New ICT Analysis with Plausible Values - Education Forum Manuscript

AUTHOR

Nirmal Ghimire, Ph.D. [iD]

AFFILIATION

Watson college of Education, University of North Carolina Wilmington

Load Data (ICT_data.Rdata)

```
# Load the data
load("ICT_data.Rdata")
# Check the structure of the data
str(ICT_data)
```

```
'data.frame':   4838 obs. of  20 variables:
 $ COM_HOM : int  1 0 1 1 1 1 1 1 1 1 ...
 $ INTERNET: int  2 2 2 2 2 2 2 2 2 2 ...
 $ ICTHOME : int  10 5 9 6 12 11 9 10 6 9 ...
 $ LOC_INFO: num  537 428 511 432 508 ...
 $ UNDERSTD: num  516 407 501 429 536 ...
 $ EVAL_REF: num  518 422 506 436 562 ...
 $ SINGLE  : num  559 429 508 470 518 ...
 $ MULTIPLE: num  556 422 503 445 537 ...
 $ READ_SCR: num  544 432 504 438 536 ...
 $ W_FSTUWT: num  647 630 614 614 614 ...
 $ PV1READ : num  590 395 501 464 575 ...
 $ PV2READ : num  517 389 491 458 567 ...
 $ PV3READ : num  556 424 516 424 513 ...
 $ PV4READ : num  520 473 498 431 519 ...
 $ PV5READ : num  575 480 505 436 513 ...
 $ PV6READ : num  526 413 532 436 552 ...
 $ PV7READ : num  555 457 498 424 529 ...
 $ PV8READ : num  561 429 499 429 510 ...
 $ PV9READ : num  496 410 505 415 561 ...
 $ PV10READ: num  547 453 494 461 521 ...
```

```
# Check the first few rows of the data
# head(ICT_data)
# Check the column names
colnames(ICT_data)
```

```
 [1] "COM_HOM"   "INTERNET" "ICTHOME"   "LOC_INFO" "UNDERSTD" "EVAL_REF"
 [7] "SINGLE"    "MULTIPLE" "READ_SCR"  "W_FSTUWT" "PV1READ"  "PV2READ"
[13] "PV3READ"   "PV4READ"  "PV5READ"   "PV6READ"  "PV7READ"  "PV8READ"
[19] "PV9READ"   "PV10READ"
```

```
# Check the number of rows and columns
dim(ICT_data)
```

```
[1] 4838    20
```

```
# Changing Following Variables to Factors
ICT_data$COM_HOM <- as.factor(ICT_data$COM_HOM)
summary(ICT_data$COM_HOM)
```

```
   0    1 NA's
 604 4170   64
```

```
ICT_data$INTERNET <- as.factor(ICT_data$INTERNET)
summary(ICT_data$INTERNET)
```

```
   0    1    2 NA's
 182  139 4261  256
```

Data Preparation

```
# Remove cases with missing values in key predictors (as you did before)
ICT_clean <- ICT_data %>%
  filter(!is.na(COM_HOM) & !is.na(INTERNET) & !is.na(ICTHOME)) %>%
  mutate(
    COM_HOM = factor(COM_HOM,
      levels = c("0", "1"),
      labels = c("No Computer", "Has Computer")
    ),
    INTERNET = factor(INTERNET,
      levels = c("0", "1", "2"),
      labels = c(
        "No Internet",
        "Has Internet (Not Used)",
        "Has Internet (Used)"
      )
    )
  )
```

```
# Create a list of 10 datasets, one for each plausible value
# This follows PISA technical standards
pv_datasets <- lapply(1:10, function(i) {
  # Create dataset with the i-th plausible value as the outcome
  temp_data <- ICT_clean
  temp_data$READ <- temp_data[[paste0("PV", i, "READ")]]
  return(temp_data)
})

# Function to run weighted regression on each plausible value
# This ensures we use the student weights (W_FSTUWT) properly
run_weighted_model <- function(formula_string, data_list) {
  # Run the model on each plausible value dataset
  models <- lapply(data_list, function(df) {
    # Create survey design object with weights
    design <- svydesign(
      ids = ~1, # no clustering in this analysis
      weights = ~W_FSTUWT,
      data = df
    )

    # Fit weighted linear model
    model <- svyglm(as.formula(formula_string), design = design)
    return(model)
  })

  # Extract coefficients and variance-covariance matrices
  coefs <- lapply(models, coef)
  vcovs <- lapply(models, vcov)

  # Combine results using Rubin's rules
  # This properly accounts for both sampling and imputation variance
  combined <- MIcombine(results = coefs, variances = vcovs)

  return(combined)
}
```

