```
options(scipen=999) #No scientific notation
options(digits = 2) #Round results two decimals

knots_color_binary <- c("#5f5758","#ff3057") #Knots theme

pacman::p_load(tidyverse, #Data manipulation
               gt, #Tables
               gtsummary, #Descriptive table
               sjmisc, #Label manipulation
               labelled, #Label manipulation
               skimr, #Data manipulation
               sjPlot, #Correlation
               corrplot, #Correlation
               Hmisc, #Correlation
               lme4, #Multilevel modeling
               reghelper, #Get ICC
               texreg, #View models
               AICcmodavg, #Best fit model
               responsePatterns, #Check items
               ggrepel, #labels plots
               ggpubr, #arrange plots
               ggExtra #marginal plots
               )


#Chile ICILS 2018 data
student_proc_2018 <-
readRDS("../../../input/proc_data/03_student_proc_2018.rds")

#Select variables
data <- student_proc_2018 |>
  select(
    #ID
    idschool,idstud,
    #Student level
    s_sex,s_pv1cil,s_speceff,s_geneff,
    #School level
    c_pv1cil,c_s_f_ratio,c_speceff,c_geneff,
    #Control variables
    s_hisced,s_homlit,s_hisei,s_nisb,
    #Items specific self-efficacy
    is2g27b,is2g27e,is2g27g,is2g27h,
    #Items general self-efficacy
    is2g27a,is2g27c,is2g27d,is2g27i,is2g27j,is2g27k,is2g27l,is2g27m,
    #Weights
    starts_with(c("wg","tot"))
    )

#Delete missing data labels
val_label(data$s_sex,8) <- NULL
val_label(data$s_sex,9) <- NULL
```

```
val_label(data$is2g27b,8) <- NULL
val_label(data$is2g27b,9) <- NULL
val_label(data$is2g27e,8) <- NULL
val_label(data$is2g27e,9) <- NULL
val_label(data$is2g27g,8) <- NULL
val_label(data$is2g27g,9) <- NULL
val_label(data$is2g27h,8) <- NULL
val_label(data$is2g27h,9) <- NULL
val_label(data$is2g27a, 8) <- NULL
val_label(data$is2g27a, 9) <- NULL
val_label(data$is2g27c, 8) <- NULL
val_label(data$is2g27c, 9) <- NULL
val_label(data$is2g27d, 8) <- NULL
val_label(data$is2g27d, 9) <- NULL
val_label(data$is2g27i, 8) <- NULL
val_label(data$is2g27i, 9) <- NULL
val_label(data$is2g27j, 8) <- NULL
val_label(data$is2g27j, 9) <- NULL
val_label(data$is2g27k, 8) <- NULL
val_label(data$is2g27k, 9) <- NULL
val_label(data$is2g27l, 8) <- NULL
val_label(data$is2g27l, 9) <- NULL
val_label(data$is2g27m, 8) <- NULL
val_label(data$is2g27m, 9) <- NULL

#Change labels geneff items
var_label(data) <- list(
  is2g27a = "Edit graphic images",
  is2g27c = "Write or edit text",
  is2g27d = "Search and find information on the Internet",
  is2g27i = "Create a multimedia presentation",
  is2g27j = "Upload multimedia information to an online profile",
  is2g27k = "Insert an image into a document/message",
  is2g27l = "Install a program/application",
  is2g27m = "Judging the veracity of information on the Internet"
)

#Change labels specific items
var_label(data) <- list(
  is2g27b = "Create a database",
  is2g27e = "Build a web page",
  is2g27g = "Create a computer program",
  is2g27h = "Setting up a local network"
)

# Create gender composition categorical variable
data <- data |>
  mutate(c_gender_type=case_when(
    c_s_f_ratio <= 0.33 ~"Masculinized school (0-33% girls)",
    c_s_f_ratio > 0.34 & c_s_f_ratio < 0.66 ~"Mixed schools (34%-66%
girls)",
    c_s_f_ratio >= 0.67 ~"Feminized school (67%-100% girls)"
  ))
```

```
# Recode CIL
data <- data |>
  mutate(s_pv1cil=s_pv1cil*0.1,
         c_pv1cil=c_pv1cil*0.1)

var_label(data)<- list(
  s_pv1cil = "Computer Literacy Score",
  s_geneff = "Basic digital self-efficacy",
  s_speceff = "Advanced digital self-efficacy",
  s_sex = "Student gender (Girl=1)",
  s_nisb = "Household Socioeconomic Level Index",
  c_pv1cil = "School average Computer literacy",
  c_gender_type = "School gender composition"
)

# Delete less than 10 schools
lost<-data |>
  group_by(idschool)|>
  summarise(num_cases = n()) %>%
  filter(num_cases < 10)

'%!in%' <- function(x,y)!('%in%'(x,y))

data <- data |>
  filter(idschool %!in% lost$idschool)
```


