Impact Evaluation of SSDP Teacher Training in Nepal

Cosimo ZATTI

2025-12-10

1 Policy Analysis and Policy Assessment: Evaluation of the School Sector Development Program (SSDP) Teacher Training in Nepal

STEP 2:

1. The Policy The intervention under evaluation is the teacher training component of the School Sector Development Program (SSDP), a comprehensive national initiative implemented by the Government of Nepal. Specifically, this policy targets mathematics and science teachers at the secondary level (Grades 9 and 10) in community (public) schools. The policy represents a significant strategic shift from the previous "needs-based" training model employed under the School Sector Reform Program (SSRP). The earlier model, which allowed Education Training Centers (ETCs) to tailor curricula based on local requests, was deemed ineffective due to heterogeneity in quality and a failure to translate training into actual classroom practice. Consequently, the SSDP introduced a centralized, standardized curriculum designed to ensure consistency and rigor across the national education system.

The intervention is defined by specific inputs and activities designed to upgrade pedagogical quality. The primary inputs include the standardized curriculum developed by the National Centre for Educational Development (NCED), the utilization of government ETC infrastructure, the deployment of trainers (both staff and external roster trainers), and financial resources estimated at approximately $130 per teacher. The core activity is a training regimen split into two phases. First, teachers are required to attend a 10-day face-to-face training session delivered at an ETC. This session focuses on clarifying difficult subject content—such as trigonometry in mathematics or the circulatory system in science—and modeling "active learning" pedagogies that utilize locally available materials for demonstrations. Following this residential phase, teachers must complete a 5-day self-study project at their home schools. This project work requires the development of 10 specific lesson plans, the creation of teaching aids, and the completion of a small action research project. The policy was implemented experimentally through the issuance of invitation letters to treatment schools, encouraging all relevant teachers to attend regardless of their prior training history or employment status.

2. Final Outcomes While the study operationalizes its analysis using standardized student test scores, it is crucial from a policy perspective to distinguish between these measured variables and the underlying theoretical constructs they represent. The original goal of the SSDP policy is not merely the elevation of test scores, but the broader enhancement of school quality and inclusivity to drive national development. Recognizing that development success requires the country's youth to acquire valuable skills, the government explicitly prioritized these outcomes to improve human capital accumulation.

The policy is explicitly designed to move instruction away from rote memorization and "chalk-and-talk" lecturing toward methodologies that foster deep conceptual understanding. Therefore, the final outcome of interest is Subject Mastery and Cognitive Competence. This encompasses the students' ability to understand complex abstract concepts in mathematics and science and to apply reasoning skills to solve problems using those concepts. The evaluation attempts to capture this construct by developing assessments that include items drawn from international item banks (such as TIMSS) and items specifically mapped to the advanced concepts emphasized in the SSDP curriculum. By normalizing these scores relative to the control group, the researchers aim to measure the shift in the latent distribution of student learning, effectively using test performance as a proxy for the broader construct of educational quality and human capital formation.

3. Theory of Change The Theory of Change (ToC) for the SSDP intervention relies on a long and complex causal chain involving three distinct sets of actors: trainers, teachers, and students. For the policy to succeed, specific assumptions regarding capacity, resources, and motivation must hold at every link in this chain.

The chain begins with Inputs (curriculum, funding, logistics) facilitating Activities, specifically the delivery of the 10-day training modules by ETC trainers. This step assumes that trainers possess the requisite subject matter expertise and pedagogical skill to deliver the curriculum effectively—an assumption the study later challenges by noting that trainers often lacked specific content knowledge and received no "training of trainers" themselves.

Successful activities are expected to produce Outputs at the teacher level: improved subject knowledge and the acquisition of new pedagogical skills. This assumes that teachers possess sufficient foundational knowledge to grasp the advanced concepts presented in the training. However, the evaluation reveals a critical failure point here, as many teachers demonstrated significant gaps in basic content knowledge (e.g., calculating the perimeter of a rectangle), rendering advanced training potentially inaccessible.

These outputs must then translate into Intermediate Outcomes, defined as a change in classroom behavior. Teachers are expected to return to their schools and implement demonstration-based methods, create lesson plans, and utilize teaching aids. This stage represents the "black box" of the classroom. The ToC assumes teachers have the time, resources, and motivation to change their habits. The study identifies this as a primary failure point: 59% of teachers cited a lack of teaching materials as a barrier. Furthermore, the lack of external accountability meant that few teachers actually completed the required self-study projects.

Finally, these intermediate outcomes lead to Final Outcomes: improved student learning. This final link assumes that students are adequately prepared to benefit from improved instruction at the Grade 9 and 10 levels. The study explicitly identifies a mechanism for unintended negative outcomes here. Because many students enter secondary school with severe learning deficits—such as Grade 9 students unable to solve Grade 3 division problems—training teachers to focus on advanced Grade 10 concepts may exacerbate the mismatch between instruction and student ability. This misalignment could actively harm learning by diverting time away from necessary remedial instruction. This negative effect is likely compounded by the logistical reality of short 40-45 minute class periods, which teachers noted were insufficient for setting up the complex demonstrations required by the new curriculum, potentially slowing down the pace of instruction for high-performing students who might otherwise have progressed faster.

4. Measurement Strategy The evaluation utilizes a mixed-methods measurement strategy that is robust in its intent but faces limitations in its execution, particularly regarding the validation of behavioral change. The strategy relies heavily on triangulation to compensate for weak administrative data systems.

Inputs and Activities were measured through a high-frequency monitoring system involving phone calls to ETCs. This was a necessary and clever adaptation to ensure accurate data on training rollout and attendance, as administrative records were often unreliable or missing.

Outputs (Teacher Knowledge) were measured using a novel indirect method. Rather than subjecting teachers to a formal test—which could induce refusal or labor unrest—the researchers asked teachers to "evaluate" student assessment items by identifying the correct answer and rating clarity. This proxy was highly effective in securing participation (only 2 refusals out of 429 teachers) while successfully revealing deep deficits in teacher content knowledge. This was a highly appropriate measurement choice given the context of potential teacher resistance, as it effectively minimized non-response bias while still capturing the necessary data on subject mastery.

Intermediate Outcomes (Teaching Practices) represent the weakest link in the measurement strategy. The study relies entirely on self-reports from teachers, head teachers, and students to assess changes in classroom behavior. While the researchers attempt to triangulate these sources, the lack of direct classroom observations at endline is a significant critical flaw. Self-reported data on teaching practices is notoriously susceptible to social desirability bias; teachers know they should be using demonstrations, even if they are not. The absence of objective verification (such as the Stallings method used at baseline but not endline) means the study cannot definitively distinguish between a failure to implement the curriculum (teachers didn't change behavior) and a failure of the curriculum itself (teachers changed behavior, but it didn't help).

Final Outcomes were measured using standardized math and science assessments developed specifically for the study using Item Response Theory (IRT). A major strength of this strategy was the inclusion of items from lower grade levels to diagnose foundational deficits. This allowed the researchers to empirically validate the "curriculum mismatch" hypothesis, providing nuance to the null result that a simple grade-level test would have missed.

