Predictive Utility of Metacognitive Reading Strategies Across Educational Models - Raw Regression Analyses

Nirmal Ghimire & Kouider Mokhtari, K-16 Literacy Center

2024-11-24

1. Understanding Data Structure

i. Data Dimension and Variables in the Dataset

```
[1] 612004     33
```

```
 [1] "CNTRYID"    "CNTSCHID"   "CNTSTUID"   "ST164Q01IA" "ST164Q02IA"
 [6] "ST164Q03IA" "ST164Q04IA" "ST164Q05IA" "ST164Q06IA" "ST165Q01IA"
[11] "ST165Q02IA" "ST165Q03IA" "ST165Q04IA" "ST165Q05IA" "ST166Q01HA"
[16] "ST166Q02HA" "ST166Q03HA" "ST166Q04HA" "ST166Q05HA" "UNDREM"
[21] "METASUM"    "METASPAM"   "W_FSTUWT"   "PV1READ"    "PV2READ"
[26] "PV3READ"    "PV4READ"    "PV5READ"    "PV6READ"    "PV7READ"
[31] "PV8READ"    "PV9READ"    "PV10READ"
```

ii. Descriptive Statistics

```
# A tibble: 19 x 5
   variable   mean      sd        min                       max
   <chr>      <dbl+lbl> <dbl+lbl> <dbl+lbl>                 <dbl+lbl>
 1 ST164Q01IA  3.53     1.61       1 [Not useful at all(1)] 6 [Very useful(6)]
 2 ST164Q02IA  3.21     1.59       1 [Not useful at all(1)] 6 [Very useful(6)]
 3 ST164Q03IA  3.69     1.66       1 [Not useful at all(1)] 6 [Very useful(6)]
 4 ST164Q04IA  4.31     1.63       1 [Not useful at all(1)] 6 [Very useful(6)]
 5 ST164Q05IA  4.28     1.60       1 [Not useful at all(1)] 6 [Very useful(6)]
 6 ST164Q06IA  3.19     1.73       1 [Not useful at all(1)] 6 [Very useful(6)]
 7 ST165Q01IA  3.53     1.65       1 [Not useful at all(1)] 6 [Very useful(6)]
 8 ST165Q02IA  2.84     1.56       1 [Not useful at all(1)] 6 [Very useful(6)]
 9 ST165Q03IA  3.84     1.56       1 [Not useful at all(1)] 6 [Very useful(6)]
10 ST165Q04IA  4.41     1.51       1 [Not useful at all(1)] 6 [Very useful(6)]
11 ST165Q05IA  4.39     1.61       1 [Not useful at all(1)] 6 [Very useful(6)]
12 ST166Q01HA  3.01     1.75       1 [Not useful at all(1)] 6 [Very useful(6)]
13 ST166Q02HA  4.07     1.73       1 [Not useful at all(1)] 6 [Very useful(6)]
14 ST166Q03HA  2.61     1.64       1 [Not useful at all(1)] 6 [Very useful(6)]
15 ST166Q04HA  3.21     1.79       1 [Not useful at all(1)] 6 [Very useful(6)]
16 ST166Q05HA  3.94     1.80       1 [Not useful at all(1)] 6 [Very useful(6)]
17 UNDREM     -0.0789   0.999     -1.64                      1.5
18 METASUM    -0.142    1.00      -1.72                      1.36
19 METASPAM   -0.160    0.985     -1.41                      1.33
```

```
  ST164Q01IA      ST164Q02IA      ST164Q03IA      ST164Q04IA
Min.   :1.00    Min.   :1.00    Min.   :1.00    Min.   :1.00
1st Qu.:2.00    1st Qu.:2.00    1st Qu.:2.00    1st Qu.:3.00
Median :3.00    Median :3.00    Median :4.00    Median :5.00
Mean   :3.53    Mean   :3.21    Mean   :3.69    Mean   :4.31
3rd Qu.:5.00    3rd Qu.:4.00    3rd Qu.:5.00    3rd Qu.:6.00
Max.   :6.00    Max.   :6.00    Max.   :6.00    Max.   :6.00
NA's   :55114   NA's   :58053   NA's   :59643   NA's   :59731
  ST164Q05IA      ST164Q06IA      ST165Q01IA      ST165Q02IA
Min.   :1.00    Min.   :1.00    Min.   :1.00    Min.   :1.00
1st Qu.:3.00    1st Qu.:2.00    1st Qu.:2.00    1st Qu.:2.00
Median :5.00    Median :3.00    Median :3.00    Median :3.00
Mean   :4.28    Mean   :3.19    Mean   :3.53    Mean   :2.84
3rd Qu.:6.00    3rd Qu.:5.00    3rd Qu.:5.00    3rd Qu.:4.00
Max.   :6.00    Max.   :6.00    Max.   :6.00    Max.   :6.00
NA's   :59651   NA's   :58931   NA's   :59850   NA's   :63228
  ST165Q03IA      ST165Q04IA      ST165Q05IA      ST166Q01HA
Min.   :1.00    Min.   :1.00    Min.   :1.00    Min.   :1.00
1st Qu.:3.00    1st Qu.:3.00    1st Qu.:3.00    1st Qu.:1.00
Median :4.00    Median :5.00    Median :5.00    Median :3.00
Mean   :3.84    Mean   :4.41    Mean   :4.39    Mean   :3.01
3rd Qu.:5.00    3rd Qu.:6.00    3rd Qu.:6.00    3rd Qu.:4.00
Max.   :6.00    Max.   :6.00    Max.   :6.00    Max.   :6.00
NA's   :64864   NA's   :63680   NA's   :62473   NA's   :63099
  ST166Q02HA      ST166Q03HA      ST166Q04HA      ST166Q05HA
Min.   :1.00    Min.   :1.00    Min.   :1.00    Min.   :1.00
1st Qu.:3.00    1st Qu.:1.00    1st Qu.:2.00    1st Qu.:2.00
Median :4.00    Median :2.00    Median :3.00    Median :4.00
Mean   :4.07    Mean   :2.61    Mean   :3.21    Mean   :3.94
3rd Qu.:6.00    3rd Qu.:4.00    3rd Qu.:5.00    3rd Qu.:6.00
Max.   :6.00    Max.   :6.00    Max.   :6.00    Max.   :6.00
NA's   :67502   NA's   :68956   NA's   :68455   NA's   :67116
    UNDREM          METASUM         METASPAM
Min.   :-1.64   Min.   :-1.72   Min.   :-1.41
1st Qu.:-0.94   1st Qu.:-0.95   1st Qu.:-1.41
Median : 0.10   Median : 0.21   Median :-0.04
Mean   :-0.08   Mean   :-0.14   Mean   :-0.16
3rd Qu.: 0.80   3rd Qu.: 0.59   3rd Qu.: 0.42
Max.   : 1.50   Max.   : 1.36   Max.   : 1.33
NA's   :77626   NA's   :77131   NA's   :85033
```

