# CAN WE PREDICT READING SCORES BASED ON PARENT AND STUDENT ACTIONS AND PERCEPTIONS, AND HOME DEMOGRAPHICS, REGARDING READING? REGRESSION ANALYSIS USING NELS:88 DATA

***Shreya Goel*, Teachers College, Columbia University**

***Lauren Romine*, Teachers College, Columbia University**

***Videesha Bordoloi*, Teachers College, Columbia University**

**Acknowledgements**

This report is the final project requirement for the course HUDM5122 (Applied Regression Analysis) at Teachers College, Columbia University for the Spring 2017 semester. We thank Professor Elizabeth Tipton, Nicholas Anderson, and Andrew Dakapolos for their support throughout the course. All errors are our own.

## I. Introduction

A study by Hewison and Tizard (1980) studied the association between number of home background factors and the reading ability in three samples of working class children. They found that home background was which related most strongly to the reading achievement was whether the mother regularly coached the child to read.

A study by Morgan and Fuchs (2007) studied the relationship between children's reading skills and reading motivation. The results of the study indicates that reading skills and motivation correlates. The study found that there is a bidirectional relationship between reading skills and motivation to read. The finding of the study suggests that children's reading and motivation influence each other bidirectionally to affect their later reading ability. This means that it is important to provide interventions which effectively counteract initially poor reading skills and low motivation.

Cipielewski and Stanovich (1992) conducted a study to predict growth in reading ability from children's exposure to print. This was a longitudinal study of growth in reading ability by employing two indicators of print exposure. The study used the author recognition test (ART) and the Magazine Recognition Test (MRT) and found that the extent to which individuals engage in literacy activities is a significant contributor to developed reading ability. Print exposure appears to be both a consequences of developed reading ability and a contributor to further growth in reading ability.

A study by Baker, Scher and Mackler (2010) conducted a study on home and family influences on motivations for reading. The studied reviewed literature on home and family influences in children's motivation for reading. The study found that children who have early encounters with literacy which are enjoyable are more likely to develop a predisposition to read frequently and broadly in subsequent years. Shared storybook reading plays an important role in

promoting reading motivations. When the socio-emotional climate at home is positive children are more interested in reading as well as view it as enjoyable. The belief held by parents about the importance of reading and how relatable it is also associated with children's motivation for reading. Parents who view reading as an enjoyable activity/source of entertainment have children who have a positive attitude towards reading.

Senechal and LeFevre (2002) conducted a longitudinal study to show the complex relations among early home literacy experiences, subsequent receptive language and emergent literacy skills and reading achievement. The result of the study showed that children who have more exposure to books have better vocabulary development and listening comprehension skills. The study also found that parents who are actively involved in teaching their children about reading and writing words had a positive impact on the development of the child's early literacy skills.

Various factors contribute to predicting of standardized reading scores: parental involvement, home background,exposure to print etc. This research paper seeks to contribute to the literature by using the data from the National Educational Longitudinal Survey 1988 to address the question of whether we can predict reading scores using demographic information related to actions and perceptions rather than other test scores or performance measures.

## II. Research Design

### 2.1 Data

The National Educational Longitudinal Survey of 1988 ("NELS:88") is a longitudinal study designed to collect a representative sample of data from 8th grade students across the United States in the spring of 1988. In 1990, 1992, 1994 and 2000, additional supplemental surveys were administered to the same students in order to gather continuous data on the cohort

and its performance during high school and into adult life, according to the survey's overview page on the National Center for Education Statistics website (https://nces.ed.gov/surveys/nels88/).

The data was collected by the Department of Education, which choose a sampling of schools and students using a "two-stage stratified probability sampling design," according to the ICPSR description. This sample included "1,734 school selections with 1,052 participating schools, including 815 public and 237 private schools. The second stage produced a random selection of 26,435 students among the sampled schools, resulting in participation by 24,599 eighth-grade students."