5. Estimands & Research Design The study employs a Cluster Randomized Controlled Trial (RCT) design, which is the gold standard for causal inference in this context. The randomization unit was the school, which is appropriate given that the intervention is delivered at the school level (invitations sent to the school) and to avoid contamination within schools. The sample consisted of 203 schools across 16 districts, stratified by district and "priority" status (schools with previously untrained teachers), ensuring the sample was nearly nationally representative.

The primary estimand is the Intent-to-Treat (ITT) effect. This estimates the causal effect of being assigned to the treatment group (receiving the invitation) on student outcomes. The ITT is the most policy-relevant estimand because the government can mandate training availability but cannot strictly force attendance or compliance. The estimator used is a Weighted Least Squares (WLS) regression of endline scores on the treatment indicator, controlling for strata fixed effects. The secondary estimand is the Local Average Treatment Effect (LATE), which estimates the effect of actual training attendance on the compliers—those teachers who attended because they were invited. This is estimated using an Instrumental Variables (IV) approach, where random assignment is used as the instrument for training participation.

The validity of these estimands rests on several critical identifying assumptions:

- **Independence:** Random assignment ensures that the treatment and control groups are comparable in expectation, eliminating selection bias. The authors validate this through balance tests, which show no systematic differences in baseline characteristics.

- **SUTVA (Stable Unit Treatment Value Assumption):** This assumption requires that the treatment status of one unit does not affect the outcomes of another. The researchers explicitly addressed this by sampling only one school per Village Development Committee (VDC), thereby creating a geographical buffer to minimize spillovers and contamination between treatment and control schools.

· **Exclusion Restriction (for IV in LATE):** This assumption requires that the instrument (invitation) affects the outcome only through the treatment (attendance). This is a strong assumption. It could be violated if the mere receipt of an invitation motivated a teacher to self-study or change behavior without actually attending the training. However, given the qualitative evidence of low teacher motivation and lack of accountability, the authors implicitly argue that this channel is negligible, supporting the validity of the LATE estimates.

· **Attrition:** A potential threat to internal validity is the high student attrition rate observed (36-43%). If attrition is correlated with treatment and potential outcomes, estimates could be biased. The authors explicitly check this and find that while attrition rates were similar, attritors in the treatment group had lower test scores than those in the control group. They argue this differential attrition would likely bias the results downward (making them more negative), suggesting that their finding of "no positive impact" is robust and perhaps even conservative.

6. Findings (Substantive vs. Statistical) The study reports null to negative impacts of the new SSDP training on academic achievement, as indicated by both the ITT and LATE regressions. While the authors conclude that there is "no evidence" that the training raised test scores, it is crucial to distinguish between the statistical and substantive significance of this finding.

Statistically, the results are null: for all estimates, the null hypothesis of zero effect cannot be rejected at the 0.05 level, though some negative coefficients are significant at the 0.10 level. However, substantively, this is a meaningful finding of ineffectiveness rather than just an imprecise "absence of evidence." The standard errors (approx. 0.07 SD) were precise enough to construct confidence intervals that ruled out positive effects larger than 0.10 standard deviations. In the context of education policy, where 0.10-0.20 SD is often considered a modest but relevant success, the ability to statistically rule out an effect of 0.10 SD is substantively significant. It allows policymakers to conclude with high confidence that the program—as currently implemented—is not generating the desired learning gains, rather than simply concluding that the study was too noisy to detect them.

The findings also reveal heterogeneity that reinforces the "curriculum mismatch" theory. While not always statistically significant due to power limitations, the negative estimates are concentrated among students with higher baseline performance. This suggests that the new pedagogy may have actively harmed learning for these students, perhaps by slowing down the pace of instruction without clarifying the advanced concepts they were ready to learn. The LATE estimates mirror the ITT results, confirming that the failure was not solely due to low attendance but due to the inefficacy of the training itself.

7. Improvements While the internal validity of the RCT is high, the study's ability to diagnose the specific mechanisms of failure is limited by its design. To provide more actionable policy intelligence, a future iteration of this study should incorporate the following improvements:

Design: 2x2 Factorial Design. The qualitative analysis identified "lack of materials" as a primary barrier to implementation, with 59% of teachers citing it as an obstacle. The current design confounds the failure of knowledge transfer (training) with the failure of resource provision (materials). To disentangle these, a 2x2 Factorial Design (Cross-Cutting Design) is necessary. Schools should be randomized into four arms:

1. Control (Status Quo)

2. Training Only

3. Materials/Grants Only

4. Training + Materials This design would allow for the estimation of the main effect of materials alone and, crucially, the interaction effect between training and materials. It tests the hypothesis that

training is necessary but insufficient, and that it only activates when paired with the physical resources required to execute the new pedagogy.

Data: High-Frequency Classroom Observations. The reliance on self-reports for teaching practices is a critical weakness that obscures the "black box" of the classroom. Future data collection must include high-frequency, unannounced classroom observations. Using a standardized protocol like the Stallings method (which was used at baseline but seemingly not at endline for impact estimation), researchers could objectively measure "Time on Task" and the frequency of demonstration-based methods. This data is essential to distinguish definitively between "implementation failure" (teachers didn't try the new methods) and "theory failure" (teachers used the methods, but they didn't work).

Testing: Power Calculations for Non-Compliance. The study suffered from lower-than-expected precision because the power calculations assumed perfect compliance, whereas actual attendance was only 60% for math and 42% for science. Future designs must structurally account for non-compliance in their ex-ante power calculations. By adjusting the Minimum Detectable Effect (MDE) to account for the participation rate (effectively scaling the MDE by the inverse of the compliance rate), researchers can ensure the sample size is sufficient to detect the LATE, preventing the ambiguity that arises when standard errors are "small enough" for ITT but too large for LATE. Additionally, a dedicated Quantile Regression analysis with sufficient power should be pre-specified to rigorously test the hypothesis that the intervention harms the top decile of students, moving beyond the suggestive evidence currently presented.

This document presents a comprehensive replication of the impact evaluation for the SSDP Teacher Training program in Nepal. It executes the full data processing and analysis pipeline to reproduce Tables 1 through 17 from the original study.

The analysis follows the methodology described in the final report, utilizing a randomized control trial (RCT) of 203 schools across 16 districts.

Note on Figures: This replication focuses exclusively on the data-driven tables (Tables 1–17). The original study includes four figures: Figure 1 (Theory of Change), Figure 2 (Map of Study Districts), Figure 3 (Random Assignment Flowchart), and Figure 4 (Study Timeline). These are conceptual diagrams or maps that rely on external GIS shapefiles or qualitative information not provided in the replication dataset. Consequently, they are not reproduced programmatically in this report.