iii. Correlation Matrix

	ST164Q01IA	ST164Q02IA	ST164Q03IA	ST164Q04IA	ST164Q05IA	ST164Q06IA
ST164Q01IA	1.000	0.411	0.258	0.266	0.233	0.162
ST164Q02IA	0.411	1.000	0.253	0.217	0.197	0.178
ST164Q03IA	0.258	0.253	1.000	0.473	0.460	0.386
ST164Q04IA	0.266	0.217	0.473	1.000	0.577	0.337
ST164Q05IA	0.233	0.197	0.460	0.577	1.000	0.390
ST164Q06IA	0.162	0.178	0.386	0.337	0.390	1.000
ST165Q01IA	0.315	0.258	0.346	0.374	0.380	0.256
ST165Q02IA	0.315	0.321	0.179	0.164	0.142	0.227
ST165Q03IA	0.299	0.286	0.352	0.396	0.369	0.281

ST165Q04IA	0.250	0.197	0.399	0.481	0.475	0.249
ST165Q05IA	0.223	0.172	0.389	0.585	0.502	0.291
ST166Q01HA	0.218	0.190	0.166	0.151	0.148	0.150
ST166Q02HA	0.177	0.144	0.274	0.281	0.277	0.146
ST166Q03HA	0.219	0.225	0.101	0.090	0.081	0.172
ST166Q04HA	0.076	0.105	0.129	0.095	0.111	0.108
ST166Q05HA	0.128	0.108	0.245	0.247	0.254	0.168
UNDREM	-0.287	-0.327	0.347	0.449	0.440	-0.125
METASUM	-0.085	-0.122	0.121	0.249	0.231	0.013
METASPAM	-0.087	-0.088	0.094	0.117	0.128	-0.003

	ST165Q01IA	ST165Q02IA	ST165Q03IA	ST165Q04IA	ST165Q05IA	ST166Q01HA
ST164Q01IA	0.315	0.315	0.299	0.250	0.223	0.218
ST164Q02IA	0.258	0.321	0.286	0.197	0.172	0.190
ST164Q03IA	0.346	0.179	0.352	0.399	0.389	0.166
ST164Q04IA	0.374	0.164	0.396	0.481	0.585	0.151
ST164Q05IA	0.380	0.142	0.369	0.475	0.502	0.148
ST164Q06IA	0.256	0.227	0.281	0.249	0.291	0.150
ST165Q01IA	1.000	0.393	0.458	0.478	0.416	0.221
ST165Q02IA	0.393	1.000	0.381	0.193	0.186	0.244
ST165Q03IA	0.458	0.381	1.000	0.543	0.477	0.209
ST165Q04IA	0.478	0.193	0.543	1.000	0.650	0.152
ST165Q05IA	0.416	0.186	0.477	0.650	1.000	0.152
ST166Q01HA	0.221	0.244	0.209	0.152	0.152	1.000
ST166Q02HA	0.279	0.083	0.258	0.372	0.315	0.374
ST166Q03HA	0.166	0.338	0.182	0.042	0.075	0.506
ST166Q04HA	0.127	0.093	0.102	0.156	0.130	-0.018
ST166Q05HA	0.231	0.081	0.229	0.326	0.294	0.324
UNDREM	0.073	-0.194	0.086	0.276	0.316	-0.069
METASUM	-0.081	-0.502	-0.013	0.442	0.470	-0.107
METASPAM	0.019	-0.199	0.025	0.215	0.170	-0.405

	ST166Q02HA	ST166Q03HA	ST166Q04HA	ST166Q05HA	UNDREM	METASUM	METASPAM
ST164Q01IA	0.177	0.219	0.076	0.128	-0.287	-0.085	-0.087
ST164Q02IA	0.144	0.225	0.105	0.108	-0.327	-0.122	-0.088
ST164Q03IA	0.274	0.101	0.129	0.245	0.347	0.121	0.094
ST164Q04IA	0.281	0.090	0.095	0.247	0.449	0.249	0.117
ST164Q05IA	0.277	0.081	0.111	0.254	0.440	0.231	0.128
ST164Q06IA	0.146	0.172	0.108	0.168	-0.125	0.013	-0.003
ST165Q01IA	0.279	0.166	0.127	0.231	0.073	-0.081	0.019
ST165Q02IA	0.083	0.338	0.093	0.081	-0.194	-0.502	-0.199
ST165Q03IA	0.258	0.182	0.102	0.229	0.086	-0.013	0.025
ST165Q04IA	0.372	0.042	0.156	0.326	0.276	0.442	0.215
ST165Q05IA	0.315	0.075	0.130	0.294	0.316	0.470	0.170
ST166Q01HA	0.374	0.506	-0.018	0.324	-0.069	-0.107	-0.405
ST166Q02HA	1.000	0.151	0.224	0.549	0.159	0.167	0.362
ST166Q03HA	0.151	1.000	0.019	0.157	-0.195	-0.239	-0.518
ST166Q04HA	0.224	0.019	1.000	0.225	0.018	0.036	0.384
ST166Q05HA	0.549	0.157	0.225	1.000	0.147	0.160	0.398
UNDREM	0.159	-0.195	0.018	0.147	1.000	0.465	0.317
METASUM	0.167	-0.239	0.036	0.160	0.465	1.000	0.390
METASPAM	0.362	-0.518	0.384	0.398	0.317	0.390	1.000

2. Creating Survey Design and Analyzing Effective Sample Size

```
options(scipen = 999)
# Define the survey design with the weights
design <- svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)
summary(design)
```

```
Independent Sampling design (with replacement)
svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)
Probabilities:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0003394 0.0227413 0.0898264 0.1939754 0.2021693 1.0000000
Data variables:
 [1] "CNTRYID"    "CNTSCHID"   "CNTSTUID"   "ST164Q01IA" "ST164Q02IA"
 [6] "ST164Q03IA" "ST164Q04IA" "ST164Q05IA" "ST164Q06IA" "ST165Q01IA"
[11] "ST165Q02IA" "ST165Q03IA" "ST165Q04IA" "ST165Q05IA" "ST166Q01HA"
[16] "ST166Q02HA" "ST166Q03HA" "ST166Q04HA" "ST166Q05HA" "UNDREM"
[21] "METASUM"    "METASPAM"   "W_FSTUWT"   "PV1READ"    "PV2READ"
[26] "PV3READ"    "PV4READ"    "PV5READ"    "PV6READ"    "PV7READ"
[31] "PV8READ"    "PV9READ"    "PV10READ"
```

```
weights_vector <- weights(design)

# Calculate effective sample size for a variable
effsize <- sum(weights_vector)^2 / sum(weights_vector^2)
effsize
```

```
[1] 92777.77
```

In this study, data were analyzed from a total sample of 612,004 participants. Given the complex survey design, individual responses were weighted to account for variability in representation across the study population. The weighting process adjusts for over- or under-representation of specific segments within the sample, ensuring that our estimates more accurately reflect the target population (Gard et al., 2023). To quantify the impact of these survey weights on our analysis, we calculated the effective sample size, which considers the distribution of the survey weights and their contribution to the variance of our estimates.