Additionally, for each student in the sample, the researchers collected data, including information from personal interviews and test scores, from four sources: Parents, School Administrators, Teachers and the Students themselves.

The collected data, therefore, includes observations on 24,599 students with varying degrees of completeness across students. There are 408 variables that explore a large assortment of topics, including but not limited to demographic information, standardized test scores, and attendance information. These variables were generally measured as either nominal or scale and specific values for missing information are well noted.

For this research regarding Reading Scores, only the initial 1988 survey results are included. A number of different variables are used to assess the best predictors of Reading Score and observations with pertinent missing data have been withheld so as not to skew the results (as the missing data input typically takes a numerical value which would significantly impact the analysis and results).

## 2.2 Statistical Analysis

The dependent variable of interest which is what we are trying to predict in this model is the Reading Standardized Score. In the NELS:88 dataset, this is variable number 377 and is listed in the data as "BYTXRSTD." It is a scale variable with values ranging from 23.098 to 67.499, with missing data coded as 999.998. Of the 23,932 observations in the dataset that had information for this variable, 882 (or 3.7%) were coded as the missing value and were therefore omitted from all of our analyses.

Given the large number of possible independent variables in this dataset, we used a combination of research papers that touched on what might influence reading scores as well as a logical review of the many available variables in the dataset that might support prediction of the Standardized Reading Score. Using this approach, we formulated three models using variables related to: home demographics regarding regarding reading; student actions or conceptions regarding reading; and parental involvement or conceptions regarding reading. We also looked at additional variables that we chose not to pursue in this iteration, such as the standardized scores for science and history; these we excluded because they were concurrent with the Reading Standardized Score and therefore likely to be (highly) correlated and, since they would generally be from the same time and same type of tests, not able to predict in advance if the model were used for additional data which did not yet include these standardized scores.

## 2.3 Model Selection

Model 1: Home Demographics Related to Reading

For demographics relating to whether reading might be supported or encouraged in the student's home, we used the following variables:

- "R's Family Has Specific Place For Study" (variable 83, coded "BYS35A," nominal with codes for Yes or No as well as missing values; recoded into dummy variable using No as reference category and Yes in regression model and omitting missing values)
- "R's Family Has A Daily Newspaper" (variable 84, coded "BYS35B," nominal with codes for Yes or No as well as missing values; recoded into dummy variable using No as reference category and Yes in regression model and omitting missing values)
- "R's Family Has More Than 50 Books" (variable 95, coded "BYS35M," nominal with codes for Yes or No as well as missing values; recoded into dummy variable using No as reference category and Yes in regression model and omitting missing values).

According to the omnibus test, this model was significant ($p < .001$) and explained roughly 5% of the variation for both adjusted $R^2$ and standard $R^2$ (see Table 5). Each of the three factors was also significant ($p < .001$) with regard to the reference categories (see Table 1). Interestingly, while having more than 50 books and receiving a daily newspaper both had positive effects compared to not having these, having a specific place to study had a negative effect compared to not having a specific place to study at home.

Models 2, 3 and 4: Student Actions and Perceptions Regarding Reading

In order to assess a student's interest, actions, and perceptions regarding reading, which would in theory contribute to Reading Standardized Scores, we looked at the following variables:

- "Number Of Hours Spent On Homework Per Week" (variable 368, coded "BYHOMEWK," nominal with eight levels for hours ranging from 0 to 21 or more hours as well as an additional level for missing values; recoded into dummy variables following the breakdown of the original data, 1=none, 2=.5 to 1.99, 3=2 to 2.99, 4=3-5.49, 5=5.50

to 10.49, 6=10.50 to 12.99, 7=13.00 to 20.99, 8=21 hrs or more and omitting missing values)