2 Comprehensive R Code & Results

2.1 The code below loads the datasets, cleans the variables, handles missing data, and performs the regression analyses to generate Tables 1 through 17.

```r
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, haven, survey, quantreg, mirt,
               AER, sandwich, lmtest, modelsummary, knitr, openxlsx)

base_dir <- "/Users/lucia/Desktop/REPLICATION PAPE/NepalTeacherTrainingPBR"
path_datasets <- file.path(base_dir, "Datasets")
path_tables   <- file.path(base_dir, "Excel_Tables")
```

```
dir.create(path_tables, showWarnings = FALSE)

knitr::opts_knit$set(root.dir = path_datasets)
```

```
# Robust Merge
stata_merge <- function(x, y, by, suffix = c(".x", ".y")) {
  merged <- full_join(x, y, by = by, suffix = suffix)
  x_flag <- x %>% select(all_of(by)) %>% mutate(in_master = 1)
  y_flag <- y %>% select(all_of(by)) %>% mutate(in_using = 1)
  flags <- full_join(x_flag, y_flag, by = by) %>%
    mutate(`_merge` = case_when(
      in_master == 1 & is.na(in_using) ~ 1,
      is.na(in_master) & in_using == 1 ~ 2,
      in_master == 1 & in_using == 1 ~ 3
    )) %>% select(all_of(by), `_merge`)
  left_join(merged, flags, by = by)
}

# Normalization
normalize_score <- function(df, target_col, design) {
  tryCatch({
    stats <- svymean(as.formula(paste0("~", target_col)),
                     design = subset(design, treat == 0), na.rm = TRUE)
    var_obj <- svyvar(as.formula(paste0("~", target_col)),
                      design = subset(design, treat == 0), na.rm = TRUE)
    return((df[[target_col]] - coef(stats)) / sqrt(coef(var_obj)))
  }, error = function(e) { return(df[[target_col]]) })
}

# Column Finder
find_ssrp_col <- function(df, subject) {
  cols <- names(df)
  matches <- cols[grepl("ssrp", cols) & grepl(subject, cols)]
  if (length(matches) == 0) return(NULL)
  return(matches[1])
}
```

```
if (!file.exists("basicdata.dta")) {
  message("Current Working Directory: ", getwd())
  message("Files in this directory: ", paste(list.files(), collapse = ", "))
  stop("CRITICAL ERROR: basicdata.dta not found. Check the paths printed
   ↪ above.")
}

# 1. Load Basics
df_basic <- read_dta("basicdata.dta") %>% rename_with(tolower)

# 2. Load Assessments
e_math_9  <- read_dta("Math09_endline_IRT.dta") %>% rename_with(tolower)
```

```r
e_sci_9   <- read_dta("Sci09_endline_IRT.dta")  %>% rename_with(tolower)
e_math_10 <- read_dta("Math10_endline_IRT.dta") %>% rename_with(tolower)
e_sci_10  <- read_dta("Sci10_endline_IRT.dta")  %>% rename_with(tolower)

# 3. School Covariates
df_school_base <- read_dta("endlinetestassent.dta") %>%
  rename_with(tolower) %>%
  left_join(df_basic, by = "schoolid") %>%
  mutate(
    dist_stratum = (district * 10) + stratum,
    TT = if_else(studyarm == 1, 1, 0),
    TTVA = if_else(studyarm == 2, 1, 0),
    treat = TT + TTVA
  )

# 4. Add HT and Management Data
df_ht_base <- read_dta("HTschoolbaseline.dta") %>% rename_with(tolower)
df_mgmt <- read_dta("management.dta") %>%
  rename_with(tolower) %>%
  rename(theta_mgmt = theta1)

df_school_base <- df_school_base %>%
  select(-any_of(names(df_ht_base)[names(df_ht_base) != "schoolid"])) %>%
  left_join(df_ht_base, by = "schoolid") %>%
  left_join(df_mgmt, by = "schoolid") %>%
  mutate(
    logstud = log(total_all),
    HTmasters = if_else(httopdegree >= 3, 1, 0),
    femtchrpct = t_female_all / teachertotal,
    goodmgmt = if_else(!is.na(theta_mgmt) & theta_mgmt > 0.06503, 1, 0)
  )

# 5. Teacher Data
df_teacher <- read_dta("Teacher_c.dta") %>%
  rename_with(tolower) %>%
  mutate(texp05yr = if_else(!is.na(t_exper) & t_exper <= 5, 1, 0))

df_ssrp <- read_dta("SSRPdata.dta") %>% rename_with(tolower)
ssrp_math_var <- find_ssrp_col(df_ssrp, "math")
ssrp_sci_var  <- find_ssrp_col(df_ssrp, "sci")

if (!is.null(ssrp_math_var)) {
  df_ssrp <- df_ssrp %>% rename(ssrpmath_std = all_of(ssrp_math_var))
} else { df_ssrp$ssrpmath_std <- NA }

if (!is.null(ssrp_sci_var)) {
  df_ssrp <- df_ssrp %>% rename(ssrpsci_std = all_of(ssrp_sci_var))
} else { df_ssrp$ssrpsci_std <- NA }

df_teacher <- stata_merge(df_teacher, df_ssrp, by = "teacherid") %>%
```

```
  select(-`_merge`)

df_ssdp <- read_dta("SSDP_VA_attendance.dta") %>%
  rename_with(tolower) %>%
  filter(teacherid != ".")

df_ssdp <- df_ssdp %>% mutate(
  ssdp_m_t = if_else((ssdp_math_days >= 6) |
                     (ssdp_math == 1 & is.na(ssdp_math_days)), 1, 0),
  ssdp_s_t = if_else((ssdp_sci_days >= 9) |
                     (ssdp_sci == 1 & is.na(ssdp_sci_days)), 1, 0)
)
df_teacher_full <- stata_merge(df_ssdp, df_teacher, by = "teacherid")

# 6. Match Teachers to Students
process_teacher_match <- function(grade_file, subject_col, teacher_data) {
  df_map <- read_dta(grade_file) %>% rename_with(tolower)
  t_col <- if(subject_col == "math") "mathteacherid" else "sciteacherid"
  t_colA <- paste0(t_col, "a"); t_colB <- paste0(t_col, "b")
  match_col <- if(subject_col == "math") "m_match_type" else "s_match_type"

  df_map <- df_map %>% filter(!!sym(match_col) != 2)
  df_single <- df_map %>% filter(!!sym(match_col) != 3) %>%
    rename(teacherid = !!sym(t_col))
  df_double <- df_map %>% filter(!!sym(match_col) == 3)

  df_d1 <- df_double %>% mutate(teacherid = !!sym(t_colA))
  df_d2 <- df_double %>% mutate(teacherid = !!sym(t_colB))
  df_combined <- bind_rows(df_single, df_d1, df_d2)

  df_merged <- left_join(df_combined, teacher_data, by = "teacherid")

  prefix <- if(subject_col == "math") "mat_t_" else "sci_t_"
  ssdp_col <- if(subject_col == "math") "ssdp_m_t" else "ssdp_s_t"
 target_ssrp <- if(subject_col == "math") "ssrpmath_std" else "ssrpsci_std"

  df_merged <- df_merged %>% rename(
    !!paste0(prefix, "perm") := t_perm,
    !!paste0(prefix, "exp05yr") := texp05yr,
    !!paste0(prefix, "ssrp_train") := !!sym(target_ssrp)
  )

  df_merged %>% group_by(stu_serial) %>%
    summarize(
      across(starts_with(prefix), \(x) mean(x, na.rm=TRUE)),
      across(all_of(ssdp_col), \(x) mean(x, na.rm=TRUE))
    )
}

df_g09_mat_t <- process_teacher_match("T_stu_sections09.dta", "math",
↪  df_teacher_full)
```

```r
df_g10_mat_t <- process_teacher_match("T_stu_sections10.dta", "math",
 ↪  df_teacher_full)
df_g09_sci_t <- process_teacher_match("T_stu_sections09.dta", "sci",
 ↪  df_teacher_full)
df_g10_sci_t <- process_teacher_match("T_stu_sections10.dta", "sci",
 ↪  df_teacher_full)
```

```r
df_turnover <- read_dta("Teacherturnover_c.dta") %>% rename_with(tolower)
 ↪  %>%
  mutate(t_teachnow = if_else(teacherid == "0346", 1, t_teachnow)) %>%
  filter(t_teachnow == 1)

df_t_q <- read_dta("Teacher_c.dta") %>% rename_with(tolower)
df_merged_1 <- stata_merge(df_turnover, df_t_q, by = "teacherid")

if ("schoolid.x" %in% names(df_merged_1)) {
  df_merged_1 <- df_merged_1 %>%
    mutate(schoolid = coalesce(schoolid.x, schoolid.y))
}
df_merged_1 <- df_merged_1 %>%
  mutate(t_quest_admin = if_else(`_merge` >= 2, 1, 0))