The effective sample size was determined to be approximately 92,778, a figure representing the equivalent number of equally weighted observations necessary to achieve a similar level of precision in our estimates. This discrepancy between the total and effective sample sizes underscores the significance of the survey weights in our analysis, indicating that, due to the weighted survey design, the actual amount of independent information available for analysis is akin to having 92,778 equally weighted observations (Heeringa et al., 2017).

3. Sampling with PV1READ Model

When dealing with plausible values like those in the PISA dataset, averaging them isn't generally recommended. This is because plausible values aren't "missing data imputed" but are drawn from a posterior distribution of proficiency, given the test data. Averaging them could result in misleading inference.

To address this, the OECD's method for analyzing plausible values is to run the analyses separately for each plausible value and then average the results of those analyses. This approach retains the variance within each plausible value.

The use of sampling weights is essential in survey data analysis, as it ensures that the sample is representative of the target population. These weights, like W_FSTUWT in PISA, account for the complex sampling design, oversampling, and non-response.

For a weighted multiple regression in R, we could use the survey package. Here's an example of how we might structure your analysis for one plausible value (PV1READ):

```
Call:
svyglm(formula = PV1READ ~ UNDREM + METASUM + METASPAM, design = design)

Survey design:
svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept) 472.2776     0.3635 1299.24 <0.0000000000000002 ***
UNDREM       16.5291     0.4053   40.78 <0.0000000000000002 ***
METASUM      23.9883     0.4111   58.35 <0.0000000000000002 ***
METASPAM     37.4345     0.3919   95.52 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 8332.086)

Number of Fisher Scoring iterations: 2
```

The output shown is a summary of the weighted linear regression model you've fit using the **svyglm** function from the survey package. This model predicts `PV1READ` (a plausible value of reading score) based on three predictor variables: `UNDREM`, `METASUM`, and `METASPAM`. Here's how to interpret the output:

The Estimate column provides the coefficients of PV1 model. The intercept, 472.2776, is the expected value of PV1READ when all predictor variables are 0. The other estimates tell us how a one-unit change in the corresponding predictor variable is associated with a change in the PV1READ score, assuming all other variables are held constant.

- For UNDREM, an increase by 1 unit is associated with an increase in PV1READ by 16.5291 units.
- For METASUM, an increase by 1 unit is associated with an increase in PV1READ by 23.9883 units.
- For METASPAM, an increase by 1 unit is associated with an increase in PV1READ by 37.4345 units.

The Dispersion parameter for the Gaussian family is the estimated scale parameter, equivalent to the estimated variance of the errors in a classical linear regression model. In this case, the dispersion parameter is 8332.086. It's important to note that this is an absolute measure, and its interpretation depends on the scale of our outcome variable (PV1READ). The root of the dispersion parameter can give us an estimate of the average absolute deviation (which, in our case, would be the square root of 8332.086), which is ~ 91.28.

4. Modeling Weighted Survey Generalized Linear Models

A. Predicting Reading Scores by Meta-Cognitive Reading Skills

i. Full Model

```
# A tibble: 4 x 4
  term        average_estimate average_std_error average_p_value
  <chr>                  <dbl>             <dbl>           <dbl>
1 (Intercept)            472.              0.363               0
2 METASPAM                37.4             0.391               0
3 METASUM                 24.0             0.410               0
4 UNDREM                  16.5             0.404               0
```

ii. Null Model

```
# A tibble: 1 x 2
  term        average_estimate
  <chr>                  <dbl>
1 (Intercept)             447.
```

iii. Model Comparison

```
Working (Rao-Scott) LRT for UNDREM METASUM METASPAM
 in svyglm(formula = as.formula(paste(.x, "~ UNDREM + METASUM + METASPAM")),
    design = design)
Working 2logLR =  30844.47 p= < 0.000000000000000222
(scale factors:  1.1 1 0.9 )
```

B. Predicting Reading Scores by Reading Strategies

i . Full Model

```
# A tibble: 17 x 5
   term        Estimate `Std. Error` `t value` `Pr(>|t|)`
   <chr>          <dbl>        <dbl>     <dbl>      <dbl>
 1 (Intercept) 377.            1.46    258.     0
 2 ST164Q01IA   -2.14          0.246    -8.67   3.37e- 17
 3 ST164Q02IA   -0.0807        0.240    -0.335  7.32e-  1
 4 ST164Q03IA    5.10          0.254    20.1    2.85e- 87
 5 ST164Q04IA    0.402         0.292     1.38   1.79e-  1
 6 ST164Q05IA    4.89          0.292    16.8    8.78e- 62
 7 ST164Q06IA   -5.12          0.233   -22.0    1.60e-102
 8 ST165Q01IA    0.523         0.269     1.95   5.75e-  2
 9 ST165Q02IA  -13.6           0.266   -51.3    0
10 ST165Q03IA   -3.56          0.285   -12.5    3.80e- 34
11 ST165Q04IA   13.9           0.347    40.1    0
12 ST165Q05IA    3.26          0.326     9.99   1.14e- 22
13 ST166Q01HA   -2.34          0.246    -9.50   1.01e- 20
14 ST166Q02HA   14.9           0.257    58.1    0
15 ST166Q03HA  -20.2           0.252   -80.1    0
16 ST166Q04HA    6.68          0.205    32.6    8.98e-230
17 ST166Q05HA    4.64          0.232    20.0    2.91e- 82
```