- "In Advanced, Enriched, Accelerated English" (variable 224, coded "BYS66A," nominal with codes for Yes or No as well as missing values; recoded into dummy variables with No as the reference category and Yes in the model, and omitting missing values)
- "Usually Look Forward To English Class" (variable 257, coded "BYS70A," nominal with codes for Strongly Agree, Agree, Disagree, Strongly Disagree, and missing values; recoded into two dummy variables for Agree or Disagree with Disagree as the reference category, and omitting missing values)
- "How Much Reading Do You Do On Your Own" (variable 288, coded "BYS80," nominal with five levels for range of less than an hour per week to six or more hours per week, as well as missing values; recoded into dummy variables following the original data, 1=none, 2=1 hr or less, 3=2 hrs or less, 4=3 hrs or less, 5=3-5 hrs, 6=6 or more hours and omitting missing values).

Using the omnibus test, this model is significant ($p < .001$) and explains 12.8% of the variation in Reading Standardized Score, according to both adjusted $R^2$ and standard $R^2$. Again, interestingly, most of the individual coefficients are significant and positive except for looking forward to English class (as compared with not looking forward to English class) and the two levels for spending less than 3 hours on homework per week (as compared with no hours on homework), which are significant but negative.

Because of a high level of collinearity, we removed one level of hours spent on homework per week (level 4: 3-5.49 hrs) which had the effect of slightly reducing the adjusted $R^2$ to 12.5% but seemed to fix the issue of collinearity within the homework hours factor while

maintaining significance for the model. However, this also made two additional levels of homework per week (levels 2 and 3, combined .5 to 2.99 hours) non-significant (p values of .892 and .812, respectively) so we removed these two levels and ran the model again. The final model for this set, model 4, maintained the 12.5% $R^2$ while achieving significance for all remaining variables (see Tables 2 and 5).

Models 5 and 6: Parent Involvement and Perceptions Regarding Reading

In order to assess the parental actions and perceptions regarding reading for students, and to see if there is a connection that can be used to predict Reading Standardized Scores, we looked at the following variables (and we focused specifically on the mother's involvement when there were overlapping variables for mother or father, as the research papers we reviewed stressed that the mother's involvement should have a higher impact):

- "Mother/Female Guardian Employment Status" (variable 5, coded "BYS4A," nominal with four levels for Currently Working, Unemployed, Retired or Disabled, as well as missing values; recoded into dummy variables for the four levels of data using Unemployed as the reference category, and omitting missing values).
- "Discuss School Activities With Parents" (variable 100, coded "BYS36B," nominal with three levels for Not At All, Once Or Twice, or Three Or More Times, as well as missing values; recoded into dummy variables following the original data using Not At All as the reference category, and omitting missing values)
- "Discuss Things Studied In Class With Parents" (variable 101, coded "BYS36C," nominal with three levels for Not At All, Once Or Twice, or Three Or More Times, as well as missing values; recoded into dummy variables following the original data, using Not At All as the reference category, and omitting missing values)

- "How Often Parents Check On R's (Student's) Homework" (variable 106, coded "BYS38A," nominal with four levels for Often, Sometimes, Rarely, Never, as well as missing values; recoded into dummy variables, 1=Never, 2=Rarely, 3=Sometimes, 4=Often, using Never as the reference category and omitting missing values)
- "Talk To Mother About Planning HS Program" (variable 145, coded "BYS50B," nominal with three levels for not at all, once or twice, or three or more times, as well as missing values; recoded into dummy variables following the original data, using Not At All as the reference category, and omitting missing values).

According to the omnibus test, this model is significant ($p < .001$) and the adjusted $R^2$ is 10.2%, showing that this model accounts for approximately 10% of the variation in Reading Standardized Scores. All of the coefficients were significant ($p < .001$), except for the retired level of mother's employment status ($p = .1$), so we removed that level for the next iteration of the model. The resulting model, model 6, maintained significance and 10.2% $R^2$ while achieving significance for all remaining variables (see Tables 3 and 5).