print("Table 1: Teacher Questionnaire Administration (0=No, 1=Yes)")
```

```
[1] "Table 1: Teacher Questionnaire Administration (0=No, 1=Yes)"
```

```r
print(table(df_merged_1$t_quest_admin))
```

```

  0   1 
 46 434 
```

```r
df_rollout <- read_dta("SSDP_VA_attendance.dta") %>%
  rename_with(tolower) %>%
  filter(!is.na(district))

table_2 <- df_rollout %>%
  mutate(
    math_complete = if_else((ssdp_math_days >= 6) |
                            (ssdp_math == 1 & is.na(ssdp_math_days)), 1, 0),
    sci_complete = if_else((ssdp_sci_days >= 9) |
                           (ssdp_sci == 1 & is.na(ssdp_sci_days)), 1, 0)
  ) %>%
  group_by(district) %>%
  summarize(
    Math_Completed = sum(math_complete, na.rm=TRUE),
    Sci_Completed = sum(sci_complete, na.rm=TRUE)
```

```
  )

print("Table 2: Completed Trainings by District")
```

```
[1] "Table 2: Completed Trainings by District"
```

```
print(head(table_2, 16))
```

```
# A tibble: 16 x 3
   district              Math_Completed Sci_Completed
   <dbl+lbl>                      <dbl>         <dbl>
 1  2 [Panchthar]                     6             5
 2  5 [Morang]                       18             7
 3 11 [Solukhumbu]                    5             2
 4 20 [Sindhuli]                      6             4
 5 24 [Kavrepalanchowk]               5             3
 6 28 [Nuwakot]                       5             4
 7 34 [Parsa]                         8             6
 8 35 [Chitwan]                       4             4
 9 37 [Lamjung]                       8             5
10 45 [Baglung]                       6             4
11 50 [Kapilbastu]                    7             4
12 51 [Arghakhanchi]                  9             5
13 55 [Salyan]                        5             4
14 60 [Dailekh]                       6             5
15 63 [Jumla]                         0             0
16 69 [Achham]                        7             5
```

```
# --- Prepare Baseline Data (Table 3) ---
ht_data <- read_dta("HTschoolbaseline.dta") %>%
  zap_labels() %>%
  rename_with(tolower) %>%
  left_join(df_basic, by = "schoolid")

if ("district.x" %in% names(ht_data)) {
  ht_data <- ht_data %>% mutate(district = coalesce(district.x,
↪   district.y))
}

ht_data <- ht_data %>%
  mutate(
    dist_stratum = (district * 10) + stratum,
    treat = if_else(studyarm == 1 | studyarm == 2, 1, 0),
    ht_masters = if_else(httopdegree >= 3, 1, 0)
  )

ht_svy <- svydesign(id = ~schoolid, strata = ~dist_stratum,
```

```
                        weights = ~sch_wght, data = ht_data, nest = TRUE)

print("Table 3: School Characteristics Balance (Means by Treatment)")
```

```
[1] "Table 3: School Characteristics Balance (Means by Treatment)"
```

```
print(svyby(~total_all + hourstoroad + ht_masters, by = ~treat,
             design = ht_svy, FUN = svymean, na.rm=TRUE))
```

```
  treat total_all hourstoroad ht_masters se.total_all se.hourstoroad
0     0  453.0851     3.47349  0.5196170     26.25454      0.5801346
1     1  402.0621     2.84936  0.6293477     20.68490      0.4234846
  se.ht_masters
0    0.05788089
1    0.05120348
```

```
# --- Student Baseline (Table 4) ---
st_data <- bind_rows(
  read_dta("Grade08baseline.dta") %>% zap_labels(),
  read_dta("Grade09baseline.dta") %>% zap_labels()
) %>%
  rename_with(tolower) %>%
  left_join(df_basic, by = "schoolid")

if ("district.x" %in% names(st_data)) {
  st_data <- st_data %>% mutate(district = coalesce(district.x,
 ↪  district.y))
}

st_data <- st_data %>%
  mutate(
    dist_stratum = (district * 10) + stratum,
    treat = if_else(studyarm == 1 | studyarm == 2, 1, 0),
    female = if_else(gender == 2, 1, 0)
  )

st_svy <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                    weights = ~sch_wght, data = st_data, nest = TRUE)

print("Table 4: Student Characteristics Balance (Means by Treatment)")
```

```
[1] "Table 4: Student Characteristics Balance (Means by Treatment)"
```

```
print(svyby(~female, by = ~treat, design = st_svy, FUN = svymean,
 ↪  na.rm=TRUE))
```

```
  treat    female          se
0     0 0.5484466 0.007598731
1     1 0.5487305 0.010690973
```

```
run_impact_analysis <- function(grade, subject, irt_file,
                               df_teacher_match, student_file) {
  print(paste("---", subject, "Grade", grade, "---"))
  df_scores <- read_dta(irt_file) %>% rename_with(tolower)

  df_merged <- df_scores %>% left_join(df_school_base, by = "schoolid")
  if ("district.x" %in% names(df_merged)) {
    df_merged <- df_merged %>% mutate(district = coalesce(district.x,
 ↪  district.y))
  }

  df_stu <- read_dta(student_file) %>% rename_with(tolower)
  df_merged <- inner_join(df_merged, df_stu, by = "stu_serial", suffix =
 ↪  c("", ".y")) %>%
    select(-ends_with(".y")) %>%
    inner_join(df_teacher_match, by = "stu_serial") %>%
    filter(!is.na(dist_stratum))

  svy_des <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                      weights = ~sch_wght, data = df_merged, nest = TRUE)

  theta_col <- if(grade==9) "theta_gr09" else "theta_gr10"
  df_merged$std_irt <- normalize_score(df_merged, theta_col, svy_des)
  svy_des <- update(svy_des, std_irt = df_merged$std_irt)

  print("Table 5 (ITT):")
  print(summary(svyglm(std_irt ~ treat + asnt_aft + math_1st, design =
   ↪  svy_des)))

  t_treat_col <- if(subject == "Math") "ssdp_m_t" else "ssdp_s_t"
  df_merged$tchr_treated <- df_merged[[t_treat_col]] * df_merged$treat

  print("Table 6 (LATE/IV):")
  try({
    m_iv <- ivreg(std_irt ~ tchr_treated + asnt_aft + math_1st +
 ↪  factor(dist_stratum) |
                  treat + asnt_aft + math_1st + factor(dist_stratum),
                  weights = sch_wght, data = df_merged)
    print(coeftest(m_iv, vcov = vcovCL, cluster = ~schoolid))
  })
  return(df_merged)
}

df_g9m <- run_impact_analysis(9, "Math", "Math09_endline_IRT.dta",
 ↪  df_g09_mat_t, "Grade09_c.dta")
```

```
[1] "--- Math Grade 9 ---"
[1] "Table 5 (ITT):"

Call:
svyglm(formula = std_irt ~ treat + asnt_aft + math_1st, design = svy_des)

Survey design:
update(svy_des, std_irt = df_merged$std_irt)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.05197    0.12016   0.432    0.666
treat       -0.03391    0.11250  -0.301    0.763
asnt_aft     0.07324    0.11911   0.615    0.540
math_1st    -0.18760    0.10993  -1.706    0.090 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.9598839)

Number of Fisher Scoring iterations: 2

[1] "Table 6 (LATE/IV):"

t test of coefficients:

                         Estimate Std. Error t value  Pr(>|t|)
(Intercept)             -0.452532   0.176545 -2.5633 0.0103933 *
tchr_treated            -0.146657   0.096817 -1.5148 0.1298790
asnt_aft                 0.045817   0.077654  0.5900 0.5552012
math_1st                -0.101662   0.062606 -1.6238 0.1044627
factor(dist_stratum)22   0.357947   0.185923  1.9252 0.0542462 .
