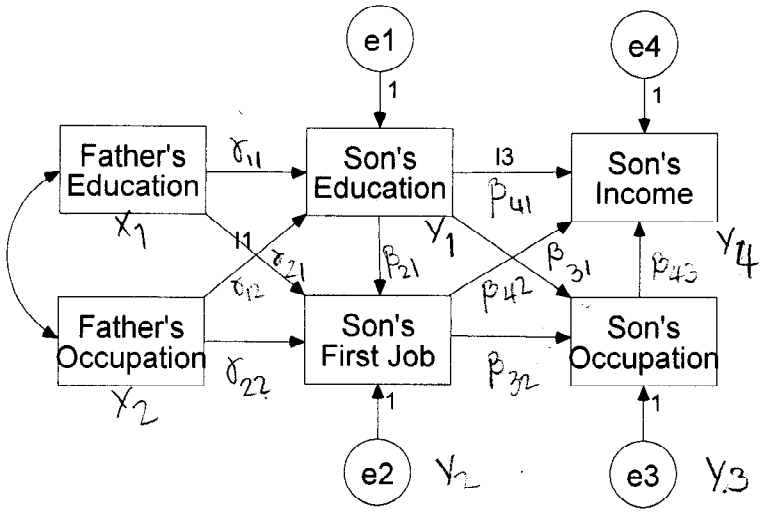
   b. If Y_1 is omitted from the model, the covariance between X_2 and ε_2 will depend on the covariance or correlation between X_2 and X_1.

$$\begin{aligned}\text{Cov}(X_2, \varepsilon_2^*) &= \text{Cov}(X_2, \beta_{21}Y_1 + \varepsilon_2) \\ &= \text{Cov}(X_2, \beta_{21}(\gamma_{11}X_1 + \varepsilon_1) + \varepsilon_2) \\ &= \beta_{21}\gamma_{11}\text{Cov}(X_2, X_1) + \beta_{21}\text{Cov}(X_2, \varepsilon_1) + \text{Cov}(X_2, \varepsilon_2) \\ &= \beta_{21}\gamma_{11}\text{Cov}(X_2, X_1)\end{aligned}$$

c. Removing an essential variable that contributes to a model and is associated with at least two variables (i.e. could be a confounder, common source, intervening variable, etc.) is a violation of the isolation assumption, that is the specified model is accurate and 'isolated' from any other variable such that the unexplained error of an endogenous variable does not covary with any predictor variable. In this example, both exogenous variables will still have a residual covariance with the unobserved error of Y2 if Y1 were removed, although the greatest covariance will be for X1 and E2. This will result in misrepresented gamma coefficients and a mispecified/ 'un-isolated' model.

Part B: Path Analysis

1.



Estimable Parameters: 17 (2 variances of exogenous variables, 4 error variances of endogenous variables, 10 direct effects, 1 double-headed arrow, correlation of exogenous variables)
Observed Parameters: 21 (6 variables*6+1variables)/2
According to the T-rule, which is a necessary but not sufficient rule for identification, the model may be identified because the number of estimable parameters is less than the number of observed parameters. A just-identified recursive model would include 4 additional paths 1) a direct effect from father's education to son's income 2) a direct effect from father's education to son's occupation 3) a direct effect from father's occupation to son's income 4) a direct effect from father's occupation to son's occupation.