Run Models

Model 0: Baseline model with no predictors

```
# Baseline model with no predictors
baseline_formula <- "READ ~ 1"
baseline_results <- run_weighted_model(baseline_formula, pv_datasets)
# Display results
summary(baseline_results)

Multiple imputation results:
    MIcombine.default(results = coefs, variances = vcovs)
```

```
             results       se  (lower    upper) missInfo
(Intercept) 509.9317 1.718493 506.562 513.3014      6 %
```

```r
# Extract Model Fit Statistics
get_model_stats <- function(model_results, data_list, formula_string) {
  aics <- sapply(1:10, function(i) {
    design <- svydesign(ids = ~1, weights = ~W_FSTUWT, data = data_list[[i]])
    model <- svyglm(as.formula(formula_string), design = design)

    # Safe AIC calculation: handle null model separately
    if (length(attr(terms(model), "term.labels")) == 0) {
      # Null model: AIC = 2k - 2LL, where k = 1 (intercept)
      logL <- as.numeric(logLik(model))
      aic_val <- 2 * 1 - 2 * logL
    } else {
      # Regular model
      aic_val <- AIC(model)
    }

    return(aic_val)
  })

  return(mean(aics))
}
# display AIC
baseline_aic <- get_model_stats(baseline_results, pv_datasets, baseline_formula)
cat("\nBaseline Model AIC (Average across plausible values):", baseline_aic, "\n")
```

```
Baseline Model AIC (Average across plausible values): 51675328
```

```r
# ---- Compute pseudo R² across replicates ----
pseudo_r2 <- sapply(seq_along(pv_datasets), function(i) {
  # set up survey design for replicate i
  design_i <- svydesign(
    ids = ~1,
    weights = ~W_FSTUWT,
    data = pv_datasets[[i]]
  )
  # fit null model
  mod_i <- svyglm(READ ~ 1, design = design_i, family = gaussian())
  # deviance-based pseudo R²
  1 - (mod_i$deviance / mod_i$null.deviance)
})
mean_r2 <- mean(pseudo_r2)
sd_r2 <- sd(pseudo_r2)

cat(
  "Pseudo R² (mean ± SD across replicates):",
  sprintf("%.4f \u00B1 %.4f\n", mean_r2, sd_r2)
)
```

```
Pseudo R² (mean ± SD across replicates): 0.0000 ± 0.0000
```

```
# ---- Compute average sample size across replicates ----
# here we count non-missing READ in each dataset
n_i <- sapply(pv_datasets, function(df) sum(!is.na(df$READ)))
mean_n <- mean(n_i)
sd_n <- sd(n_i)

cat(
  "Sample size (mean ± SD across replicates):",
  sprintf("%.0f ± %.0f\n", mean_n, sd_n)
)
```

```
Sample size (mean ± SD across replicates): 4565 ± 0
```

Model 1: Computer at home only

```
# Model 1: Computer at home only (replicating your fit1)
model1_formula <- "READ ~ COM_HOM"
model1_results <- run_weighted_model(model1_formula, pv_datasets)

# Display results
summary(model1_results)
```

```
Multiple imputation results:
      MIcombine.default(results = coefs, variances = vcovs)
                      results       se    (lower    upper) missInfo
(Intercept)          448.3566 4.743414 439.03381 457.67945    15 %
COM_HOMHas Computer   69.6754 5.023828  59.81114  79.53966    12 %
```

```
Model 1 Pseudo R² (mean ± SD across replicates): 0.0441 ± 0.0022
```

```
Model 1 Sample size (mean ± SD across replicates): 4565 ± 0
```