INVALSI

Do female students underestimate their technological skills?

Gender differences in general and specialized digital self-efficacy at school

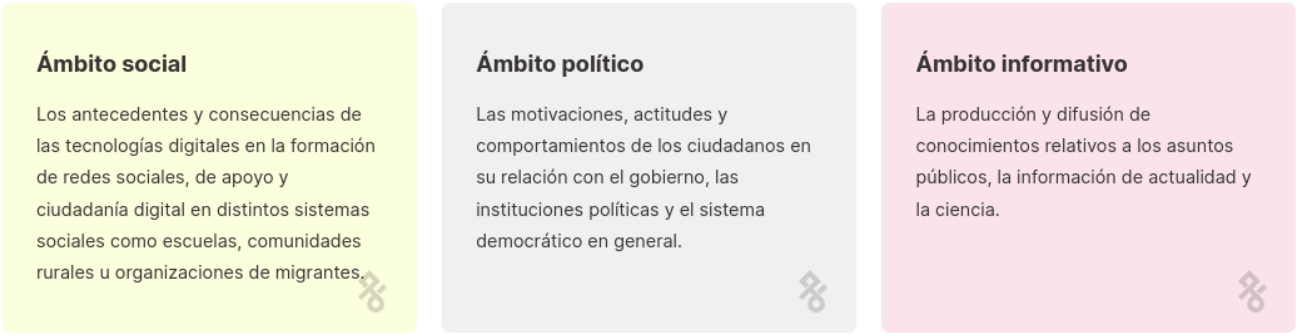
Juan Carlos Castillo, Daniel Miranda & Nicolas Tobar

Universidad de Chile & Millennium Nucleus of Inequalities and Digital Opportunities (*nudos.cl*)

IX Seminar Data of and for the Educational System: Tools for Research and Teaching- Rome, 17 – 18 – 19 October 2024

{width="80%" fig-align="center"}

Líneas de investigación:



nudos.cl

Introduction

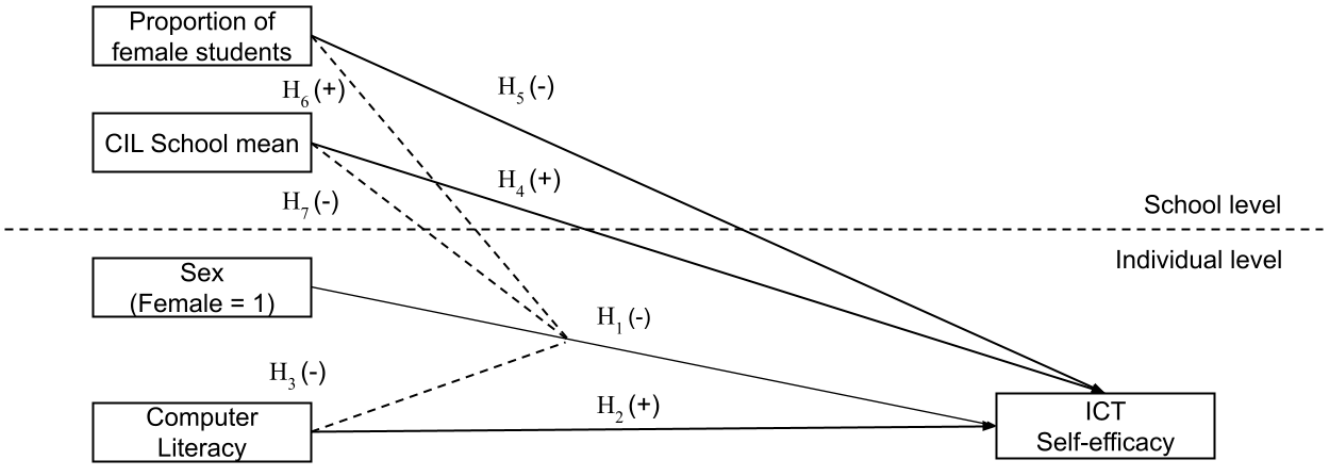
- Mass use of ICTs has attenuated differences in use by gender [@hargittaiDifferencesActualPerceived2006].
- Female students have higher literacy rates but less confidence in their abilities than male students [@punterGenderDifferencesComputer2017].
- Self-efficacy: "*a functional self-awareness in which students reflect on their personal efficacy, thoughts, actions, the meaning of their pursuits, and make corrective adjustments if necessary*" [@banduraPsychologyHumanAgency2006].