Models 7 and 8: Final Models

Using models 1, 4, and 6, we combined all of the significant variables in order to make a full model. Using these 25 variables, we achieved model significance ($p < .001$) and an adjusted $R^2$ of 20.2%, accounting for 20.2% of the variation in Reading Standardized Scores (see Tables 4 and 5). One of the variables in the full model, how often parents check on the student's homework (level 2 = rarely) had a non-significant p-value in this model, so we removed it and reran the model. The final resulting model with 24 variables maintained an adjusted $R^2$ of 20.2% while maintaining model significance and achieving significance for all remaining variables (see Tables 4 and 5).

## III. Results

We attempted to use demographic, involvement or action, and perception related variables in order to predict Reading Standardized Scores rather than other traditional methods such as other test scores or school performance. Based on our model, we were able to predict 20.2% of the variation in the NELS:88 Reading Standardized Scores using a combination of home demographic information relating to reading, parent involvement and perceptions regarding reading, and student actions and perceptions regarding reading. Given the variables we used, this is a significant prediction model. For future research, it would be logical to try other variables and/or combinations from this dataset as well as potentially exploring other research surveys that might provide additional demographic or other related information that was not available in this particular dataset.

## IV. Conclusion

This analysis of NELS:88 data was used to predict the reading scores of students based on parent and student actions and perceptions, and home demographics regarding reading. This study sought to contribute to existing literature about factors which can be used to predict student reading scores apart from traditional methods like test scores or school performance. The result of the study and analysis suggest that home demographics related to reading, student's perceptions and actions regarding reading, and parental involvement and parents' perceptions regarding reading contribute significantly to predicting the reading scores of the student. Among these three sets of variables, student's perceptions and actions regarding reading, and parental involvement and parents' perceptions has a more of an effect on reading scores than home demographics. The three models together predict 20.2% of the variation in Standardized Reading Scores. The results of the study are in line with existing literature mentioned in the introduction.

**References**


Cipielewski, J., & Stanovich, K. E. (1992). Predicting Growth in Reading Ability from Children's Exposure to Print. *Journal of Experimental Child Psychology,54*, 74-89.

Hewison, J., & Tizard, J. (1980). Parental Involvement And Reading Attainment. British Journal of Educational Psychology,50(3), 209-215. doi:10.1111/j.2044-8279.1980.tb00803.x

Linda Baker , Deborah Scher & Kirsten Mackler (1997) Home and family influences on motivations for reading, Educational Psychologist, 32:2, 69-82, DOI: 10.1207/s15326985ep3202_2

Morgan, P. L., & Fuchs, D. (2007). Is There a Bidirectional Relationship Between Children's Reading Skills and Reading Motivation? *Council for Exceptional Children,73*(2), 165-183.

National Education Longitudinal Study of 1988 (NELS:88). (n.d.). Retrieved from https://nces.ed.gov/surveys/nels88/

Senechal, M., & LeFevre, J. (2002). Parental Involvement in the Development of Children's Reading Skill: A Five-Year Longitudinal Study. *Child Development,73*(2), 445-460. doi:10.1111/1467-8624.00417

Zadovny, M. (2006). Does watching television rot your mind? Estimates of the effect on Test scores *Economics of Education Review*, 25(5), 565-573.

## Appendix 1: Tables and Figures

**Table 1: Model 1 - Home Demographics Related to Reading**

| Variables | B | Sig. | Tolerance | VIF |
|---|---|---|---|---|
| (Constant) | 44.294 | .000 | | |
| Has Specific Place for Study=Have | -.807 | .000 | .988 | 1.012 |
| Family Has a Daily Newspaper=Have | 1.380 | .000 | .976 | 1.025 |
| Family Has More Than 50 Books=Have | 6.466 | .000 | .974 | 1.027 |