Model 2: Internet at home only

```
# Model 2: Computer, Internet, and their interaction (replicating your fit2)
model2_formula <- "READ ~ COM_HOM + INTERNET + COM_HOM:INTERNET"
model2_results <- run_weighted_model(model2_formula, pv_datasets)

summary(model2_results)
```

```
Multiple imputation results:
      MIcombine.default(results = coefs, variances = vcovs)
                                                       results
(Intercept)                                           437.89376
COM_HOMHas Computer                                    29.40122
INTERNETHas Internet (Not Used)                       -26.19420
INTERNETHas Internet (Used)                            15.49669
COM_HOMHas Computer:INTERNETHas Internet (Not Used)   -41.60946
```

```
COM_HOMHas Computer:INTERNETHas Internet (Used)         39.26007
                                                            se
(Intercept)                                            11.30968
COM_HOMHas Computer                                    17.56243
INTERNETHas Internet (Not Used)                        18.97410
INTERNETHas Internet (Used)                            12.41597
COM_HOMHas Computer:INTERNETHas Internet (Not Used) 25.54750
COM_HOMHas Computer:INTERNETHas Internet (Used)         18.36446
                                                         (lower
(Intercept)                                           415.689811
COM_HOMHas Computer                                    -5.037991
INTERNETHas Internet (Not Used)                       -63.396235
INTERNETHas Internet (Used)                            -8.870895
COM_HOMHas Computer:INTERNETHas Internet (Not Used) -91.706707
COM_HOMHas Computer:INTERNETHas Internet (Used)         3.250161
                                                         upper)
(Intercept)                                           460.097706
COM_HOMHas Computer                                    63.840424
INTERNETHas Internet (Not Used)                        11.007833
INTERNETHas Internet (Used)                            39.864280
COM_HOMHas Computer:INTERNETHas Internet (Not Used)  8.487795
COM_HOMHas Computer:INTERNETHas Internet (Used)         75.269976
                                                       missInfo
(Intercept)                                               11 %
COM_HOMHas Computer                                        6 %
INTERNETHas Internet (Not Used)                            5 %
INTERNETHas Internet (Used)                               10 %
COM_HOMHas Computer:INTERNETHas Internet (Not Used)      6 %
COM_HOMHas Computer:INTERNETHas Internet (Used)            6 %
Model 2 Pseudo R² (mean ± SD across replicates): 0.0769 ± 0.0028


Model 2 Sample size (mean ± SD across replicates): 4565 ± 0
```

Model 3: Full model with ICTHOME and all interactions (replicating fit3)

```
# Model 3: Full model with ICTHOME and all interactions (replicating your fit3)
model3_formula <- "READ ~ COM_HOM + INTERNET + ICTHOME +
                 COM_HOM:INTERNET + COM_HOM:ICTHOME +
                 INTERNET:ICTHOME + COM_HOM:INTERNET:ICTHOME"
model3_results <- run_weighted_model(model3_formula, pv_datasets)


summary(model3_results)
```

```
Multiple imputation results:
      MIcombine.default(results = coefs, variances = vcovs)
                                                      results
(Intercept)                                         378.19979
COM_HOMHas Computer                                  94.26215
INTERNETHas Internet (Not Used)                      66.86832
INTERNETHas Internet (Used)                         118.12102
ICTHOME                                              11.99039
```

```
COM_HOMHas Computer:INTERNETHas Internet (Not Used)                      -134.04214
COM_HOMHas Computer:INTERNETHas Internet (Used)                           -53.06729
COM_HOMHas Computer:ICTHOME                                               -12.77843
INTERNETHas Internet (Not Used):ICTHOME                                   -15.87025
INTERNETHas Internet (Used):ICTHOME                                       -17.51114
COM_HOMHas Computer:INTERNETHas Internet (Not Used):ICTHOME    16.09720
COM_HOMHas Computer:INTERNETHas Internet (Used):ICTHOME         16.72572
                                                                          se
(Intercept)                                                         25.240677
COM_HOMHas Computer                                                 48.019618
INTERNETHas Internet (Not Used)                                     45.953584
INTERNETHas Internet (Used)                                         31.687676
ICTHOME                                                              4.784431
COM_HOMHas Computer:INTERNETHas Internet (Not Used)                 80.585229
COM_HOMHas Computer:INTERNETHas Internet (Used)                     52.707009
COM_HOMHas Computer:ICTHOME                                          7.507736
INTERNETHas Internet (Not Used):ICTHOME                             7.053102
INTERNETHas Internet (Used):ICTHOME                                 5.317299
COM_HOMHas Computer:INTERNETHas Internet (Not Used):ICTHOME 10.439960
COM_HOMHas Computer:INTERNETHas Internet (Used):ICTHOME         7.946854
                                                                         (lower
(Intercept)                                                        328.70833828
COM_HOMHas Computer                                                  0.08623722
INTERNETHas Internet (Not Used)                                    -23.46558796
INTERNETHas Internet (Used)                                         56.00359633
ICTHOME                                                              2.60260312
COM_HOMHas Computer:INTERNETHas Internet (Not Used)               -292.61315923
COM_HOMHas Computer:INTERNETHas Internet (Used)                   -156.41866609
COM_HOMHas Computer:ICTHOME                                        -27.50335118
INTERNETHas Internet (Not Used):ICTHOME                            -29.72622485
INTERNETHas Internet (Used):ICTHOME                                -27.93994858
COM_HOMHas Computer:INTERNETHas Internet (Not Used):ICTHOME   -4.43099184
COM_HOMHas Computer:INTERNETHas Internet (Used):ICTHOME         1.14042416
                                                                          upper)
(Intercept)                                                        427.691248
COM_HOMHas Computer                                                188.438069
INTERNETHas Internet (Not Used)                                    157.202224
INTERNETHas Internet (Used)                                        180.238446
ICTHOME                                                             21.378168
COM_HOMHas Computer:INTERNETHas Internet (Not Used)                 24.528884
COM_HOMHas Computer:INTERNETHas Internet (Used)                     50.284087
COM_HOMHas Computer:ICTHOME                                          1.946483
INTERNETHas Internet (Not Used):ICTHOME                             -2.014268
INTERNETHas Internet (Used):ICTHOME                                 -7.082332
COM_HOMHas Computer:INTERNETHas Internet (Not Used):ICTHOME 36.625385
COM_HOMHas Computer:INTERNETHas Internet (Used):ICTHOME        32.311012
                                                                        missInfo
(Intercept)                                                            6 %
COM_HOMHas Computer                                                    7 %
INTERNETHas Internet (Not Used)                                       15 %
INTERNETHas Internet (Used)                                            4 %
ICTHOME                                                                9 %
COM_HOMHas Computer:INTERNETHas Internet (Not Used)                   18 %
COM_HOMHas Computer:INTERNETHas Internet (Used)                        6 %
```

```
COM_HOMHas Computer:ICTHOME                                      7 %
INTERNETHas Internet (Not Used):ICTHOME                         13 %
INTERNETHas Internet (Used):ICTHOME                              7 %
COM_HOMHas Computer:INTERNETHas Internet (Not Used):ICTHOME     16 %
COM_HOMHas Computer:INTERNETHas Internet (Used):ICTHOME          7 %
Model 3 Pseudo R² (mean ± SD across replicates): 0.0802 ± 0.0027


Model 3 Sample size (mean ± SD across replicates): 4565 ± 0
```