3. Fathed = X1 Fathocc = X2 Grade = Y1 Firstjob = Y2 Jobnow= Y3 Income = Y4

Grade = γ_{11}Fathed + γ_{12}Fathocc + e1

Firstjob = γ_{21}Fathed + γ_{22}Fathocc + β_{21}Grade + e2

Jobnow = β_{31}Grade + β_{32}Firstjob + e3

Income = β_{41}Grade + β_{42}Firstjob + β_{43}Jobnow + e4

Path Formulas

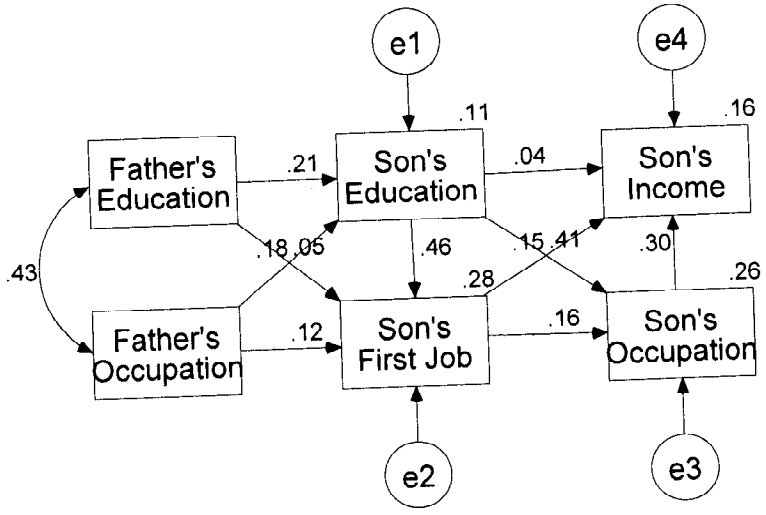
Exogenous variable: Father's Education (X1)

Endogenous Variable	**Direct Effect**	**Indirect Effect**	**Indirect Effect via Curved Arrow**
Grade	γ_{11}	--	r(X1, X2)* γ_{12}
Firstjob	γ_{21}	γ_{11}*β_{21}	r(X1, X2)*(γ_{22} + γ_{12}* β_{21})
Jobnow	--	$\gamma_{11}(\beta_{31}+ \beta_{21}*\beta_{32}) + \gamma_{21}*\beta_{32}$	r(X1, X2)*[γ_{22}*β_{32} + $\gamma_{12}(\beta_{31}+ \beta_{21}*\beta_{32})$]
Income	--	$\gamma_{11}[\beta_{41}+ \beta_{31}*\beta_{43}+ \beta_{21}(\beta_{42} +\beta_{32}*\beta_{43})] + \gamma_{21}(\beta_{42}+ \beta_{32}*\beta_{43})$	r(X1, X2)*[γ_{12}*(β_{41}+ β_{31}*β_{43}+ $\beta_{21}(\beta_{42}$ +β_{32}*β_{43}))+ γ_{22}*(β_{42}+ β_{32}*β_{43})]

Exogenous variable: Father's Occupation (X2)

Endogenous Variable	**Direct Effect**	**Indirect Effect**	**Indirect Effect via Curved Arrow**
Grade	γ_{12}	--	r(X2, X1)* γ_{11}
Firstjob	γ_{22}	γ_{12}*β_{21}	r(X2, X1)*(γ_{21} + γ_{11}* β_{21})
Jobnow	--	$\gamma_{12}(\beta_{31}+ \beta_{21}*\beta_{32}) + \gamma_{22}*\beta_{32}$	r(X2, X1)*[γ_{21}*β_{32} + $\gamma_{11}(\beta_{31}+ \beta_{21}*\beta_{32})$]
Income	--	$\gamma_{12}[\beta_{41}+ \beta_{31}*\beta_{43}+ \beta_{21}(\beta_{42} +\beta_{32}*\beta_{43})] + \gamma_{22}(\beta_{42}+ \beta_{32}*\beta_{43})$	r(X2, X1)*[γ_{11}*(β_{41}+ β_{31}*β_{43}+ $\beta_{21}(\beta_{42}$ +β_{32}*β_{43}))+ γ_{21}*(β_{42}+ β_{32}*β_{43})]

4. Standardized Coefficients



Maximum Likelihood Estimates

Regression Weights:

	Estimate	S.E.	C.R.	Label
GRADE <--- FATHED	0.138	0.030	4.638	par-3
GRADE <--- FATHOCC	0.017	0.004	3.940	par-10
FIRSTJOB <--- FATHOCC	0.118	0.043	2.742	par-4
FIRSTJOB <--- GRADE	5.046	0.437	11.559	par-7
FIRSTJOB <--- FATHED	0.364	0.299	1.217	I1
JOBNOW <--- FIRSTJOB	0.151	0.043	3.540	par-6
JOBNOW <--- GRADE	4.332	0.467	9.268	par-9
INCOME <--- JOBNOW	0.036	0.006	6.375	par-5
INCOME <--- GRADE	0.046	0.065	0.705	I3
INCOME <--- FIRSTJOB	0.018	0.006	3.225	par-11

Standardized Regression Weights:

	Estimate
GRADE <--- FATHED	0.214
GRADE <--- FATHOCC	0.182
FIRSTJOB <--- FATHOCC	0.116
FIRSTJOB <--- GRADE	0.461
FIRSTJOB <--- FATHED	0.052
JOBNOW <--- FIRSTJOB	0.156
JOBNOW <--- GRADE	0.409
INCOME <--- JOBNOW	0.299
INCOME <--- GRADE	0.036
INCOME <--- FIRSTJOB	0.153

Covariances:

	Estimate	S.E.	C.R.	Label
FATHED <--> FATHOCC	35.124	3.895	9.018	par-8

Correlations:

	Estimate
FATHED <--> FATHOCC	0.434

Variances:

	Estimate	S.E.	C.R.	Label
FATHED	11.704	0.731	16.016	par-12
FATHOCC	559.469	34.933	16.016	par-13
e1	4.284	0.267	16.016	par-14
e2	418.887	26.155	16.016	par-15
e3	401.617	25.077	16.016	par-16
e4	6.698	0.418	16.016	par-17

Squared Multiple Correlations:

	Estimate
GRADE	0.113
FIRSTJOB	0.277
JOBNOW	0.257
INCOME	0.163

Total Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.017	0.138	0.000	0.000	0.000
FIRSTJOB	0.203	1.058	5.046	0.000	0.000
JOBNOW	0.104	0.756	5.094	0.151	0.000
INCOME	0.008	0.053	0.322	0.023	0.036

Standardized Total Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.182	0.214	0.000	0.000	0.000
FIRSTJOB	0.200	0.150	0.461	0.000	0.000
JOBNOW	0.106	0.111	0.481	0.156	0.000
INCOME	0.069	0.064	0.250	0.200	0.299

Direct Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.017	0.138	0.000	0.000	0.000
FIRSTJOB	0.118	0.364	5.046	0.000	0.000
JOBNOW	0.000	0.000	4.332	0.151	0.000
INCOME	0.000	0.000	0.046	0.018	0.036

Standardized Direct Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.182	0.214	0.000	0.000	0.000
FIRSTJOB	0.116	0.052	0.461	0.000	0.000
JOBNOW	0.000	0.000	0.409	0.156	0.000
INCOME	0.000	0.000	0.036	0.153	0.299

Indirect Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.000	0.000	0.000	0.000	0.000
FIRSTJOB	0.085	0.694	0.000	0.000	0.000
JOBNOW	0.104	0.756	0.762	0.000	0.000
INCOME	0.008	0.053	0.276	0.005	0.000

Standardized Indirect Effects

	FATHOCC	FATHED	GRADE	FIRSTJOB	JOBNOW
GRADE	0.000	0.000	0.000	0.000	0.000
FIRSTJOB	0.084	0.099	0.000	0.000	0.000
JOBNOW	0.106	0.111	0.072	0.000	0.000
INCOME	0.069	0.064	0.214	0.047	0.000

Summary of models

Model	NPAR	CMIN	DF	P	CMIN/DF
Default Model	17	3.977	4	0.409	0.994
Model Number 2	15	5.952	6	0.429	0.992
Saturated model	21	0.000	0		
Independence model	6	582.593	15	0.000	38.840

Model	RMR	GFI	AGFI	PGFI
Default Model	6.110	0.997	0.987	0.190
Model Number 2	6.179	0.996	0.987	0.285
Saturated model	0.000	1.000		
Independence model	62.260	0.665	0.531	0.475

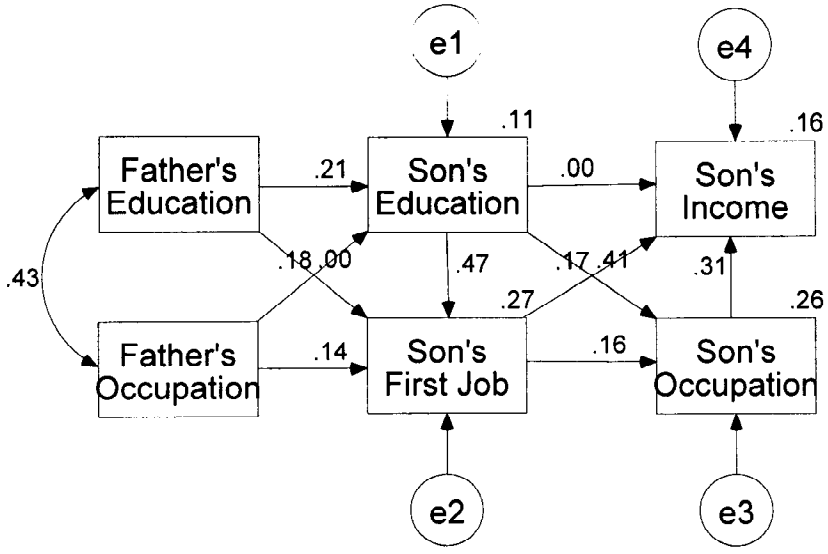
Model	AIC	BCC	BIC	CAIC
Default Model	37.977	38.447	140.554	127.094
Model Number 2	35.952	36.367	126.462	114.585
Saturated model	42.000	42.581	168.714	152.087
Independence model	594.593	594.759	630.797	626.046