ii. UNDREM Variables and Reading Scores Model

```
# A tibble: 7 x 5
  term        Estimate `Std. Error` `t value` `Pr(>|t|)`
  <chr>          <dbl>        <dbl>     <dbl>      <dbl>
1 (Intercept)   407.          1.49      273.   0
2 ST164Q01IA     -5.80        0.289     -20.1  2.04e- 87
3 ST164Q02IA     -5.86        0.283     -20.7  9.95e- 93
4 ST164Q03IA     10.4         0.300      34.6  5.81e-258
5 ST164Q04IA      6.43        0.323      19.9  1.21e- 85
6 ST164Q05IA     12.3         0.332      37.0  2.35e-295
7 ST164Q06IA     -9.93        0.275     -36.1  1.88e-278
```

iii. METASUM Variables and Reading Scores Model

```
# A tibble: 6 x 5
  term        Estimate `Std. Error` `t value` `Pr(>|t|)`
  <chr>          <dbl>        <dbl>     <dbl>      <dbl>
1 (Intercept)   395.          1.30     305.     0
2 ST165Q01IA      2.69        0.300      8.94   2.12e-18
3 ST165Q02IA    -22.4         0.278    -80.6    0
4 ST165Q03IA     -4.79        0.324    -14.8    9.50e-48
5 ST165Q04IA     23.2         0.367     63.2    0
6 ST165Q05IA      7.02        0.332     21.2    6.30e-97
```

iv. METASPAM Variables and Reading Scores Model

```
# A tibble: 6 x 5
  term        Estimate `Std. Error` `t value` `Pr(>|t|)`
  <chr>          <dbl>        <dbl>     <dbl>      <dbl>
1 (Intercept)   402.          1.12      360.   0
2 ST166Q01HA     -4.01        0.246     -16.3  3.63e- 58
3 ST166Q02HA     20.4         0.251      81.4  0
4 ST166Q03HA    -24.8         0.243    -102.   0
5 ST166Q04HA      6.44        0.213      30.2  2.29e-197
6 ST166Q05HA      7.35        0.234      31.4  1.48e-207
```

C. Metacognitive Skill Models

i. Predicting Understanding and Remembering Using Associated Reading Strategies

```
Call:
svyglm(formula = UNDREM ~ ST164Q01IA + ST164Q02IA + ST164Q03IA +
    ST164Q04IA + ST164Q05IA + ST164Q06IA, design = design)

Survey design:
svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
(Intercept) -0.557857   0.009046  -61.67 <0.0000000000000002 ***
ST164Q01IA  -0.218367   0.001284 -170.04 <0.0000000000000002 ***
ST164Q02IA  -0.226458   0.001339 -169.10 <0.0000000000000002 ***
```

```
ST164Q03IA   0.191887   0.001213  158.14 <0.0000000000000002 ***
ST164Q04IA   0.234926   0.001281  183.44 <0.0000000000000002 ***
ST164Q05IA   0.227304   0.001312  173.22 <0.0000000000000002 ***
ST164Q06IA  -0.232416   0.001308 -177.75 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.3039362)

Number of Fisher Scoring iterations: 2


The R-squared value for the undrem_model is 0.703
```

ii. Metacognitive Awareness of Summarizing

```
Call:
svyglm(formula = METASUM ~ ST165Q01IA + ST165Q02IA + ST165Q03IA +
    ST165Q04IA + ST165Q05IA, design = design)

Survey design:
svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
(Intercept) -0.799123   0.008344  -95.77 <0.0000000000000002 ***
ST165Q01IA  -0.120372   0.001488  -80.87 <0.0000000000000002 ***
ST165Q02IA  -0.345791   0.001488 -232.33 <0.0000000000000002 ***
ST165Q03IA  -0.111643   0.001600  -69.78 <0.0000000000000002 ***
ST165Q04IA   0.284037   0.001736  163.63 <0.0000000000000002 ***
ST165Q05IA   0.276417   0.001500  184.28 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.3347374)

Number of Fisher Scoring iterations: 2


The R-squared value for the undrem_model is 0.675
```

iii. Metacognitive Awareness of Assessing Credibility

```
Call:
svyglm(formula = METASPAM ~ ST166Q01HA + ST166Q02HA + ST166Q03HA +
    ST166Q04HA + ST166Q05HA, design = design)

Survey design:
svydesign(ids = ~1, data = meta_read_data, weights = ~W_FSTUWT)

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
```

```
(Intercept) -0.810046   0.006772  -119.6 <0.0000000000000002 ***
ST166Q01HA  -0.246523   0.001125  -219.2 <0.0000000000000002 ***
ST166Q02HA   0.195534   0.001162   168.3 <0.0000000000000002 ***
ST166Q03HA  -0.237765   0.001165  -204.1 <0.0000000000000002 ***
ST166Q04HA   0.123404   0.001026   120.3 <0.0000000000000002 ***
ST166Q05HA   0.201061   0.001064   188.9 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2627529)

Number of Fisher Scoring iterations: 2


The R-squared value for the undrem_model is 0.736
```

5. Pairwise Correlation Analysis

A. UNDREM vs. Reading Strategies and Reading Scores

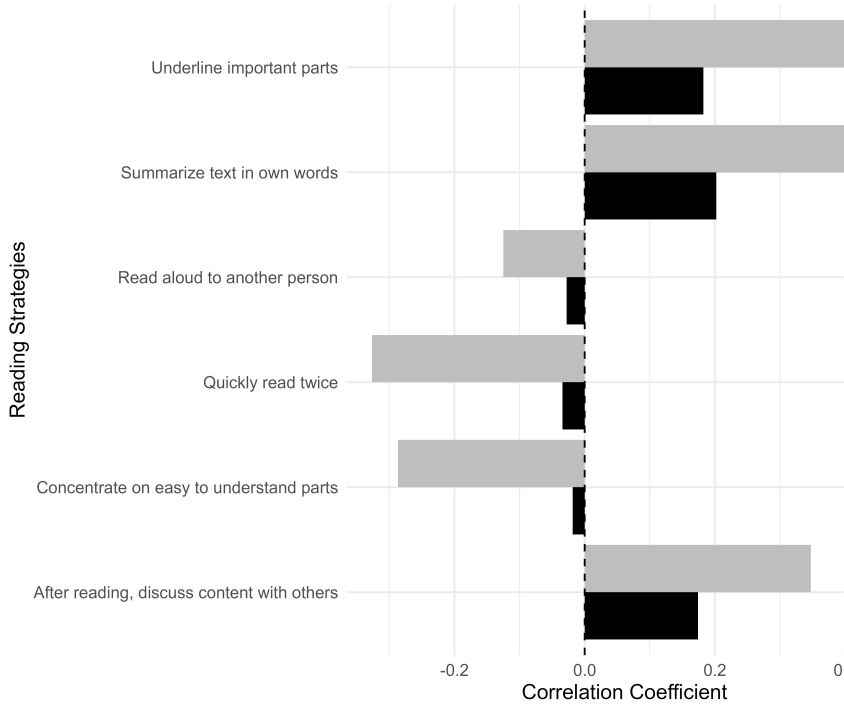
- *UNDREM* = "Metacognitive awareness of understanding and remembering";
- *ST164Q01IA* = "I concentrate on the parts of the text that are easy to understand";
- *ST164Q02IA* = "I quickly read through the text twice";
- *ST164Q03IA* = "After reading the text, I discuss its content with other people";
- *ST164Q04IA* = "I underline important parts of the text";
- *ST164Q05IA* = "I summarize the text in my own words";
- *ST164Q06IA* = "I read the text aloud to another person",