Model 4: Non-linear relationship with ICTHOME, adding a squared term

```
# This addresses Reviewer A's comment about proper specification
model4_formula <- "READ ~ COM_HOM + INTERNET + ICTHOME + I(ICTHOME^2) +
                   COM_HOM:INTERNET"
model4_results <- run_weighted_model(model4_formula, pv_datasets)

summary(model4_results)
```

```
Multiple imputation results:
      MIcombine.default(results = coefs, variances = vcovs)
                                                              results
(Intercept)                                                290.015079
COM_HOMHas Computer                                          6.579310
INTERNETHas Internet (Not Used)                            -52.856037
INTERNETHas Internet (Used)                                -17.262059
ICTHOME                                                     46.360254
I(ICTHOME^2)                                                -2.737135
COM_HOMHas Computer:INTERNETHas Internet (Not Used)        -17.191237
COM_HOMHas Computer:INTERNETHas Internet (Used)             59.667791
                                                                   se
(Intercept)                                                19.8797655
COM_HOMHas Computer                                        17.5416997
INTERNETHas Internet (Not Used)                            19.8321597
INTERNETHas Internet (Used)                                13.0193877
ICTHOME                                                     5.2628044
I(ICTHOME^2)                                                0.2979637
COM_HOMHas Computer:INTERNETHas Internet (Not Used)        25.7768896
COM_HOMHas Computer:INTERNETHas Internet (Used)            18.1829901
                                                               (lower
(Intercept)                                                 251.02681
COM_HOMHas Computer                                         -27.82185
INTERNETHas Internet (Not Used)                             -91.74324
INTERNETHas Internet (Used)                                 -42.81271
ICTHOME                                                      36.04110
I(ICTHOME^2)                                                 -3.32143
COM_HOMHas Computer:INTERNETHas Internet (Not Used)         -67.73517
COM_HOMHas Computer:INTERNETHas Internet (Used)             24.01333
                                                               upper)
(Intercept)                                                329.003346
COM_HOMHas Computer                                         40.980473
INTERNETHas Internet (Not Used)                            -13.968836
INTERNETHas Internet (Used)                                  8.288594
```

```
ICTHOME                                                     56.679410
I(ICTHOME^2)                                               -2.152840
COM_HOMHas Computer:INTERNETHas Internet (Not Used)  33.352698
COM_HOMHas Computer:INTERNETHas Internet (Used)      95.322256
                                                          missInfo
(Intercept)                                                    7 %
COM_HOMHas Computer                                            7 %
INTERNETHas Internet (Not Used)                               6 %
INTERNETHas Internet (Used)                                  10 %
ICTHOME                                                       6 %
I(ICTHOME^2)                                                  6 %
COM_HOMHas Computer:INTERNETHas Internet (Not Used)           6 %
COM_HOMHas Computer:INTERNETHas Internet (Used)              6 %


Model 4 Pseudo R² (mean ± SD across replicates): 0.0973 ± 0.0029


Model 4 Sample size (mean ± SD across replicates): 4565 ± 0
```

