Question 1

a)

```
. *1. a)
. reg testscr str

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(1, 418)       =     22.58
       Model |  7794.11004         1  7794.11004   Prob > F        =    0.0000
    Residual |  144315.484       418  345.252353   R-squared       =    0.0512
-------------+----------------------------------   Adj R-squared   =    0.0490
       Total |  152109.594       419  363.030056   Root MSE        =    18.581

------------------------------------------------------------------------------
     testscr | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .4798256    -4.75   0.000    -3.22298   -1.336637
       _cons |   698.933   9.467491     73.82   0.000    680.3231    717.5428
------------------------------------------------------------------------------
```

We have a negative coefficient for STR which implies that a rise in student-teacher ratio (i.e., rise in number of students for a given number of teachers or a fall in the number of teachers for a given number of students) leads to a fall in test scores.
This is not a credible estimate of the causal effect on test scores as there can be variables that influence test scores but are hiding in the error and are correlated with the regressor.

b) We are overestimating the coefficient of STR. The coefficient is more negative due to the omitted variable bias.

c)

```
. *1. c)
. reg testscr str el_pct

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(2, 417)       =    155.01
       Model |  64864.3011         2  32432.1506   Prob > F        =    0.0000
    Residual |  87245.2925       417  209.221325   R-squared       =    0.4264
-------------+----------------------------------   Adj R-squared   =    0.4237
       Total |  152109.594       419  363.030056   Root MSE        =    14.464

------------------------------------------------------------------------------
     testscr | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -1.101296   .3802783    -2.90   0.004    -1.848797   -.3537945
      el_pct |  -.6497768   .0393425   -16.52   0.000    -.7271112   -.5724423
       _cons |   686.0322   7.411312    92.57   0.000     671.4641    700.6004
------------------------------------------------------------------------------
```

The coefficient on STR reduces in absolute terms because by not including english learner percentage, we were overestimating the causal effect of STR on test scores.

Here we find that the test statistic of F is calculated to be 14.94. On comparing this with the table value of F, which is 2.3, we can reject the null i.e., B1=B2=0.
Even the p values give the same conclusion as 0.000 is less than our level of significance.
Thus we know that coefficients of str and expn_stu are jointly significant.
We should not exclude either of the variables as that can lead to omitted variable bias. We should just accept multicollinearity as it is less worse of a problem when compared to omitted variable bias.