(a)

Here is the Eviews output of the model:

Dependent Variable: LOGW
Method: Least Squares
Date: 11/29/15 Time: 16:41
Sample: 1 3010
Included observations: 3010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.611014	0.067895	67.91387	0.0000
EDUC	**0.081580**	**0.003499**	**23.31486**	**0.0000**
EXPER	0.083836	0.006774	12.37698	0.0000
EXPER^2	-0.002202	0.000324	-6.800161	0.0000
SMSA	0.150801	0.015836	9.522648	0.0000
SOUTH	-0.175176	0.014649	-11.95852	0.0000

Estimated β_2 coefficient is 0.082 (highlighted in yellow). It is highly significant with a p-value (highlighted in green) that is close to zero.

Holding all other independent variables constant, ***with one more year of education we can expect wage to increase by nearly 8.2% on average.***

(b)

Variables "*educ"* and "*exper"* might be endogenous because there might be other variables that affect these variables (while not affecting wage directly). As they are inconsistent, OLS estimates from part (a) are wrong estimates. In this case, as the sample size increases OLS estimators converge to the wrong value. There might be some (instrumental) variables that can definitely have an effect on these variables. ***For instance, being closer to a college may indicate more years in education because being close to a college can make education cheaper as one can save a lot by staying with his/her family, compared to living in another city to attend college.***

Although it may give us some ideas about the relationship between the variables, we can say that **we cannot totally trust OLS estimates**. Therefore, the OLS estimates cannot be a total use at this point. We have to make further investigations by using instrumental variables to reach to more meaningful/rightful conclusions.

(c)

As that motivation cannot be acquired at the first sight, looking at the original work of D. Card (1993)[1] helps here (pages 12&13). Here is the main motivation:

[1] Original working paper's 1993 version can be found in here: **http://www.nber.org/papers/w4483.pdf**

Experience and its square are constructed from observed age and education. <u>If education is measured with error so is experience. Following this, if education (educ) is really endogenous in the log wage (logw) equation then we can say for sure that experience is also endogenous (because it is mechanically related to education).</u> Therefore, to eliminate this problem, one can instrument experience and its square with age and age square.

(d)

As far as I can understand, for the first stage regression, I take 'educ' and regress it on the variables: *constant*, *age*, age^2, *nearc*, *dadeduc*, *momeduc*.

Here is the Eviews output for this regression:

Dependent Variable: EDUC
Method: Least Squares
Date: 11/29/15 Time: 18:56
Sample: 1 3010
Included observations: 3010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.923273	4.010502	-1.476940	0.1398
AGE	0.992550	0.281060	3.531459	0.0004
AGE^2	-0.017075	0.004878	-3.500252	0.0005
NEARC	0.528751	0.092698	5.704029	0.0000
DADED	0.202048	0.015665	12.89827	0.0000
MOMED	0.248379	0.017036	14.58002	0.0000

R-squared	0.233003	Mean dependent var	13.26346
Adjusted R-squared	0.231727	S.D. dependent var	2.676913
S.E. of regression	2.346346	Akaike info criterion	4.545587
Sum squared resid	16538.05	Schwarz criterion	4.557567
Log likelihood	-6835.109	Hannan-Quinn criter.	4.549896
F-statistic	182.5149	Durbin-Watson stat	1.777563
Prob(F-statistic)	0.000000		

All the individual coefficients of the instruments are highly significant (p-values are highlighted in green). **From this result, we can say that instruments that we want to use affect number of years of schooling significantly. They seem to be suitable for using as instruments.**

(Note: I did not include the result in here, but they are also jointly significant too).

(e)

As E-views has the ability to directly estimate 2SLS models, I took advantage of it and here is the result:

Dependent Variable: LOGW
Method: Two-Stage Least Squares
Date: 11/29/15 Time: 18:23
Sample: 1 3010
Included observations: 3010
Instrument specification: C SMSA SOUTH AGE AGE^2 NEARC DADED MOMED

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.416904	0.115421	38.26784	0.0000
EDUC	0.099843	0.006574	15.18793	0.0000
EXPER	0.072867	0.016713	4.359793	0.0000
EXPER^2	-0.001639	0.000838	-1.955853	0.0506
SMSA	0.134937	0.016770	8.046563	0.0000
SOUTH	-0.158987	0.015685	-10.13595	0.0000

R-squared	0.251174	Mean dependent var	6.261832
Adjusted R-squared	0.249927	S.D. dependent var	0.443798
S.E. of regression	0.384359	Sum squared resid	443.7856
F-statistic	175.8540	Durbin-Watson stat	1.862787
Prob(F-statistic)	0.000000	Second-Stage SSR	462.7457
J-statistic	3.695008	Instrument rank	8
Prob(J-statistic)	0.157630		

Instrumental variables (age, age^2, nearc, daded, momed) are highlighted in pink along with the exogenous variables (constant, smsa, south).

Coefficients are highlighted in green, to compare them with part (a), I extract the necessary parts from both tables:

2SLS Results:	
Variable	Coefficient
C	4.416904
EDUC	**0.099843**
EXPER	0.072867
EXPER^2	-0.001639
SMSA	0.134937
SOUTH	-0.158987

OLS Results		
Variable	Coefficient	Coefficient
C	4.611014	
EDUC	**0.081580**	
EXPER	0.083836	
EXPER^2	-0.002202	
SMSA	0.150801	
SOUTH	-0.175176	

As the tables show 2SLS results are different than OLS ones. The coefficient of education (educ) is now relatively higher for 2SLS, but other coefficients are lower for 2SLS.

(f)

There is no direct way to compute Sargan test in Eviews. **But by saving residuals from the 2SLS estimation from part (e) (I name it as "*res_2sls*") and regressing it on the instruments and exogenous variables, we can obtain the R^2 value that we need (which is a more practical way than the one explained in Training Exercise 4.4 (c) by the way).** After that, by multiplying it with the number of observations (3010 in our case) we can obtain the Sargan test statistic.

Here is the result of the regression:

Dependent Variable: RES_2SLS
Method: Least Squares
Date: 11/29/15 Time: 20:33
Sample: 1 3010
Included observations: 3010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.125821	0.656812	0.191564	0.8481
AGE	-0.009332	0.046038	-0.202693	0.8394
AGE^2	0.000159	0.000799	0.199054	0.8422
NEARC	0.013508	0.016367	0.825326	0.4093
DADED	-0.004105	0.002579	-1.591980	0.1115
MOMED	0.004113	0.002813	1.462481	0.1437
SMSA	-0.003346	0.016766	-0.199595	0.8418
SOUTH	0.002226	0.015037	0.148032	0.8823

R-squared	0.001230	Mean dependent var	-6.75E-16
Adjusted R-squared	-0.001099	S.D. dependent var	0.384039
S.E. of regression	0.384250	Akaike info criterion	0.927608
Sum squared resid	443.2398	Schwarz criterion	0.943581
Log likelihood	-1388.050	Hannan-Quinn criter.	0.933353
F-statistic	0.528157	Durbin-Watson stat	1.863636
Prob(F-statistic)	0.813842		

The values we need are highlighted in blue.

Sargan test statistic=0.00123*3010=3.7023

Now as we have 8 instruments and exogenous variables and 5 explanatory variables from the model in part (a).

The chi-square value at 0.05 level that we need to compare with the Sargan t-stat is:

$\chi^2(5)=7.815$

As 3.7023<7.815 we <u>cannot reject the Null Hypothesis that the instruments are valid</u>.

In summary, we can state that ***the instruments which we used are valid!***