Model 5: ICTHOME Curve Moderated by Both Computer Ownership and Internet Status

```
# --- Define Enhanced Model 5b ---
model5b_formula_str <- "READ ~ COM_HOM + INTERNET + ICTHOME + I(ICTHOME^2) + COM_HOM:INTERNET +
        COM_HOM:ICTHOME + COM_HOM:I(ICTHOME^2) + INTERNET:ICTHOME + INTERNET:I(ICTHOME^2)"
model5b_results <- run_weighted_model(model5b_formula_str, pv_datasets)
summary(model5b_results)
```

```
Multiple imputation results:
     MIcombine.default(results = coefs, variances = vcovs)
                                                               results
(Intercept)                                                 313.31996262
COM_HOMHas Computer                                          -49.63664693
INTERNETHas Internet (Not Used)                              23.12219773
INTERNETHas Internet (Used)                                 -70.98306748
ICTHOME                                                       55.78102935
I(ICTHOME^2)                                                  -5.05444686
COM_HOMHas Computer:INTERNETHas Internet (Not Used) -73.70465830
COM_HOMHas Computer:INTERNETHas Internet (Used)       6.85358188
COM_HOMHas Computer:ICTHOME                                   13.40793895
COM_HOMHas Computer:I(ICTHOME^2)                              -0.08095963
INTERNETHas Internet (Not Used):ICTHOME                     -23.26482986
INTERNETHas Internet (Used):ICTHOME                           6.09768475
INTERNETHas Internet (Not Used):I(ICTHOME^2)                 2.55884233
INTERNETHas Internet (Used):I(ICTHOME^2)                     0.94381189
                                                                    se
(Intercept)                                                 33.3922099
COM_HOMHas Computer                                          47.9530916
INTERNETHas Internet (Not Used)                             82.5737961
INTERNETHas Internet (Used)                                 53.3091080
ICTHOME                                                      12.4723182
I(ICTHOME^2)                                                  1.0706723
COM_HOMHas Computer:INTERNETHas Internet (Not Used) 28.2917071
```

```
COM_HOMHas Computer:INTERNETHas Internet (Used)        20.4067801
COM_HOMHas Computer:ICTHOME                            12.5822311
COM_HOMHas Computer:I(ICTHOME^2)                        0.7621303
INTERNETHas Internet (Not Used):ICTHOME               23.5509472
INTERNETHas Internet (Used):ICTHOME                   16.1621289
INTERNETHas Internet (Not Used):I(ICTHOME^2)           1.6244813
INTERNETHas Internet (Used):I(ICTHOME^2)               1.2452044
                                                       (lower
(Intercept)                                          247.8605149
COM_HOMHas Computer                                  -143.6827848
INTERNETHas Internet (Not Used)                      -139.4040713
INTERNETHas Internet (Used)                          -175.5181933
ICTHOME                                                31.3179641
I(ICTHOME^2)                                           -7.1556373
COM_HOMHas Computer:INTERNETHas Internet (Not Used) -129.1772351
COM_HOMHas Computer:INTERNETHas Internet (Used)       -33.1743383
COM_HOMHas Computer:ICTHOME                           -11.2648925
COM_HOMHas Computer:I(ICTHOME^2)                       -1.5754422
INTERNETHas Internet (Not Used):ICTHOME              -69.5604769
INTERNETHas Internet (Used):ICTHOME                  -25.6020887
INTERNETHas Internet (Not Used):I(ICTHOME^2)          -0.6335548
INTERNETHas Internet (Used):I(ICTHOME^2)              -1.4992754
                                                       upper)
(Intercept)                                          378.779410
COM_HOMHas Computer                                   44.409491
INTERNETHas Internet (Not Used)                      185.648467
INTERNETHas Internet (Used)                           33.552058
ICTHOME                                               80.244095
I(ICTHOME^2)                                           -2.953256
COM_HOMHas Computer:INTERNETHas Internet (Not Used) -18.232081
COM_HOMHas Computer:INTERNETHas Internet (Used)       46.881502
COM_HOMHas Computer:ICTHOME                           38.080770
COM_HOMHas Computer:I(ICTHOME^2)                        1.413523
INTERNETHas Internet (Not Used):ICTHOME              23.030817
INTERNETHas Internet (Used):ICTHOME                  37.797458
INTERNETHas Internet (Not Used):I(ICTHOME^2)          5.751240
INTERNETHas Internet (Used):I(ICTHOME^2)              3.386899
                                                      missInfo
(Intercept)                                             4 %
COM_HOMHas Computer                                     7 %
INTERNETHas Internet (Not Used)                        18 %
INTERNETHas Internet (Used)                             6 %
ICTHOME                                                 7 %
I(ICTHOME^2)                                           10 %
COM_HOMHas Computer:INTERNETHas Internet (Not Used)     5 %
COM_HOMHas Computer:INTERNETHas Internet (Used)         8 %
COM_HOMHas Computer:ICTHOME                             6 %
COM_HOMHas Computer:I(ICTHOME^2)                        6 %
INTERNETHas Internet (Not Used):ICTHOME               15 %
INTERNETHas Internet (Used):ICTHOME                    7 %
INTERNETHas Internet (Not Used):I(ICTHOME^2)          14 %
INTERNETHas Internet (Used):I(ICTHOME^2)               9 %


Enhanced Model 5b Pseudo R² (mean ± SD across replicates): 0.1063 ± 0.0031
```