i. UNDREM vs. Reading Strategies

```
UNDREM ~ ST164Q01IA : Correlation = -0.2867328 p-value = 0
UNDREM ~ ST164Q02IA : Correlation = -0.3266454 p-value = 0
UNDREM ~ ST164Q03IA : Correlation = 0.3472629 p-value = 0
UNDREM ~ ST164Q04IA : Correlation = 0.4490406 p-value = 0
UNDREM ~ ST164Q05IA : Correlation = 0.4396146 p-value = 0
UNDREM ~ ST164Q06IA : Correlation = -0.1247665 p-value = 0
```

ii. Reading Strategies Vs. Reading Scores

```
 ST164Q01IA  ST164Q02IA  ST164Q03IA  ST164Q04IA  ST164Q05IA  ST164Q06IA
-0.01833729 -0.03404063  0.17417845  0.18238375  0.20229358 -0.02774301
```



iii. Plotting the Correlations for UNDREM

B. METASUM vs. Reading Strategies and Reading Scores

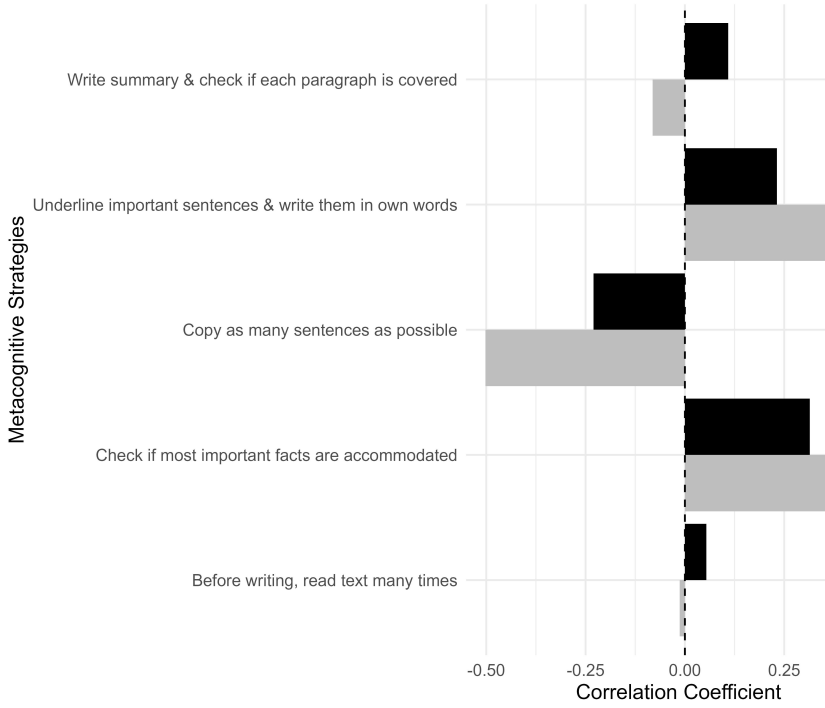
- *METASUM* = "Metacognitive awareness of summarizing";
- *ST165Q01IA* = "I write a summary. Then I check that each paragraph is covered in the summary, because the content of each paragraph should be included";
- *ST165Q02IA* = "I try to copy out accurately as many sentences as possible";
- *ST165Q03IA* = "Before writing the summary, I read the text as many times as possible";
- *ST165Q04IA* = "I carefully check whether the most important facts in the text are represented in the summary";
- *ST165Q05IA* = "I read through the text, underlining the most important sentences. Then I write them in my own words as a summary";

i. METASUM vs. Reading Strategies

```
METASUM ~ ST165Q01IA : Correlation = -0.08107716 p-value = 0
METASUM ~ ST165Q02IA : Correlation = -0.5021068 p-value = 0
METASUM ~ ST165Q03IA : Correlation = -0.01311923 p-value = 0.00000000000000000000008381113
METASUM ~ ST165Q04IA : Correlation = 0.4421182 p-value = 0
METASUM ~ ST165Q05IA : Correlation = 0.4701468 p-value = 0
```

ii. METASUM vs. Reading Scores

```
 ST165Q01IA  ST165Q02IA  ST165Q03IA  ST165Q04IA  ST165Q05IA
 0.10879998 -0.22987239  0.05402973  0.31454011  0.23178539
```



iii. Plotting the Correlations for METASUM

C. METASPAM vs. Reading Strategies and Reading Scores

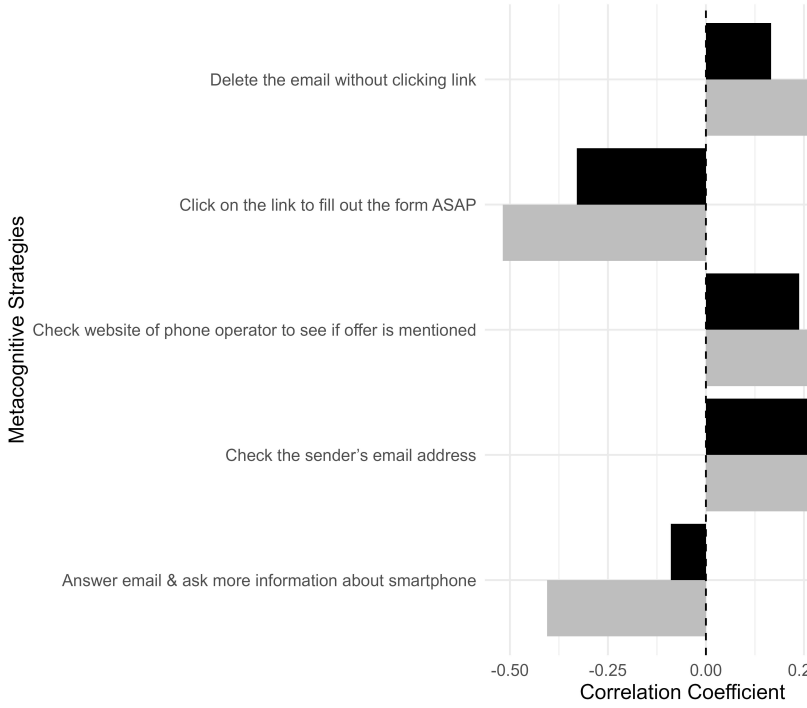
- *METASPAM* = "Metacognitive awareness of assessing credibility";
- *ST166Q01HA* = "Answer the email and ask for more information about the smartphone";
- *ST166Q02HA* = "Check the sender's email address";
- *ST166Q03HA* = "Click on the link to fill out the form as soon as possible";
- *ST166Q04HA* = "Delete the email without clicking on the link";
- *ST166Q05HA* = "Check the website of the mobile phone operator to see whether the smartphone offer is mentioned"