The Model (Default Model) appears to fit the data well. The CMIN chi-square value not significantly different from the saturated model indicates little discrepancy between the observed and estimated variance-covariance matrix. The GFI is very close to 1 indicating near perfect fit relative to a saturated model. The AIC, BCC, BIC, and CAIC values are all lower than the saturated model which also indicate good fit and parsimony. The BIC and CAIC may be better measures of goodness of fit for these data because the sample size is relatively large (n=514).

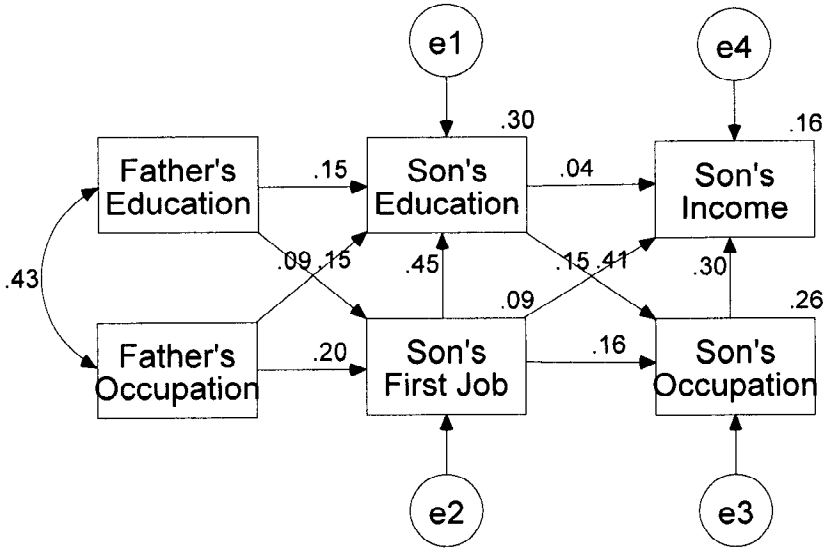
5. Indirect effect of Father's Education (X1) on First Job (Y2) with Unstandardized Coefficients
Indirect Effect
$\gamma_{11}*\beta_{21} = .138*5.046 = .69$ (this is shown in AMOS output)
Indirect Effect via curved arrow
$r(X1, X2)*(\gamma_{22} + \gamma_{12}* \beta_{21}) = .434*(.118+.017*5.046) = .09$
Total Indirect Effect = .78
Interpretation: For a one unit increase in paternal education, the occupational status of son's first job increases by .78 units through other than direct paths (intervening variables or curved arrows).

Indirect effect of Father's Occupation (X2) on Job now (Y3) with Unstandardized Coefficients
Indirect Effect
$\gamma_{12}(\beta_{31}+ \beta_{21}*\beta_{32}) + \gamma_{22}*\beta_{32} = .017(4.332+5.046*.151) + .118*.151 = .104$ (shown in output)
Indirect Effect via curved arrow
$r(X2, X1)*[\gamma_{21}*\beta_{32} + \gamma_{11}(\beta_{31}+ \beta_{21}*\beta_{32})] = .434[.364*.151 + .138(4.332+5.046*.151)] = .329$
Total Indirect Effect = .43
Interpretation: For a one unit increase in paternal occupational status, the son's eventual occupational status increases by .43 units through multiple indirect pathways, including its correlation with paternal education (curved arrow).