Enhanced Model 5b Sample size (mean ± SD across replicates): 4565 ± 0

MODEL DIAGNOSTICS AND COMPARISONS

```r
# Formula Strings
model1_formula <- "READ ~ COM_HOM"
model2_formula <- "READ ~ COM_HOM + INTERNET + COM_HOM:INTERNET"
model3_formula <- "READ ~ COM_HOM + INTERNET + ICTHOME + COM_HOM:INTERNET + COM_HOM:ICTHOME +
        INTERNET:ICTHOME + COM_HOM:INTERNET:ICTHOME"
model4_formula <- "READ ~ COM_HOM + INTERNET + ICTHOME + I(ICTHOME^2) + COM_HOM:INTERNET"
model5b_formula_str <- "READ ~ COM_HOM + INTERNET + ICTHOME + I(ICTHOME^2) + COM_HOM:INTERNET +
        COM_HOM:ICTHOME + COM_HOM:I(ICTHOME^2) + INTERNET:ICTHOME + INTERNET:I(ICTHOME^2)"

# Function to extract model fit statistics
get_model_stats <- function(data_list, formula_string) {
  aics <- sapply(seq_along(data_list), function(i) {
    design <- svydesign(ids = ~1, weights = ~W_FSTUWT, data = data_list[[i]])
    model <- svyglm(as.formula(formula_string), design = design, family = gaussian())
    AIC(model)
  })
  return(mean(aics))
}

# Create the data frame for model comparison
model_comparison <- data.frame(
  Model = c(
    "Model 1: Computer only",
    "Model 2: Computer + Internet + Interaction",
    "Model 3: Full ICT Interactions",
    "Model 4: Quadratic ICTHOME",
    "Model 5: ICTHOME Curve Moderated by COM_HOM & INTERNET"
  ),
  Formula_String = c(
    model1_formula, # Use the formula string variables directly
    model2_formula,
    model3_formula,
    model4_formula,
    model5b_formula_str
  ),
  AIC = NA # Initialize AIC column
)

# Calculate AIC for each model using the CORRECTED function calls
model_comparison$AIC[1] <- get_model_stats(pv_datasets, model_comparison$Formula_String[1])
model_comparison$AIC[2] <- get_model_stats(pv_datasets, model_comparison$Formula_String[2])
model_comparison$AIC[3] <- get_model_stats(pv_datasets, model_comparison$Formula_String[3])
model_comparison$AIC[4] <- get_model_stats(pv_datasets, model_comparison$Formula_String[4])
model_comparison$AIC[5] <- get_model_stats(pv_datasets, model_comparison$Formula_String[5])

# View the comparison table
print(model_comparison)
```

```
                                                        Model
1                                      Model 1: Computer only
2                        Model 2: Computer + Internet + Interaction
3                                  Model 3: Full ICT Interactions
4                                    Model 4: Quadratic ICTHOME
5 Model 5: ICTHOME Curve Moderated by COM_HOM & INTERNET


1
2
3                                     READ ~ COM_HOM + INTERNET + ICTHOME + COM_HOM:INTERNET + COM_HOI
4                                                                        READ ~ COM_HOM
5 READ ~ COM_HOM + INTERNET + ICTHOME + I(ICTHOME^2) + COM_HOM:INTERNET + COM_HOM:ICTHOME + COM_HOM::
        AIC
1 18455.83
2 18406.63
3 18407.00
4 18374.81
5 18364.65
```
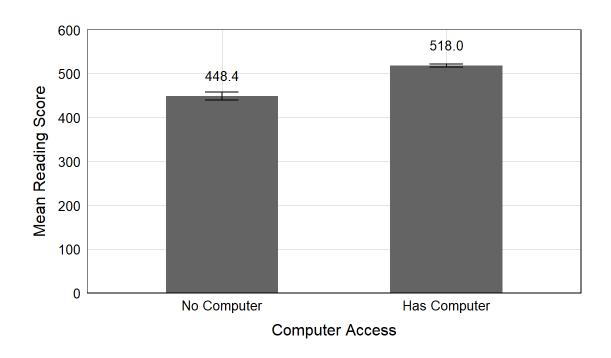
Visualization and Analysis

```r