a. Dependent Variable: READING STANDARDIZED SCORE

**Table 2: Model 4 - Student Actions and Perceptions Related to Reading**

| Variables | B | Sig. | Tolerance | VIF |
|---|---|---|---|---|
| (Constant) | 46.516 | .000 | | |
| In advanced, enriched, accelerated English=Yes | .469 | .001 | .981 | 1.020 |
| Usually look forward to English class=1.0 | -1.367 | .000 | .962 | 1.039 |
| How much reading do you do on your own=1 hr or less per wk | 3.192 | .000 | .565 | 1.770 |
| How much reading do you do on your own=2 hours | 4.737 | .000 | .611 | 1.635 |
| How much reading do you do on your own=3 hours | 6.581 | .000 | .709 | 1.411 |
| How much reading do you do on your own=4-5 hours | 8.801 | .000 | .780 | 1.281 |
| How much reading do you do on your own=6 hrs or more per wk | 9.819 | .000 | .738 | 1.355 |
| Number of hrs spent on homework per week=5.50-10.49 hrs | 2.033 | .000 | .945 | 1.058 |
| Number of hrs spent on homework per week=10.50-12.99 hrs | 3.686 | .000 | .973 | 1.028 |
| Number of hrs spent on homework per week=13.00-20.99 hrs | 4.250 | .000 | .952 | 1.050 |
| Number of hrs spent on homework per week=21.00 and up hrs | 2.547 | .000 | .973 | 1.027 |

a. Dependent Variable: READING STANDARDIZED SCORE

**Table 3: Model 6 – Parents Involvement Related to Reading**

| Variables | B | Sig. | Tolerance | VIF |
|---|---|---|---|---|
| (Constant) | 42.766 | | | |
| Mother/female guardian employment status=Currently working | 3.166 | .000 | .874 | 1.144 |
| Mother/female guardian employment status=Disabled | -1.661 | .000 | .880 | 1.136 |
| Discuss school activities with parents=Once or twice | 1.563 | .002 | .290 | 3.447 |
| Discuss school activities with parents=3 or more times | 4.254 | .000 | .266 | 3.764 |
| Discuss things studied in class with parents=Once or twice | 1.723 | .000 | .337 | 2.968 |
| Discuss things studied in class with parents=3 or more times | 4.568 | .000 | .304 | 3.285 |
| Talk to mother about planning h.s. prog=Once or twice | 1.137 | .000 | .334 | 2.998 |
| Talk to mother about planning h.s. prog=3 or more times | 1.797 | .000 | .307 | 3.258 |
| How often parents check on r's homework=2 | -.512 | .000 | .440 | 2.270 |
| How often parents check on r's homework=3 | -2.367 | .044 | .349 | 2.863 |
| How often parents check on r's homework =4 | -3.005 | .000 | .314 | 3.187 |