i. METASPAM vs. Reading Strategies

```
METASPAM ~ ST166Q01HA : Correlation = -0.4050722 p-value = 0
METASPAM ~ ST166Q02HA : Correlation = 0.3620675 p-value = 0
METASPAM ~ ST166Q03HA : Correlation = -0.5179707 p-value = 0
METASPAM ~ ST166Q04HA : Correlation = 0.3835217 p-value = 0
METASPAM ~ ST166Q05HA : Correlation = 0.3983277 p-value = 0
```

ii. METASPAM vs. Reading Scores

```
 ST166Q01HA  ST166Q02HA  ST166Q03HA  ST166Q04HA  ST166Q05HA
-0.08965924  0.31437917 -0.32927337  0.16644737  0.23787345
```



iii. Plotting the Correlations for METASPAM

D. Metacognitive Reading Skills and Reading Scores

```
   UNDREM   METASUM  METASPAM
0.3470385 0.4070833 0.4435966
```

6. Decomposition of Categorical Variables Regression Models

A. Metacognition Skill Models

i. Understanding and Remembering

```
Call:
lm(formula = UNDREM ~ ST164Q01IA + ST164Q02IA + ST164Q03IA +
    ST164Q04IA + ST164Q05IA + ST164Q06IA, data = undrem_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.37294 -0.27165  0.04536  0.30361  2.80628

Coefficients:
             Estimate Std. Error t value             Pr(>|t|)
(Intercept) -1.056654   0.003040 -347.58 < 0.0000000000000002 ***
ST164Q01IA2 -0.075835   0.002616  -28.99 < 0.0000000000000002 ***
ST164Q01IA3 -0.238032   0.002504  -95.05 < 0.0000000000000002 ***
ST164Q01IA4 -0.438648   0.002584 -169.74 < 0.0000000000000002 ***
ST164Q01IA5 -0.679463   0.002855 -237.97 < 0.0000000000000002 ***
ST164Q01IA6 -1.040817   0.002740 -379.91 < 0.0000000000000002 ***
```

```
ST164Q02IA2 -0.015482   0.002270   -6.82     0.00000000000911 ***
ST164Q02IA3 -0.232372   0.002324 -100.00 < 0.0000000000000002 ***
ST164Q02IA4 -0.462376   0.002452 -188.56 < 0.0000000000000002 ***
ST164Q02IA5 -0.682210   0.002720 -250.77 < 0.0000000000000002 ***
ST164Q02IA6 -1.113324   0.002781 -400.28 < 0.0000000000000002 ***
ST164Q03IA2  0.238561   0.002804   85.07 < 0.0000000000000002 ***
ST164Q03IA3  0.450372   0.002767  162.77 < 0.0000000000000002 ***
ST164Q03IA4  0.681296   0.002796  243.65 < 0.0000000000000002 ***
ST164Q03IA5  0.892551   0.002897  308.11 < 0.0000000000000002 ***
ST164Q03IA6  0.967161   0.002855  338.79 < 0.0000000000000002 ***
ST164Q04IA2  0.239988   0.003625   66.19 < 0.0000000000000002 ***
ST164Q04IA3  0.467979   0.003513  133.19 < 0.0000000000000002 ***
ST164Q04IA4  0.784974   0.003416  229.80 < 0.0000000000000002 ***
ST164Q04IA5  1.024071   0.003408  300.50 < 0.0000000000000002 ***
ST164Q04IA6  1.166283   0.003299  353.57 < 0.0000000000000002 ***
ST164Q05IA2  0.267879   0.003724   71.94 < 0.0000000000000002 ***
ST164Q05IA3  0.484286   0.003599  134.57 < 0.0000000000000002 ***
ST164Q05IA4  0.790146   0.003526  224.07 < 0.0000000000000002 ***
ST164Q05IA5  1.069484   0.003499  305.66 < 0.0000000000000002 ***
ST164Q05IA6  1.181320   0.003448  342.63 < 0.0000000000000002 ***
ST164Q06IA2 -0.191172   0.002227  -85.83 < 0.0000000000000002 ***
ST164Q06IA3 -0.404867   0.002296 -176.36 < 0.0000000000000002 ***
ST164Q06IA4 -0.603983   0.002410 -250.63 < 0.0000000000000002 ***
ST164Q06IA5 -0.834834   0.002598 -321.29 < 0.0000000000000002 ***
ST164Q06IA6 -1.235487   0.002490 -496.16 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4844 on 534347 degrees of freedom
  (77626 observations deleted due to missingness)
Multiple R-squared:  0.7649,    Adjusted R-squared:  0.7649
F-statistic: 5.796e+04 on 30 and 534347 DF,  p-value: < 0.00000000000000022
```