# Check data structure
str(ICT_data)
```

```
'data.frame':   4838 obs. of  20 variables:
 $ COM_HOM : Factor w/ 2 levels "0","1": 2 1 2 2 2 2 2 2 2 2 ...
 $ INTERNET: Factor w/ 3 levels "0","1","2": 3 3 3 3 3 3 3 3 3 3 ...
 $ ICTHOME : int  10 5 9 6 12 11 9 10 6 9 ...
 $ LOC_INFO: num  537 428 511 432 508 ...
 $ UNDERSTD: num  516 407 501 429 536 ...
 $ EVAL_REF: num  518 422 506 436 562 ...
 $ SINGLE  : num  559 429 508 470 518 ...
 $ MULTIPLE: num  556 422 503 445 537 ...
 $ READ_SCR: num  544 432 504 438 536 ...
 $ W_FSTUWT: num  647 630 614 614 614 ...
 $ PV1READ : num  590 395 501 464 575 ...
 $ PV2READ : num  517 389 491 458 567 ...
 $ PV3READ : num  556 424 516 424 513 ...
 $ PV4READ : num  520 473 498 431 519 ...
 $ PV5READ : num  575 480 505 436 513 ...
 $ PV6READ : num  526 413 532 436 552 ...
 $ PV7READ : num  555 457 498 424 529 ...
 $ PV8READ : num  561 429 499 429 510 ...
 $ PV9READ : num  496 410 505 415 561 ...
 $ PV10READ: num  547 453 494 461 521 ...
```
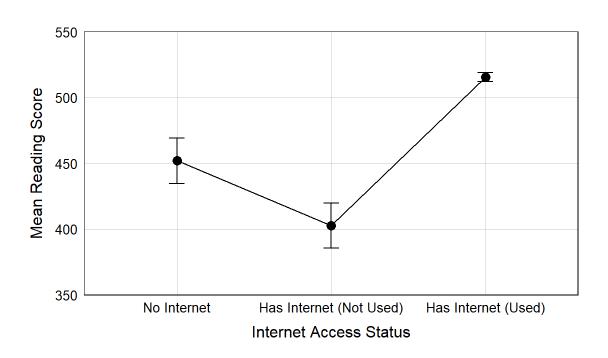
```r
names(ICT_data)
```

```
 [1] "COM_HOM"   "INTERNET" "ICTHOME"  "LOC_INFO" "UNDERSTD" "EVAL_REF"
 [7] "SINGLE"    "MULTIPLE" "READ_SCR" "W_FSTUWT" "PV1READ"  "PV2READ"
[13] "PV3READ"   "PV4READ"  "PV5READ"  "PV6READ"  "PV7READ"  "PV8READ"
[19] "PV9READ"   "PV10READ"
```

```r
# Check missing values
colSums(is.na(ICT_data))
```

```
 COM_HOM INTERNET  ICTHOME LOC_INFO UNDERSTD EVAL_REF   SINGLE
      64      256      192        0        0        0        0
MULTIPLE READ_SCR W_FSTUWT  PV1READ  PV2READ  PV3READ  PV4READ
       0        0        0        0        0        0        0
 PV5READ  PV6READ  PV7READ  PV8READ  PV9READ PV10READ
       0        0        0        0        0        0
```

```r
# Create clean dataset
ICT_clean <- ICT_data %>%
  filter(!is.na(COM_HOM) & !is.na(INTERNET) & !is.na(ICTHOME))

dim(ICT_clean)
```

```
[1] 4565   20
```

```r
# Create plausible value datasets
pv_datasets <- lapply(1:10, function(i) {
  temp_data <- ICT_clean
  temp_data$READ <- temp_data[[paste0("PV", i, "READ")]]
  return(temp_data)
})

# Function to calculate weighted means with plausible values
calculate_weighted_stats <- function(group_var) {
  results_list <- lapply(1:10, function(i) {
    df <- pv_datasets[[i]]
    design <- svydesign(ids = ~1, weights = ~W_FSTUWT, data = df)
    means <- svyby(~READ, as.formula(paste0("~", group_var)),
      design, svymean,
      na.rm = TRUE
    )
    return(means)
  })

  groups <- results_list[[1]][[group_var]]
  combined_results <- data.frame()

  for (g in 1:length(groups)) {
    group_means <- sapply(results_list, function(x) x$READ[g])

    group_ses <- sapply(results_list, function(x) {
      se_vals <- SE(x)
      if (is.matrix(se_vals)) {
```

```r
      se_vals[g, "READ"]
    } else {
      se_vals[g]
    }
  })

  mean_est <- mean(group_means)
  within_var <- mean(group_ses^2)
  between_var <- var(group_means) * (1 + 1 / 10)
  total_var <- within_var + between_var