a. Dependent Variable: READING STANDARDIZED SCORE

**Table 4: Model 8 – All significant variables related to reading from our analysis**

| Variables | B | Sig. | Tolerance | VIF |
|---|---|---|---|---|
| (Constant) | 38.441 | .000 | | |
| R's family has specific place for study=Have | -1.446 | .000 | .941 | 1.062 |
| R's family has a daily newspaper=Have | .745 | .000 | .968 | 1.034 |
| R's family has more than 50 books=Have | 3.802 | .000 | .936 | 1.068 |
| In advanced, enriched, accelerated English=Yes | .752 | .000 | .975 | 1.026 |
| Usually look forward to English class=1.0 | -1.176 | .000 | .944 | 1.060 |
| How much reading do you do on your own=1 hr or less per wk | 2.421 | .000 | .543 | 1.840 |
| How much reading do you do on your own =2 hours | 3.706 | .000 | .584 | 1.712 |
| How much reading do you do on your own =3 hours | 5.300 | .000 | .680 | 1.471 |
| How much reading do you do on your own =4-5 hours | 7.178 | .000 | .751 | 1.331 |
| How much reading do you do on your own =6 hrs or more per wk | 8.053 | .000 | .706 | 1.416 |
| BYHOMEWK=5.50 to 10.49 hours | 1.607 | .000 | .936 | 1.069 |
| BYHOMEWK=10.50 to 12.99 hours | 2.827 | .000 | .965 | 1.036 |
| BYHOMEWK=13.00 to 20.99 hours | 3.310 | .000 | .937 | 1.067 |
| BYHOMEWK=21.00 and up hours | 1.692 | .000 | .965 | 1.036 |
| Mother/female guardian employment status =Currently working | 2.595 | .000 | .874 | 1.144 |
| Mother/female guardian employment status=Disabled | -1.240 | .023 | .881 | 1.135 |
| Discuss school activities with parents=Once or twice | 1.252 | .000 | .274 | 3.654 |
| Discuss school activities with parents=3 or more times | 3.344 | .000 | .250 | 3.999 |
| Discuss things studied in class with parents=Once or twice | 1.109 | .000 | .322 | 3.103 |
| Discuss things studied in class with parents =3 or more times | 3.111 | .000 | .288 | 3.475 |
| Talk to mother about planning h.s. prog=Once or twice | .722 | .002 | .321 | 3.113 |
| Talk to mother about planning h.s. prog=3 or more times | .913 | .000 | .293 | 3.408 |
| How often parents check on r's homework=3 | -1.790 | .000 | .667 | 1.498 |
| How often parents check on r's homework=4 | -2.504 | .000 | .640 | 1.561 |

a. Dependent Variable: READING STANDARDIZED SCORE

**Table 5: Model Statistics**

| Models | $R^2$ | Adjusted $R^2$ | F-test |
|---|---|---|---|
| 1 | .047 | .047 | .000 |
| 2 | .128 | .128 | .000 |
| 3 | .125 | .125 | .000 |
| 4 | .125 | .125 | .000 |
| 5 | .102 | .102 | .000 |
| 6 | .102 | .102 | .000 |
| 7 | .203 | .202 | .000 |
| **8** | **.203** | **.202** | **.000** |

## Appendix 2: Annotated SPSS Syntax

*__Model 1__ - Home Demographics and Standardized reading scores
*__Cleaning Y variable__ - using select cases which are less than 100, missing values are coded as 999.998

```
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(BYTXRSTD < 100).
VARIABLE LABELS filter_$ 'BYTXRSTD < 100 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

*__x1=BYS35A__. Select cases less than 3 to omit missing values. 1=Have, 2=Don't Have. Dummy variables=2, reference category = don't have.

```
USE ALL.
COMPUTE filter_$=(BYS35A < 3).
VARIABLE LABELS filter_$ 'BYS35A < 3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS35A
ROOTNAME1=BYS35A
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="x".
```

*__x2=BYS35B__. Select cases less than 3 to omit missing values. 1=Have, 2=Don't Have. Dummy variables=2, reference category = don't have.

```
USE ALL.
COMPUTE filter_$=(BYS35B < 3).
VARIABLE LABELS filter_$ 'BYS35B < 3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS35B
ROOTNAME1=BYS35B
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="x1".
```

*__x3=BYS35M__. Select cases less than 3 to omit missing values. 1=Have, 2=Don't Have. Dummy variables=2, reference category = don't have.

```
USE ALL.
```

```
COMPUTE filter_$=(BYS35M < 3).
VARIABLE LABELS filter_$ 'BYS35M < 3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS35M
ROOTNAME1=BYS35M
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="x".
```

*__Model 1: Regression__ Home Demographics regarding Reading (reference category is don't haves and is stated above)

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS35A_1 BYS35B_1 BYS35M_1
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Models 2,3,4:__ Student Actions and Perceptions regarding reading

*__x1=BYS66A__. In advanced Enriched English. 1=Yes, 2=No. So, select Cases less than 3 to omit missing values. Dummy Code the variable. Reference is "NO"

```
USE ALL.
COMPUTE filter_$=(BYS66A < 3).
VARIABLE LABELS filter_$ 'BYS66A < 3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS66A
ROOTNAME1=BYS66A
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="IN_ADVANCED_ENGLISH".
```

*__x2=BYS70A__. Select cases less than 5 to omit missing values. Recode - Agree and Strongly agree as 1, disagree and strongly disagree as 2. Reference category is 2 - disagree.