factor(dist_stratum)51   0.385465   0.194724  1.9795 0.0477999 *
factor(dist_stratum)52   0.366778   0.192279  1.9075 0.0564991 .
factor(dist_stratum)111  0.724203   0.308025  2.3511 0.0187493 *
factor(dist_stratum)112  0.557690   0.250999  2.2219 0.0263282 *
factor(dist_stratum)201  0.145769   0.223171  0.6532 0.5136699
factor(dist_stratum)202  0.256838   0.168228  1.5267 0.1268826
factor(dist_stratum)241  0.805696   0.220845  3.6482 0.0002663 ***
factor(dist_stratum)242  0.792131   0.236264  3.3527 0.0008051 ***
factor(dist_stratum)281  0.990480   0.232856  4.2536 2.135e-05 ***
factor(dist_stratum)282  1.623133   0.253703  6.3978 1.696e-10 ***
factor(dist_stratum)341  0.592077   0.234957  2.5199 0.0117632 *
factor(dist_stratum)342  0.530320   0.204466  2.5937 0.0095183 **
factor(dist_stratum)351  1.220607   0.210269  5.8050 6.767e-09 ***
factor(dist_stratum)352  1.134369   0.163292  6.9469 4.129e-12 ***
factor(dist_stratum)371  1.172764   0.251713  4.6591 3.244e-06 ***
factor(dist_stratum)372  1.674037   0.268643  6.2315 4.934e-10 ***
factor(dist_stratum)451  0.848760   0.184605  4.5977 4.359e-06 ***
factor(dist_stratum)452  1.220807   0.278982  4.3759 1.230e-05 ***
```

```
factor(dist_stratum)501  0.370546   0.204868  1.8087 0.0705476 .
factor(dist_stratum)502  0.368305   0.278420  1.3228 0.1859394
factor(dist_stratum)511  0.849256   0.216062  3.9306 8.568e-05 ***
factor(dist_stratum)512  0.556767   0.221381  2.5150 0.0119300 *
factor(dist_stratum)551 -0.358009   0.203707 -1.7575 0.0788885 .
factor(dist_stratum)552 -0.027700   0.174073 -0.1591 0.8735707
factor(dist_stratum)601 -0.182374   0.204237 -0.8930 0.3719193
factor(dist_stratum)602 -0.087931   0.190593 -0.4614 0.6445629
factor(dist_stratum)631 -0.195805   0.241114 -0.8121 0.4167760
factor(dist_stratum)632 -0.131159   0.187881 -0.6981 0.4851431
factor(dist_stratum)691  0.448577   0.362163  1.2386 0.2155399
factor(dist_stratum)692  0.401525   0.198431  2.0235 0.0430657 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
df_g9s <- run_impact_analysis(9, "Sci", "Sci09_endline_IRT.dta",
 ↪  df_g09_sci_t, "Grade09_c.dta")
```

```
[1] "--- Sci Grade 9 ---"
[1] "Table 5 (ITT):"

Call:
svyglm(formula = std_irt ~ treat + asnt_aft + math_1st, design = svy_des)

Survey design:
update(svy_des, std_irt = df_merged$std_irt)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02863    0.10040   0.285    0.776
treat       -0.06521    0.09564  -0.682    0.496
asnt_aft     0.06760    0.09433   0.717    0.475
math_1st    -0.13252    0.09548  -1.388    0.167

(Dispersion parameter for gaussian family taken to be 0.9192169)

Number of Fisher Scoring iterations: 2

[1] "Table 6 (LATE/IV):"

t test of coefficients:

                      Estimate Std. Error t value  Pr(>|t|)
(Intercept)          -0.2951468  0.1800927 -1.6389 0.1012955
tchr_treated         -0.2830896  0.1883457 -1.5030 0.1328840
asnt_aft              0.0349372  0.0821038  0.4255 0.6704700
math_1st             -0.0733339  0.0741216 -0.9894 0.3225209
factor(dist_stratum)22  0.7376445  0.2629096  2.8057 0.0050372 **
factor(dist_stratum)51  0.0221251  0.1895318  0.1167 0.9070738
```

```
factor(dist_stratum)52   0.0383254  0.2636697  0.1454 0.8844365
factor(dist_stratum)111  0.6452753  0.2688768  2.3999 0.0164305 *
factor(dist_stratum)112  0.4305272  0.2994204  1.4379 0.1505242
factor(dist_stratum)201  0.1190974  0.2406938  0.4948 0.6207535
factor(dist_stratum)202  0.1856504  0.2122725  0.8746 0.3818349
factor(dist_stratum)241  0.9795988  0.3265249  3.0001 0.0027104 **
factor(dist_stratum)242  0.3728728  0.2530999  1.4732 0.1407439
factor(dist_stratum)281  0.8152870  0.2311537  3.5270 0.0004234 ***
factor(dist_stratum)282  1.2730129  0.1893558  6.7229 1.949e-11 ***
factor(dist_stratum)341  0.1418392  0.2111424  0.6718 0.5017562
factor(dist_stratum)342  0.0646790  0.2522575  0.2564 0.7976503
factor(dist_stratum)351  0.9340258  0.2121239  4.4032 1.085e-05 ***
factor(dist_stratum)352  0.5557119  0.1868482  2.9741 0.0029500 **
factor(dist_stratum)371  0.5903301  0.2337445  2.5255 0.0115780 *
factor(dist_stratum)372  0.8959984  0.2478923  3.6145 0.0003035 ***
factor(dist_stratum)451  0.6708601  0.2411441  2.7820 0.0054198 **
factor(dist_stratum)452  0.8894642  0.2382791  3.7329 0.0001911 ***
factor(dist_stratum)501  0.2108639  0.2062871  1.0222 0.3067344
factor(dist_stratum)502  0.1077576  0.2549532  0.4227 0.6725614
factor(dist_stratum)511  0.6013875  0.2045183  2.9405 0.0032894 **
factor(dist_stratum)512  0.4298331  0.2416595  1.7787 0.0753447 .
factor(dist_stratum)551 -0.4260246  0.2038480 -2.0899 0.0366682 *
factor(dist_stratum)552  0.0072073  0.3476039  0.0207 0.9834584
factor(dist_stratum)601 -0.2048351  0.2219966 -0.9227 0.3562037
factor(dist_stratum)602  0.1741834  0.1880392  0.9263 0.3543203
factor(dist_stratum)631 -0.3368857  0.2428705 -1.3871 0.1654634
factor(dist_stratum)632 -0.1646376  0.2176726 -0.7564 0.4494668
factor(dist_stratum)691  0.3670710  0.3429010  1.0705 0.2844439
factor(dist_stratum)692  0.4291941  0.2052808  2.0908 0.0365916 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
df_g10m <- run_impact_analysis(10, "Math", "Math10_endline_IRT.dta",
  df_g10_mat_t, "Grade10_c.dta")
```

```
[1] "--- Math Grade 10 ---"
[1] "Table 5 (ITT):"

Call:
svyglm(formula = std_irt ~ treat + asnt_aft + math_1st, design = svy_des)

Survey design:
update(svy_des, std_irt = df_merged$std_irt)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04766    0.12262   0.389    0.698
treat       -0.02309    0.12785  -0.181    0.857
asnt_aft     0.01785    0.13371   0.134    0.894
```

```
math_1st    -0.11225    0.12356  -0.908    0.365

(Dispersion parameter for gaussian family taken to be 0.9923289)

Number of Fisher Scoring iterations: 2

[1] "Table 6 (LATE/IV):"

t test of coefficients:

                          Estimate Std. Error t value  Pr(>|t|)    
(Intercept)              -0.399179   0.148686 -2.6847 0.0072829 ** 
tchr_treated             -0.100004   0.109756 -0.9111 0.3622604    
asnt_aft                 -0.114306   0.080430 -1.4212 0.1553222    
math_1st                  0.011427   0.066267  0.1724 0.8631036    
factor(dist_stratum)22    1.214399   0.759000  1.6000 0.1096612    
factor(dist_stratum)51    0.278403   0.204993  1.3581 0.1744896    
factor(dist_stratum)52    0.038326   0.181232  0.2115 0.8325251    
factor(dist_stratum)111   0.977332   0.267064  3.6595 0.0002552 ***
factor(dist_stratum)112   0.426710   0.165202  2.5830 0.0098233 ** 
factor(dist_stratum)201   0.128715   0.249074  0.5168 0.6053375    
factor(dist_stratum)202   0.545615   0.194692  2.8025 0.0050907 ** 
factor(dist_stratum)241   0.850537   0.209157  4.0665 4.845e-05 ***
factor(dist_stratum)242   0.521663   0.249429  2.0914 0.0365390 *  
factor(dist_stratum)281   0.671538   0.173679  3.8665 0.0001118 ***
factor(dist_stratum)282   1.782132   0.165003 10.8006 < 2.2e-16 ***