ii. Summarizing

```
Call:
lm(formula = METASUM ~ ST165Q01IA + ST165Q02IA + ST165Q03IA +
    ST165Q04IA + ST165Q05IA, data = metasum_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19456 -0.32371  0.04747  0.31179  2.67864

Coefficients:
             Estimate Std. Error  t value             Pr(>|t|)
(Intercept) -1.385818   0.003102 -446.761 < 0.0000000000000002 ***
ST165Q01IA2 -0.001651   0.002720   -0.607              0.54391
ST165Q01IA3 -0.075451   0.002687  -28.079 < 0.0000000000000002 ***
ST165Q01IA4 -0.142884   0.002780  -51.395 < 0.0000000000000002 ***
ST165Q01IA5 -0.266401   0.002970  -89.696 < 0.0000000000000002 ***
ST165Q01IA6 -0.568927   0.002937 -193.715 < 0.0000000000000002 ***
ST165Q02IA2 -0.096819   0.002072  -46.736 < 0.0000000000000002 ***
```

```
ST165Q02IA3 -0.456021   0.002211 -206.295 < 0.0000000000000002 ***
ST165Q02IA4 -0.855805   0.002440 -350.799 < 0.0000000000000002 ***
ST165Q02IA5 -1.216336   0.002879 -422.510 < 0.0000000000000002 ***
ST165Q02IA6 -1.803360   0.003041 -593.049 < 0.0000000000000002 ***
ST165Q03IA2  0.117271   0.003457   33.924 < 0.0000000000000002 ***
ST165Q03IA3  0.077216   0.003373   22.894 < 0.0000000000000002 ***
ST165Q03IA4  0.009804   0.003417    2.869              0.00412 **
ST165Q03IA5 -0.114456   0.003529  -32.429 < 0.0000000000000002 ***
ST165Q03IA6 -0.457561   0.003547 -129.014 < 0.0000000000000002 ***
ST165Q04IA2  0.394345   0.004595   85.822 < 0.0000000000000002 ***
ST165Q04IA3  0.712492   0.004488  158.760 < 0.0000000000000002 ***
ST165Q04IA4  1.107345   0.004401  251.597 < 0.0000000000000002 ***
ST165Q04IA5  1.409993   0.004408  319.848 < 0.0000000000000002 ***
ST165Q04IA6  1.576456   0.004379  359.987 < 0.0000000000000002 ***
ST165Q05IA2  0.184582   0.004077   45.268 < 0.0000000000000002 ***
ST165Q05IA3  0.377337   0.003962   95.245 < 0.0000000000000002 ***
ST165Q05IA4  0.722075   0.003908  184.762 < 0.0000000000000002 ***
ST165Q05IA5  1.047580   0.003858  271.566 < 0.0000000000000002 ***
ST165Q05IA6  1.242921   0.003754  331.088 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4912 on 534846 degrees of freedom
  (77132 observations deleted due to missingness)
Multiple R-squared:  0.7588,    Adjusted R-squared:  0.7588
F-statistic: 6.732e+04 on 25 and 534846 DF,  p-value: < 0.00000000000000022
```

iii. Assessing Credibility

```
Call:
lm(formula = METASPAM ~ ST166Q01HA + ST166Q02HA + ST166Q03HA +
    ST166Q04HA + ST166Q05HA, data = metaspm_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.02742 -0.33129  0.00248  0.32268  2.21306

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
(Intercept) -1.021562   0.002227 -458.68 <0.0000000000000002 ***
ST166Q01HA2 -0.176914   0.002275  -77.78 <0.0000000000000002 ***
ST166Q01HA3 -0.479399   0.002359 -203.20 <0.0000000000000002 ***
ST166Q01HA4 -0.725774   0.002526 -287.32 <0.0000000000000002 ***
ST166Q01HA5 -0.931450   0.002842 -327.79 <0.0000000000000002 ***
ST166Q01HA6 -1.281677   0.002606 -491.75 <0.0000000000000002 ***
ST166Q02HA2  0.432498   0.003138  137.82 <0.0000000000000002 ***
ST166Q02HA3  0.619102   0.003089  200.39 <0.0000000000000002 ***
ST166Q02HA4  0.887752   0.003059  290.24 <0.0000000000000002 ***
ST166Q02HA5  1.068204   0.003096  345.00 <0.0000000000000002 ***
ST166Q02HA6  1.111573   0.002788  398.63 <0.0000000000000002 ***
ST166Q03HA2 -0.166350   0.002236  -74.40 <0.0000000000000002 ***
ST166Q03HA3 -0.466319   0.002298 -202.93 <0.0000000000000002 ***
```

```
ST166Q03HA4 -0.772451   0.002516 -307.01 <0.0000000000000002 ***
ST166Q03HA5 -1.031412   0.002948 -349.81 <0.0000000000000002 ***
ST166Q03HA6 -1.319816   0.002848 -463.39 <0.0000000000000002 ***
ST166Q04HA2  0.153460   0.002182   70.33 <0.0000000000000002 ***
ST166Q04HA3  0.277912   0.002230  124.63 <0.0000000000000002 ***
ST166Q04HA4  0.484154   0.002476  195.56 <0.0000000000000002 ***
ST166Q04HA5  0.603372   0.002669  226.06 <0.0000000000000002 ***
ST166Q04HA6  0.579282   0.002149  269.54 <0.0000000000000002 ***
ST166Q05HA2  0.256510   0.002919   87.88 <0.0000000000000002 ***
ST166Q05HA3  0.447597   0.002797  160.04 <0.0000000000000002 ***
ST166Q05HA4  0.706357   0.002803  251.96 <0.0000000000000002 ***
ST166Q05HA5  0.930271   0.002806  331.54 <0.0000000000000002 ***
ST166Q05HA6  0.992001   0.002490  398.36 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.471 on 526944 degrees of freedom
  (85034 observations deleted due to missingness)
Multiple R-squared:  0.7715,    Adjusted R-squared:  0.7714
F-statistic: 7.115e+04 on 25 and 526944 DF,  p-value: < 0.00000000000000022
```

B. Reading Scores Model

i. Understanding and Remembering

	term	estimate	std.error	conf.low	conf.high	p.value
1	(Intercept)	408.79	0.62	407.58	410.01	0.0000
2	ST164Q01IA2	12.85	0.53	11.81	13.90	0.0000
3	ST164Q01IA3	18.13	0.51	17.14	19.13	0.0000
4	ST164Q01IA4	11.56	0.53	10.52	12.59	0.0000
5	ST164Q01IA5	2.18	0.58	1.04	3.32	0.0003
6	ST164Q01IA6	-20.86	0.56	-21.96	-19.77	0.0000
7	ST164Q02IA2	-5.80	0.46	-6.71	-4.90	0.0000
8	ST164Q02IA3	-5.47	0.47	-6.40	-4.54	0.0000
9	ST164Q02IA4	-7.71	0.50	-8.69	-6.73	0.0000
10	ST164Q02IA5	-12.31	0.56	-13.40	-11.22	0.0000
11	ST164Q02IA6	-27.53	0.57	-28.65	-26.42	0.0000
12	ST164Q03IA2	8.71	0.57	7.59	9.83	0.0000
13	ST164Q03IA3	15.32	0.56	14.22	16.43	0.0000
14	ST164Q03IA4	31.29	0.57	30.17	32.41	0.0000
15	ST164Q03IA5	42.65	0.59	41.49	43.81	0.0000
16	ST164Q03IA6	40.83	0.58	39.69	41.97	0.0000
17	ST164Q04IA2	-5.49	0.74	-6.93	-4.04	0.0000
18	ST164Q04IA3	4.75	0.72	3.35	6.16	0.0000
19	ST164Q04IA4	18.27	0.70	16.90	19.63	0.0000
20	ST164Q04IA5	33.07	0.69	31.71	34.44	0.0000
21	ST164Q04IA6	23.29	0.67	21.97	24.61	0.0000
22	ST164Q05IA2	2.87	0.76	1.38	4.36	0.0003
23	ST164Q05IA3	19.22	0.73	17.78	20.66	0.0000
24	ST164Q05IA4	35.25	0.72	33.84	36.66	0.0000
25	ST164Q05IA5	44.84	0.71	43.44	46.24	0.0000
26	ST164Q05IA6	50.58	0.70	49.20	51.96	0.0000
27	ST164Q06IA2	-3.60	0.45	-4.49	-2.71	0.0000

```
28 ST164Q06IA3    -9.31      0.47   -10.23     -8.39  0.0000
29 ST164Q06IA4   -17.20      0.49   -18.17    -16.24  0.0000
30 ST164Q06IA5   -26.59      0.53   -27.63    -25.55  0.0000
31 ST164Q06IA6   -49.08      0.51   -50.08    -48.08  0.0000
```