```
USE ALL.
COMPUTE filter_$=(BYS70A < 5).
VARIABLE LABELS filter_$ 'BYS70A < 5 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

```
RECODE BYS70A (1=1) (2=1) (3=2) (4=2) INTO BYS70A_New.
VARIABLE LABELS BYS70A_New 'UsuallyLooksForwardToEnglishClass'.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS70A_New
ROOTNAME1=BYS70A_
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="UsuallyLooksForwardToEnglishClass".
```

*__x3=BYS80__. How much reading do you do on your own. 0=None, 1=1 hour or less. 2=2 hours, 3=3 hours, 4=4-5 hours, 5=6 hours. Select cases less than 6 to omit missing values. Dummy Code the variable. Reference category is 1 which is none.

```
USE ALL.
COMPUTE filter_$=(BYS80 < 6).
VARIABLE LABELS filter_$ 'BYS80 < 6 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS80
ROOTNAME1=BYS80_
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="AmountOfReadingSelf".
```

*__x4=BYHOMEWK__. Select cases less than 9 to omit missing values. Create dummy variables: 1=None, 2=.5-1.99 hours, 3=2-2.99 hours, 4=3-5.99 hours, 5=5.5-10.49 hours, 6=10.5-12.99 hours, 7=13-20.99 hours, 8=21 or more hours. Reference category is None.

```
USE ALL.
COMPUTE filter_$=(BYHOMEWK < 9).
VARIABLE LABELS filter_$ 'BYHOMEWK < 9 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYHOMEWK
ROOTNAME1=BYHOMEWK_
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="HoursSpentOnHomework".
```

*__Model 2: Regression__. Student Actions and Perceptions regarding reading (reference category stated in each variable above)

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
```

```
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS66A_1 BYS70A_1 BYS80_2 BYS80_3 BYS80_4 BYS80_5 BYS80_6
BYHOMEWK__2
   BYHOMEWK__3 BYHOMEWK__4 BYHOMEWK__5 BYHOMEWK__6 BYHOMEWK__7
BYHOMEWK__8
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 3: Regression.__ Student Actions and Perceptions regarding reading. Due to high VIFs we remove BYHOMEWK_4 to see if the Collinearity improves.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS66A_1 BYS70A_1 BYS80_2 BYS80_3 BYS80_4 BYS80_5 BYS80_6
BYHOMEWK__2 BYHOMEWK__3 BYHOMEWK__5 BYHOMEWK__6 BYHOMEWK__7
BYHOMEWK__8
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 4__: __Regression__. Student Actions and Perceptions regarding reading. Remove BYHOMEWK_2 and BYHOMEWK_3 due to non-significance.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS66A_1 BYS70A_1 BYS80_2 BYS80_3 BYS80_4 BYS80_5 BYS80_6
BYHOMEWK__5 BYHOMEWK__6 BYHOMEWK__7 BYHOMEWK__8
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 5,6__: Parents' Involvement and Perceptions regarding reading.

*__x1= BYS4A__ . Select Cases less than 5. Dummy Code. 1=working, 2=unemployed, 3=retired, 4=disabled. Missing values omitted, 2=unemployed used as reference category.

```
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(BYS4A < 5).
VARIABLE LABELS filter_$ 'BYS4A < 4 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS4A
ROOTNAME1=BYS4A
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="MotherEmploymentStatus".
```

*__x2=BYS36B__. Discuss School Activities With Parents. Select cases less than 4. Dummy Code. 1=No, 2=once or twice, 3=three or more times, reference category 1=No.

```
USE ALL.
COMPUTE filter_$=(BYS36B < 4).
VARIABLE LABELS filter_$ 'BYS36B < 4 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS36B
ROOTNAME1=BYS36B
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="DiscussSchoolActivitiesWithParents".
```