factor(dist_stratum)341   0.622603   0.153791  4.0484 5.234e-05 ***
factor(dist_stratum)342   0.802449   0.185676  4.3218 1.577e-05 ***
factor(dist_stratum)351   1.430927   0.188969  7.5723 4.332e-14 ***
factor(dist_stratum)352   0.738605   0.172335  4.2859 1.854e-05 ***
factor(dist_stratum)371   1.314537   0.246111  5.3412 9.634e-08 ***
factor(dist_stratum)372   0.969576   0.149330  6.4928 9.228e-11 ***
factor(dist_stratum)451   0.658446   0.162775  4.0451 5.307e-05 ***
factor(dist_stratum)452   1.006815   0.202985  4.9600 7.278e-07 ***
factor(dist_stratum)501   0.667856   0.249753  2.6741 0.0075177 ** 
factor(dist_stratum)502   0.190460   0.162671  1.1708 0.2417212    
factor(dist_stratum)511   0.759024   0.169459  4.4791 7.660e-06 ***
factor(dist_stratum)512   0.766425   0.330171  2.3213 0.0203104 *  
factor(dist_stratum)551  -0.347772   0.177494 -1.9593 0.0501277 .  
factor(dist_stratum)552   0.548811   0.158573  3.4609 0.0005427 ***
factor(dist_stratum)601  -0.352360   0.158932 -2.2170 0.0266639 *  
factor(dist_stratum)602   0.256882   0.161591  1.5897 0.1119651    
factor(dist_stratum)691   0.110731   0.211360  0.5239 0.6003711    
factor(dist_stratum)692   0.746529   0.285583  2.6141 0.0089739 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
df_g10s <- run_impact_analysis(10, "Sci", "Sci10_endline_IRT.dta",
 ↪  df_g10_sci_t, "Grade10_c.dta")
```

```
[1] "--- Sci Grade 10 ---"
[1] "Table 5 (ITT):"

Call:
svyglm(formula = std_irt ~ treat + asnt_aft + math_1st, design = svy_des)

Survey design:
update(svy_des, std_irt = df_merged$std_irt)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01303    0.12838   0.102    0.919
treat       -0.01251    0.10710  -0.117    0.907
asnt_aft     0.06290    0.13136   0.479    0.633
math_1st    -0.10191    0.11125  -0.916    0.361

(Dispersion parameter for gaussian family taken to be 0.9583229)

Number of Fisher Scoring iterations: 2

[1] "Table 6 (LATE/IV):"

t test of coefficients:

                          Estimate Std. Error t value  Pr(>|t|)
(Intercept)             -0.2114821  0.1550476 -1.3640 0.1726352
tchr_treated            -0.0498718  0.1747816 -0.2853 0.7753971
asnt_aft                 0.0030923  0.0987384  0.0313 0.9750170
math_1st                -0.0287668  0.0816325 -0.3524 0.7245580
factor(dist_stratum)22   0.9189529  0.5707644  1.6100 0.1074528
factor(dist_stratum)51   0.0204319  0.3148732  0.0649 0.9482646
factor(dist_stratum)52  -0.1930012  0.2494086 -0.7738 0.4390650
factor(dist_stratum)111  0.9060161  0.2514259  3.6035 0.0003170 ***
factor(dist_stratum)112  0.1274507  0.1509524  0.8443 0.3985364
factor(dist_stratum)201  0.0184632  0.2158478  0.0855 0.9318369
factor(dist_stratum)202  0.3897603  0.1565232  2.4901 0.0128027 *
factor(dist_stratum)241  0.4812907  0.3422602  1.4062 0.1597234
factor(dist_stratum)242  0.4858362  0.2839635  1.7109 0.0871600 .
factor(dist_stratum)281  0.2673338  0.1957411  1.3658 0.1720785
factor(dist_stratum)282  1.0771401  0.3514102  3.0652 0.0021870 **
factor(dist_stratum)341  0.1176844  0.1537056  0.7656 0.4439223
factor(dist_stratum)342  0.0259967  0.2081724  0.1249 0.9006233
factor(dist_stratum)351  1.0448632  0.2452357  4.2606 2.076e-05 ***
factor(dist_stratum)352  0.6276726  0.3201447  1.9606 0.0499826 *
factor(dist_stratum)371  0.9136327  0.2417061  3.7799 0.0001587 ***
factor(dist_stratum)372  0.2476686  0.2551344  0.9707 0.3317262
factor(dist_stratum)451  0.4501674  0.1566821  2.8731 0.0040815 **
factor(dist_stratum)452  0.7300646  0.1585093  4.6058 4.211e-06 ***
factor(dist_stratum)501  0.2507473  0.2609477  0.9609 0.3366442
factor(dist_stratum)502  0.2370147  0.2401423  0.9870 0.3237023
```

```
factor(dist_stratum)511  0.4838129  0.1859532  2.6018 0.0093011 **
factor(dist_stratum)512  0.4745071  0.3997011  1.1872 0.2352232
factor(dist_stratum)551 -0.5110011  0.1643663 -3.1089 0.0018884 **
factor(dist_stratum)552  0.6619861  0.2825698  2.3427 0.0191822 *
factor(dist_stratum)601 -0.4570397  0.1534413 -2.9786 0.0029097 **
factor(dist_stratum)602 -0.1227874  0.1901407 -0.6458 0.5184575
factor(dist_stratum)691  0.0604362  0.2150056  0.2811 0.7786519
factor(dist_stratum)692  0.4619547  0.2072222  2.2293 0.0258403 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
run_het_analysis <- function(df, label) {
  print(paste("--- Heterogeneity:", label, "---"))
  df <- df %>%
    mutate(
      girl = if_else(s_gender == 2, 1, 0),
      parent_sec_ed = if_else((f_educlevel >= 2 & f_educlevel <= 5) |
                              (m_educlevel >= 2 & m_educlevel <= 5), 1, 0,
 ↪  missing = 0)
    )

  svy_des <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                       weights = ~sch_wght, data = df, nest = TRUE)

  print("Table 7: Gender Interaction")
  print(summary(svyglm(std_irt ~ treat * girl + asnt_aft + math_1st,
                       design = svy_des)))

  print("Table 7: Parent Ed Interaction")
  print(summary(svyglm(std_irt ~ treat * parent_sec_ed + asnt_aft +
   ↪  math_1st,
                       design = svy_des)))

  print("Table 8: Median Regression (Weighted / Non-Clustered)")
  try(print(summary(rq(std_irt ~ treat, tau = 0.5,
                       data = df, weights = sch_wght), se = "nid")))
}

run_het_analysis(df_g9m, "Grade 9 Math")
```

```
[1] "--- Heterogeneity: Grade 9 Math ---"
[1] "Table 7: Gender Interaction"

Call:
svyglm(formula = std_irt ~ treat * girl + asnt_aft + math_1st,
    design = svy_des)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
```

```
    data = df, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.24087    0.12039   2.001   0.0472 *
treat       -0.04860    0.11156  -0.436   0.6637
girl        -0.33960    0.05862  -5.793 3.96e-08 ***
asnt_aft     0.07239    0.11723   0.617   0.5379
math_1st    -0.17594    0.10894  -1.615   0.1084
treat:girl   0.02518    0.07178   0.351   0.7262
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.9336059)

Number of Fisher Scoring iterations: 2

[1] "Table 7: Parent Ed Interaction"

Call:
svyglm(formula = std_irt ~ treat * parent_sec_ed + asnt_aft +
    math_1st, design = svy_des)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df, nest = TRUE)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.12112    0.11831  -1.024   0.3076
treat               -0.02473    0.11792  -0.210   0.8342
parent_sec_ed        0.33532    0.06150   5.453 2.01e-07 ***
asnt_aft             0.06376    0.11592   0.550   0.5831
math_1st            -0.18012    0.10511  -1.714   0.0887 .
treat:parent_sec_ed -0.01646    0.08797  -0.187   0.8518
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.9331498)

Number of Fisher Scoring iterations: 2

[1] "Table 8: Median Regression (Weighted / Non-Clustered)"

Call: rq(formula = std_irt ~ treat, tau = 0.5, data = df, weights = sch_wght)

tau: [1] 0.5

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