The R-squared value for the UNDREM Read model is 0.118

The average sample size used for the UNDREM Read model is 529091

ii. Reading Model - Summarizing

```
          term estimate std.error conf.low conf.high p.value
1  (Intercept)   399.80      0.59   398.65    400.95   0.000
2  ST165Q01IA2    18.98      0.51    17.98     19.99   0.000
3  ST165Q01IA3    33.20      0.51    32.21     34.20   0.000
4  ST165Q01IA4    38.31      0.53    37.28     39.34   0.000
5  ST165Q01IA5    41.41      0.56    40.31     42.51   0.000
6  ST165Q01IA6    23.14      0.56    22.05     24.23   0.000
7  ST165Q02IA2   -21.27      0.39   -22.04    -20.50   0.000
8  ST165Q02IA3   -41.61      0.42   -42.43    -40.79   0.000
9  ST165Q02IA4   -61.53      0.46   -62.43    -60.62   0.000
10 ST165Q02IA5   -83.64      0.55   -84.71    -82.57   0.000
11 ST165Q02IA6  -103.20      0.58  -104.34   -102.07   0.000
12 ST165Q03IA2     1.88      0.65     0.60      3.16   0.006
13 ST165Q03IA3     0.65      0.64    -0.60      1.90   0.315
14 ST165Q03IA4     1.32      0.65     0.05      2.58   0.051
15 ST165Q03IA5    -5.28      0.67    -6.59     -3.97   0.000
16 ST165Q03IA6   -26.29      0.67   -27.61    -24.97   0.000
17 ST165Q04IA2     4.01      0.87     2.31      5.72   0.000
18 ST165Q04IA3    32.86      0.85    31.19     34.52   0.000
19 ST165Q04IA4    56.61      0.83    54.98     58.25   0.000
20 ST165Q04IA5    86.46      0.84    84.82     88.10   0.000
21 ST165Q04IA6   101.28      0.83    99.65    102.91   0.000
22 ST165Q05IA2    -1.28      0.77    -2.79      0.24   0.121
23 ST165Q05IA3     9.10      0.75     7.63     10.57   0.000
24 ST165Q05IA4    17.59      0.74    16.14     19.04   0.000
25 ST165Q05IA5    22.34      0.73    20.91     23.77   0.000
26 ST165Q05IA6    22.39      0.71    21.00     23.79   0.000
```

The R-squared value for the METASUM Read model is 0.220

The average sample size used for the METASUM Read model is 529570

iii. Reading Model - Assessing Credibility

```
          term estimate std.error conf.low conf.high p.value
1  (Intercept)   425.11      0.42   424.28    425.93   0.000
2  ST166Q01HA2     0.43      0.43    -0.41      1.28   0.342
3  ST166Q01HA3     5.68      0.45     4.80      6.56   0.000
4  ST166Q01HA4     5.24      0.48     4.30      6.18   0.000
5  ST166Q01HA5    -2.52      0.54    -3.58     -1.46   0.000
```

6	ST166Q01HA6	-24.58	0.49	-25.55	-23.62	0.000
7	ST166Q02HA2	15.19	0.59	14.03	16.35	0.000
8	ST166Q02HA3	45.96	0.58	44.81	47.11	0.000
9	ST166Q02HA4	66.62	0.58	65.48	67.75	0.000
10	ST166Q02HA5	80.38	0.59	79.23	81.53	0.000
11	ST166Q02HA6	86.21	0.53	85.17	87.24	0.000
12	ST166Q03HA2	-44.65	0.42	-45.48	-43.82	0.000
13	ST166Q03HA3	-67.06	0.44	-67.92	-66.21	0.000
14	ST166Q03HA4	-81.35	0.48	-82.29	-80.42	0.000
15	ST166Q03HA5	-97.14	0.56	-98.24	-96.05	0.000
16	ST166Q03HA6	-115.75	0.54	-116.81	-114.69	0.000
17	ST166Q04HA2	11.26	0.41	10.45	12.07	0.000
18	ST166Q04HA3	7.45	0.42	6.62	8.28	0.000
19	ST166Q04HA4	-0.82	0.47	-1.74	0.10	0.093
20	ST166Q04HA5	10.17	0.51	9.17	11.16	0.000
21	ST166Q04HA6	25.07	0.41	24.27	25.86	0.000
22	ST166Q05HA2	3.16	0.55	2.08	4.24	0.000
23	ST166Q05HA3	15.41	0.53	14.38	16.45	0.000
24	ST166Q05HA4	22.67	0.53	21.63	23.71	0.000
25	ST166Q05HA5	30.59	0.53	29.55	31.63	0.000
26	ST166Q05HA6	38.69	0.47	37.77	39.62	0.000

The R-squared value for the METASPM Read model is 0.283

The average sample size used for the METASPM Read model is 521701

iv. Average Reading Scores

	mean_read_scores
1	456.1230
2	456.1145
3	456.0685
4	456.1056
5	456.1728
6	456.2014
7	456.1219
8	456.0431
9	456.0649
10	456.0797

	SD_read_scores
1	108.0475
2	107.9959
3	108.0242
4	108.0002
5	107.9128
6	107.9264
7	108.0235
8	107.8982
9	108.0124
10	107.9995

```
   min_read_scores
1            0.000
2           28.726
3            0.341
4            0.000
5           16.891
6           31.955
7           14.165
8            0.000
9            0.000
10           0.000
```

```
   max_read_scores
1          887.692
2          898.478
3          888.223
4          885.259
5          885.244
6          873.895
7          890.932
8          928.687
9          862.252
10         884.019
```