*__x3=BYS36C__. Discuss Things Studied in Class With Parents. Select cases less than 4. Dummy Code: 1=No, 2=once or twice, 3=three or more times. Missing values omitted, reference category 1=No.

```
USE ALL.
COMPUTE filter_$=(BYS36C < 4).
VARIABLE LABELS filter_$ 'BYS36C < 4 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS36C
ROOTNAME1=BYS36C
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="DiscussStudiesWithParents".
```

*__x4 = BYS38A__. How often parents check on R's homework. Recoded as 1=never, 2=rarely, 3=sometimes, 4=often. Missing values omitted, 1=never, used as reference category.

```
USE ALL.
COMPUTE filter_$=(BYS38A < 5).
VARIABLE LABELS filter_$ 'BYS38A < 5 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

STRING BYS38A_New (A8).
RECODE BYS38A (1='4') (2='3') (3='2') (4='1') INTO BYS38A_New.
VARIABLE LABELS BYS38A_New 'HowOftenParentsCheckOnHomework'.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS38A_New
ROOTNAME1=BYS38A
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
```

```
MACRONAME1="HowOftenParentsCheckHomework".
```

*__x5=BYS50B__. Talks to Mother About Planning High School Courses. 1=No, 2=once or twice, 3=three or more times; 1=No used as reference category, missing values omitted..

```
USE ALL.
COMPUTE filter_$=(BYS50B < 4).
VARIABLE LABELS filter_$ 'BYS50B < 4 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SPSSINC CREATE DUMMIES VARIABLE=BYS50B
ROOTNAME1=BYS50B
/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO
MACRONAME1="TalksToMotherAboutChoosingHighSchoolCourses".
```

*__Model 5: Regression__ Parents' Involvement and Perceptions regarding reading.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS4A_1 BYS4A_3 BYS4A_4 BYS36B_2 BYS36B_3 BYS36C_2 BYS36C_3
BYS50B_2 BYS50B_3
   BYS38A_3 BYS38A_4 BYS38A_5
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 6__: __Regression__ Parents' Involvement and Perceptions regarding reading. Remove BYS4A_3 Retired due to non-significance.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS4A_1 BYS4A_4 BYS36B_2 BYS36B_3 BYS36C_2 BYS36C_3 BYS50B_2
BYS50B_3
   BYS38A_3 BYS38A_4 BYS38A_5
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 7__: __Regression__. Full Model, including all variables from models 1,4,6

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
```

```
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS35A_1 BYS35B_1 BYS35M_1
BYS66A_1 BYS70A__1 BYS80__2 BYS80__3 BYS80__4 BYS80__5 BYS80__6 BYHOMEWK__5
BYHOMEWK__6 BYHOMEWK__7 BYHOMEWK__8
BYS4A_1 BYS4A_4 BYS36B_2 BYS36B_3 BYS36C_2 BYS36C_3 BYS50B_2 BYS50B_3
   BYS38A_3 BYS38A_4 BYS38A_5
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

*__Model 8: Regression__ Full, Final Model. remove BYS38A_3 because p value is 0.163 not significant and r square and adjusted r square don't change which is ideal

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT BYTXRSTD
 /METHOD=ENTER BYS35A_1 BYS35B_1 BYS35M_1
BYS66A_1 BYS70A__1 BYS80__2 BYS80__3 BYS80__4 BYS80__5 BYS80__6 BYHOMEWK__5
BYHOMEWK__6 BYHOMEWK__7 BYHOMEWK__8
BYS4A_1 BYS4A_4 BYS36B_2 BYS36B_3 BYS36C_2 BYS36C_3 BYS50B_2 BYS50B_3
   BYS38A_4 BYS38A_5
 /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```