  total_se <- sqrt(total_var)

  combined_results <- rbind(combined_results, data.frame(
    group = groups[g],
    mean = mean_est,
    se = total_se,
    ci_lower = mean_est - 1.96 * total_se,
    ci_upper = mean_est + 1.96 * total_se
  ))
  }

  return(combined_results)
}

# Calculate statistics
computer_stats <- calculate_weighted_stats("COM_HOM")
internet_stats <- calculate_weighted_stats("INTERNET")
icthome_stats <- calculate_weighted_stats("ICTHOME")

# APA 7 theme function
theme_apa <- function() {
  theme_minimal() +
    theme(
      # Text elements
      text = element_text(size = 12),
      plot.title = element_text(size = 12, face = "bold", hjust = 0, margin = margin(b = 12)),
      axis.title = element_text(size = 12, face = "plain"),
      axis.title.x = element_text(margin = margin(t = 8)),
      axis.title.y = element_text(margin = margin(r = 8)),
      axis.text = element_text(size = 10, color = "black"),

      # Legend
      legend.title = element_text(size = 11, face = "plain"),
      legend.text = element_text(size = 10),
      legend.position = "top",
      legend.justification = "left",
      legend.box.margin = margin(0, 0, 0, 0),

      # Panel elements
      panel.grid.major = element_line(color = "grey80", size = 0.25),
      panel.grid.minor = element_blank(),
      panel.border = element_rect(color = "black", fill = NA, size = 0.5),

      # Spacing
```
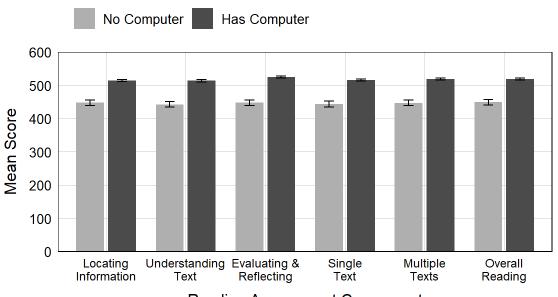
```r
      plot.margin = unit(c(1, 1, 1, 1), "cm")
    )
}

# Check if objects exist, if not create them
if (!exists("computer_stats")) {
  computer_stats <- calculate_weighted_stats("COM_HOM")
}
if (!exists("internet_stats")) {
  internet_stats <- calculate_weighted_stats("INTERNET")
}
if (!exists("icthome_stats")) {
  icthome_stats <- calculate_weighted_stats("ICTHOME")
}

# Ensure subscales data exists
if (!exists("computer_subscales_df")) {
  # Subscale analysis function
  calculate_subscale_stats <- function(outcome_var, group_var, data) {
    design <- svydesign(ids = ~1, weights = ~W_FSTUWT, data = data)

    formula_str <- paste0("~", outcome_var)
    group_formula <- as.formula(paste0("~", group_var))

    means <- svyby(as.formula(formula_str), group_formula,
      design, svymean,
      na.rm = TRUE
    )

    se_vals <- SE(means)

    result <- data.frame(
      group = means[[group_var]],
      mean = means[[outcome_var]],
      se = if (is.matrix(se_vals)) se_vals[, outcome_var] else se_vals
    )
    result$ci_lower <- result$mean - 1.96 * result$se
    result$ci_upper <- result$mean + 1.96 * result$se
    result$outcome <- outcome_var

    return(result)
  }

  # Calculate for all subscales
  subscales <- c("LOC_INFO", "UNDERSTD", "EVAL_REF", "SINGLE", "MULTIPLE", "READ_SCR")

  computer_subscales <- lapply(subscales, function(x) {
    calculate_subscale_stats(x, "COM_HOM", ICT_clean)
  })
  computer_subscales_df <- do.call(rbind, computer_subscales)
}
```

Plotting the Data