```
(Intercept) -0.07003  0.02570   -2.72510  0.00645
treat       -0.00451  0.03807   -0.11854  0.90564
```

```
run_het_analysis(df_g9s, "Grade 9 Science")
```

```
[1] "--- Heterogeneity: Grade 9 Science ---"
[1] "Table 7: Gender Interaction"

Call:
svyglm(formula = std_irt ~ treat * girl + asnt_aft + math_1st,
    design = svy_des)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22536    0.10482   2.150   0.0332 *
treat       -0.11266    0.10115  -1.114   0.2672
girl        -0.35372    0.05246  -6.742 3.21e-10 ***
asnt_aft     0.06582    0.09332   0.705   0.4817
math_1st    -0.12112    0.09411  -1.287   0.2001
treat:girl   0.08639    0.06577   1.313   0.1911
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.8952229)

Number of Fisher Scoring iterations: 2

[1] "Table 7: Parent Ed Interaction"

Call:
svyglm(formula = std_irt ~ treat * parent_sec_ed + asnt_aft +
    math_1st, design = svy_des)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df, nest = TRUE)

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.1476844  0.1001280  -1.475    0.142
treat               -0.0002294  0.1036666  -0.002    0.998
parent_sec_ed        0.3333013  0.0579817   5.748 4.92e-08 ***
asnt_aft             0.0670176  0.0919150   0.729    0.467
math_1st            -0.1294755  0.0918330  -1.410    0.161
treat:parent_sec_ed -0.1195555  0.0956541  -1.250    0.213
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.8997185)

Number of Fisher Scoring iterations: 2

[1] "Table 8: Median Regression (Weighted / Non-Clustered)"

Call: rq(formula = std_irt ~ treat, tau = 0.5, data = df, weights = sch_wght)

tau: [1] 0.5

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept) -0.07882  0.02660   -2.96275  0.00306
treat       -0.03792  0.03467   -1.09360  0.27417
```

```
ht_teacher <- read_dta("HTteacher_c.dta") %>% rename_with(tolower) %>%
  mutate(t_learning_recoded = if_else(t_learning %in% c(1,2), 0,
                                      if_else(t_learning==3, 1, NA_real_)))
ht_data <- inner_join(df_school_base, ht_teacher, by = "schoolid")
ht_design <- svydesign(id = ~schoolid, weights = ~sch_wght,
                       strata = ~dist_stratum, data = ht_data, nest=TRUE)

print("Table 9: Teacher Interest in Learning")
```

```
[1] "Table 9: Teacher Interest in Learning"
```

```
m_a18 <- svyglm(t_learning >= 3 ~ treat, design = ht_design,
                family = quasibinomial())
print(summary(m_a18))
```

```
Call:
svyglm(formula = t_learning >= 3 ~ treat, design = ht_design,
    family = quasibinomial())

Survey design:
svydesign(id = ~schoolid, weights = ~sch_wght, strata = ~dist_stratum,
    data = ht_data, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1243     0.1646  -0.755    0.451
treat        -0.0746     0.2676  -0.279    0.781

(Dispersion parameter for quasibinomial family taken to be 1.000816)
```

```
Number of Fisher Scoring iterations: 3
```

```r
# 1. Enumerator Obs
df_att <- read_dta("Teacher_c.dta") %>% rename_with(tolower) %>%
  mutate(t_absent_obs = if_else(t_absence == 1, 1, 0)) %>%
  inner_join(df_school_base, by = "schoolid")

att_design <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                        weights = ~sch_wght, data = df_att, nest = TRUE)
print("Table 10 (Enumerator Obs): Impact on Absence")
```