6. The coefficients from father's education to son's first job (γ_{21}) and from son's education to income (β_{41}) were not significant. Removing these non-significant paths did not considerably change other path coefficients. The CMIN and GFI values were not considerably different the original model, but AIC, BCC, BIC, and CAIC were lower indicating a slightly better fitting and more parsimonious model (Model Number 2 in preceding output). The differences are fairly marginal, except possibly for BIC and CAIC which assign a greater penality for complexity and thus has a greater tendency to reward parsimony.



7.



Summary of models

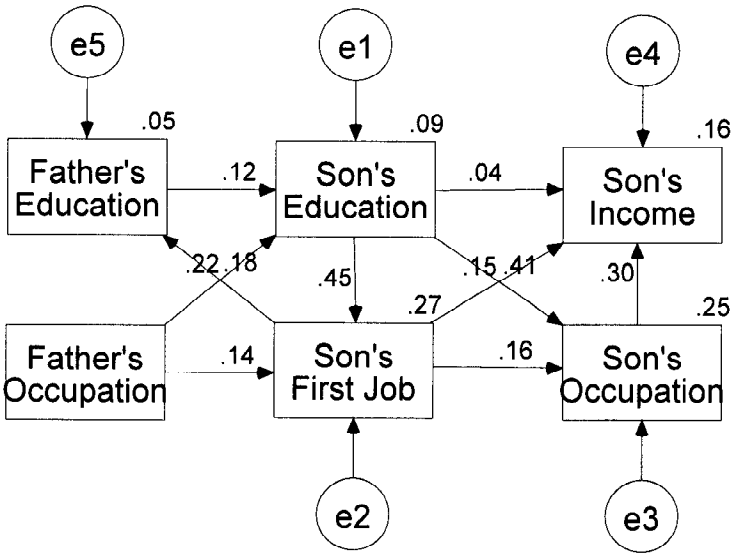
Model	NPAR	CMIN	DF	P	CMIN/DF
Q7 Model	17	3.977	4	0.409	0.994
Saturated model	21	0.000	0		
Independence model	6	582.593	15	0.000	38.840

Model	RMR	GFI	AGFI	PGFI
Q7 Model	6.110	0.997	0.987	0.190
Saturated model	0.000	1.000		
Independence model	62.260	0.665	0.531	0.475

Model	AIC	BCC	BIC	CAIC
Q7 Model	37.977	38.447	140.554	127.094
Saturated model	42.000	42.581	168.714	152.087
Independence model	594.593	594.759	630.797	626.046

Without greater information regarding the temporal sequence of first job and educational attainment, the interpretability of this model, in which son's first job is predicting son's educational attainment, may be as satisfactory as the original reversed model. However, assuming that first job refers to first job post-educational completion, the model would not make much sense. The fit of this model is exactly the same as the original model with the causal arrow pointing in the reverse direction, which highlights the necessity for strong a priori knowledge in specifying causal direction. SEM and path analysis should not be used for exploratory cross-sectional data analysis. The path coefficients that predict son's first job and son's education are also now reversed relative to the original model.

8.



Maximum Likelihood Estimates

Regression Weights:

	Estimate	S.E.	C.R.	Label
FIRSTJOB <----- FATHOCC	0.143	0.039	3.637	par-2
GRADE <-------- FATHOCC	0.021	0.004	5.317	par-8
JOBNOW <------ FIRSTJOB	0.151	0.043	3.541	par-5
JOBNOW <---------- GRADE	4.332	0.470	9.222	par-7
INCOME <-------- JOBNOW	0.036	0.006	6.375	par-3
INCOME <---------- GRADE	0.046	0.066	0.702	par-4
INCOME <------ FIRSTJOB	0.018	0.006	3.226	par-9
GRADE <---------- FATHED	0.078	0.030	2.585	par-1
FIRSTJOB <-------- GRADE	4.959	0.434	11.432	par-6
FATHED <------ FIRSTJOB	0.026	0.007	3.854	par-10

Standardized Regression Weights:

	Estimate
FIRSTJOB <----- FATHOCC	0.141
GRADE <-------- FATHOCC	0.225
JOBNOW <------ FIRSTJOB	0.156
JOBNOW <---------- GRADE	0.406
INCOME <-------- JOBNOW	0.298