```
[1] "Table 10 (Enumerator Obs): Impact on Absence"
```

```r
print(summary(svyglm(t_absent_obs ~ treat, design = att_design)))
```

```
Call:
svyglm(formula = t_absent_obs ~ treat, design = att_design)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df_att, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.91530    0.02047  44.705   <2e-16 ***
treat       -0.01043    0.03005  -0.347    0.729
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.08183608)

Number of Fisher Scoring iterations: 2
```

```r
# 2. Student Report (Math Teacher)
df_stu_att <- read_dta("Grade09_c.dta") %>% rename_with(tolower) %>%
  mutate(t_absent_stu = if_else(ma_t_absent >= 3, 1, 0)) %>%
  inner_join(df_school_base, by = "schoolid")

stu_att_design <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                            weights = ~sch_wght, data = df_stu_att, nest =
                             ↪ TRUE)
print("Table 10 (Student Report - Math): Impact on Frequent Absence")
```

```
[1] "Table 10 (Student Report - Math): Impact on Frequent Absence"
```

```r
print(summary(svyglm(t_absent_stu ~ treat, design = stu_att_design)))
```

```

Call:
svyglm(formula = t_absent_stu ~ treat, design = stu_att_design)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df_stu_att, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22333    0.01976  11.300   <2e-16 ***
treat       -0.02576    0.02921  -0.882    0.379
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.1700311)

Number of Fisher Scoring iterations: 2
```

```r
# Math (Table 11)
print("Table 11 (Math): Gives Homework Everyday")
```

```
[1] "Table 11 (Math): Gives Homework Everyday"
```

```r
df_stu_prac <- read_dta("Grade09_c.dta") %>% rename_with(tolower) %>%
  mutate(hw_everyday = if_else(ma_hwfreq == 5, 1, 0)) %>%
  inner_join(df_school_base, by = "schoolid")

prac_design <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                      weights = ~sch_wght, data = df_stu_prac, nest = TRUE)
print(summary(svyglm(hw_everyday ~ treat, design = prac_design)))
```

```

Call:
svyglm(formula = hw_everyday ~ treat, design = prac_design)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df_stu_prac, nest = TRUE)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.819292   0.023837  34.371   <2e-16 ***
treat       -0.004291   0.034669  -0.124    0.902
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.1530276)

Number of Fisher Scoring iterations: 2
```

```
# Science (Table 12)
print("Table 12 (Science): Gives Homework Everyday")
```

```
[1] "Table 12 (Science): Gives Homework Everyday"
```

```
df_stu_prac_s <- read_dta("Grade09_c.dta") %>% rename_with(tolower) %>%
  mutate(hw_everyday_s = if_else(sc_hwfreq == 5, 1, 0)) %>%
  inner_join(df_school_base, by = "schoolid")

prac_design_s <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                          weights = ~sch_wght, data = df_stu_prac_s, nest =
                            ↪ TRUE)
print(summary(svyglm(hw_everyday_s ~ treat, design = prac_design_s)))
```

```
Call:
svyglm(formula = hw_everyday_s ~ treat, design = prac_design_s)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df_stu_prac_s, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.47944    0.03876   12.37   <2e-16 ***
treat        0.02227    0.05436    0.41    0.683
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2558342)

Number of Fisher Scoring iterations: 2
```

```
print("Table 13: Uses Local Materials (Teacher Report)")
```

```
[1] "Table 13: Uses Local Materials (Teacher Report)"
```

```
df_t_prac <- read_dta("Teacher_c.dta") %>% rename_with(tolower) %>%
  mutate(use_local = if_else(t_ms_prep_material == 3, 1, 0)) %>%
  inner_join(df_school_base, by = "schoolid")
```

```r
t_prac_design <- svydesign(id = ~schoolid, strata = ~dist_stratum,
                           weights = ~sch_wght, data = df_t_prac, nest = TRUE)
print(summary(svyglm(use_local ~ treat, design = t_prac_design)))
```

```
Call:
svyglm(formula = use_local ~ treat, design = t_prac_design)

Survey design:
svydesign(id = ~schoolid, strata = ~dist_stratum, weights = ~sch_wght,
    data = df_t_prac, nest = TRUE)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11960    0.02897   4.129 5.74e-05 ***
treat       -0.05675    0.03673  -1.545    0.124
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.08382066)

Number of Fisher Scoring iterations: 2
```

```r
process_teacher_evals <- function(filename, subject) {
  data <- read_dta(filename) %>% zap_labels()
  long_data <- data %>%
    select(starts_with(paste0("t_", subject))) %>%
    pivot_longer(everything(), names_to = "q", values_to = "val") %>%
    filter(grepl("_b$", q))

  print(paste("Table:", subject, "Teacher Knowledge (% Correct)"))
  print(long_data %>% summarize(pct_correct = mean(val == 1, na.rm = TRUE)
  ↪  * 100))
}

print("Table 14 (Math):")
```

```
[1] "Table 14 (Math):"
```

```r
process_teacher_evals("Teachereval_math_c.dta", "math")
```

```
[1] "Table: math Teacher Knowledge (% Correct)"
# A tibble: 1 x 1
  pct_correct
        <dbl>
1        51.7
```

```r
print("Table 15 (Science):")
```

```
[1] "Table 15 (Science):"
```

```r
process_teacher_evals("Teachereval_science_c.dta", "science")
```

```
[1] "Table: science Teacher Knowledge (% Correct)"
# A tibble: 1 x 1
  pct_correct
        <dbl>
1        35.2
```

```r
process_student_perf <- function(filename, prefix) {
  data <- read_dta(filename) %>% zap_labels()
  cols <- names(data)[grepl(paste0("^", prefix, "_\\d+$"), names(data))]
  results <- data %>%
    summarize(across(all_of(cols), ~mean(. == 1, na.rm=TRUE))) %>%
    pivot_longer(everything(), names_to="Question",
     ↪  values_to="Pct_Correct")

  print(paste("Student Performance on Below-Grade Items:", prefix))
  print(head(results, 8))
}

print("Table 16 (Math G9):")
```

```
[1] "Table 16 (Math G9):"
```

```r
process_student_perf("Math09_endline_IRT.dta", "Math")
```

```
[1] "Student Performance on Below-Grade Items: Math"
# A tibble: 8 x 2
  Question Pct_Correct
  <chr>          <dbl>
1 Math_002       0.442
2 Math_003       0.868
3 Math_004       0.874
4 Math_005       0.487
5 Math_006       0.460
6 Math_007       0.529
7 Math_008       0.395
8 Math_009       0.441
```

```r
print("Table 17 (Science G9):")
```

```
[1] "Table 17 (Science G9):"
```

```r
process_student_perf("Sci09_endline_IRT.dta", "Sci")
```

```
[1] "Student Performance on Below-Grade Items: Sci"
# A tibble: 8 x 2
  Question Pct_Correct
  <chr>          <dbl>
1 Sci_002        0.796
2 Sci_003        0.273
3 Sci_004        0.853
4 Sci_005        0.368
5 Sci_006        0.816
6 Sci_007        0.657
7 Sci_008        0.749
8 Sci_009        0.415
```