INCOME <--------- GRADE	0.035
INCOME <------ FIRSTJOB	0.152
GRADE <--------- FATHED	0.122
FIRSTJOB <------- GRADE	0.450
FATHED <------ FIRSTJOB	0.184

Variances:

	Estimate	S.E.	C.R.	Label
FATHOCC	559.469	34.933	16.016	par-11
e1	4.318	0.272	15.889	par-12
e2	420.264	26.252	16.009	par-13
e5	11.079	0.694	15.958	par-14
e3	401.617	25.077	16.016	par-15
e4	6.698	0.418	16.016	par-16

Squared Multiple Correlations:

	Estimate
FIRSTJOB	0.267
FATHED	0.053
GRADE	0.087
JOBNOW	0.253
INCOME	0.162

Stability index for the following variables is 0.047

FIRSTJOB
FATHED
GRADE

Total Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.248	0.010	0.389	5.010	0.000
FATHED	0.007	0.027	0.010	0.132	0.000
GRADE	0.021	0.002	0.079	0.010	0.000
JOBNOW	0.129	0.162	0.399	5.133	0.000
INCOME	0.010	0.024	0.025	0.323	0.036

Standardized Total Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.245	0.010	0.056	0.455	0.000
FATHED	0.045	0.186	0.010	0.084	0.000
GRADE	0.230	0.023	0.124	0.010	0.000
JOBNOW	0.132	0.167	0.059	0.481	0.000
INCOME	0.085	0.204	0.030	0.249	0.298

Direct Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.143	0.000	0.000	4.959	0.000
FATHED	0.000	0.026	0.000	0.000	0.000
GRADE	0.021	0.000	0.078	0.000	0.000
JOBNOW	0.000	0.151	0.000	4.332	0.000
INCOME	0.000	0.018	0.000	0.046	0.036

Standardized Direct Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.141	0.000	0.000	0.450	0.000
FATHED	0.000	0.184	0.000	0.000	0.000
GRADE	0.225	0.000	0.122	0.000	0.000
JOBNOW	0.000	0.156	0.000	0.406	0.000
INCOME	0.000	0.152	0.000	0.035	0.298

Indirect Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.105	0.010	0.389	0.051	0.000
FATHED	0.007	0.000	0.010	0.132	0.000
GRADE	0.001	0.002	0.001	0.010	0.000
JOBNOW	0.129	0.010	0.399	0.801	0.000
INCOME	0.010	0.006	0.025	0.277	0.000

Standardized Indirect Effects

	FATHOCC	FIRSTJOB	FATHED	GRADE	JOBNOW
FIRSTJOB	0.104	0.010	0.056	0.005	0.000
FATHED	0.045	0.002	0.010	0.084	0.000
GRADE	0.006	0.023	0.001	0.010	0.000
JOBNOW	0.132	0.011	0.059	0.075	0.000
INCOME	0.085	0.052	0.030	0.213	0.000

Summary of models

Model	NPAR	CMIN	DF	P	CMIN/DF
Model Number 3	16	99.039	5	0.000	19.808
Saturated model	21	0.000	0		
Independence model	6	582.593	15	0.000	38.840

Model	RMR	GFI	AGFI	PGFI
Model Number 3	11.699	0.944	0.763	0.225
Saturated model	0.000	1.000		
Independence model	62.260	0.665	0.531	0.475

Model	AIC	BCC	BIC	CAIC
Model Number 3	131.039	131.481	227.582	214.914
Saturated model	42.000	42.581	168.714	152.087
Independence model	594.593	594.759	630.797	626.046

Relative to the original model and the saturated model, this model in which son's first job predicts to father's educational level is a poor fit. The Information Criteria scores are all much higher than the original model and even the saturated model despite its greater parsimony. The GFI is lower, and the CMIN is high and statistically different from the saturated model, indicating a significant discrepancy between the observed and estimated variance-covariance matrix. Although the model is non-recursive, the feedback loop is stable since the stability index of 0.047 is within the stable range of -1 and 1. The path coefficient between son's first job and father's education level is .18 (sig) compared to the original path coefficient in the opposite direction of .05 (non-sig), but we know the association is spurious because son's first job does

not precede father's education level. Although there are some good warning signs (poor goodness of fit statistics), there will never be explicit statement that the causal structure is incorrectly specified so it is imperative to build models on sound theory and/or empirical evidence and carefully inspect model statistics.