Why do girls underestimate themselves?

- Female Technological Stereotype Threat [@turnerSocialIdentityTheory2004; @steeleStereotypeThreatIntellectual1995]. Does it persist as women increase literacy?
- Evidence attributes attitudinal inequality to factors external to school [@vekiriGenderIssuesTechnology2008].
- Composition effects: women increase performance and satisfaction when surrounded by women [@kirschGenderCompositionCorporate2018]. Girls support girls?

How do the gender composition of the school and the level of digital literacy affect the levels of technological self-efficacy of eighth-grade students in Chile? {data-background-color="#5f5758"}

Hypothesis



{width="90%" fig-align="center"}

ICILS data

- Measurement in November 2017, Publication 2018.
- 46,561 8th grade students from 2226 schools in 12 countries. Complex three-stage sample.
- 3092 students and 178 schools in Chile.
- CIL test and questionnaire on access, use, perceptions and attitudes towards ICT.
- 14 schools and 86 students are eliminated for n<10.
- Exploration with multilevel regression models (Context variables are considered)

Variable Self-efficacy {.smaller}

To what extent can you perform each of these tasks using ICT?

Answers: 1. I know how to do it. 2. I don't know how to do it but I feel capable of learning it. 3. No, I don't feel capable of learning it.

Basic

- Edit digital photographs or other graphic images.
- Write or edit text for a school assignment.
- Search and find relevant information on the Internet for a school project.
- Create a multimedia presentation (with sound, images or video).
- Upload text, images or video to an online profile.\
- Insert an image into a document or message.
- Install a program or [app].
- Judge whether you can trust the information you find on the Internet.

Advanced

- Create a database (for example, using [Microsoft Access®]).
- Build or edit a web page.
- Create a computer program, macro, or [app] (for example, in [Basic, Visual Basic]).
- Set up a local area network of computers or other ICT.

Results

Level 1 descriptives

```
data |>
  select(
  s_sex,
  s_geneff,
  s_speceff,
  s_pv1cil,
  s_nisb
) |>
  mutate(
    s_sex=to_label(s_sex))|>
  tbl_summary(
    statistic = list(
              all_continuous() ~ "{mean} ({sd}) [{median}]",
              all_categorical() ~ "{n} / {N} ({p}%)"
              ),
    include = everything(),
    missing_text = "Missing values (NA)"
  )|>
  modify_header(label="**Variables**") |>
  as_gt()|>
  tab_options(
    table.border.top.color = "white",
    heading.title.font.size = px(16),
    column_labels.border.top.width = 3,
    column_labels.border.top.color = "black",
    column_labels.border.bottom.width = 3,
    column_labels.border.bottom.color = "black",
    table_body.border.bottom.color = "black",
    table.border.bottom.color = "white",
    table.width = pct(100),
    table.background.color = "white",
    table.font.size = 15
  ) |>
  # center column text
  cols_align(align="center")|>
  tab_style(
    style = list(
      # remove horizontal lines
```

```
      cell_borders(
        sides = c("top", "bottom"),
        color = "white",
        weight = px(1)
      ),
      #center text
      cell_text(
        align="center"
      ),
    # remove row striping in Markdown documents
    cell_fill(color = "white", alpha = NULL)
      ),
      #do this for all columns and rows
    locations = cells_body(
      columns = everything(),
      rows = everything()
    ))
```

Level 2 descriptives

```
list_school <- list(
     aggregate(s_pv1cil ~ idschool, data=data, mean),
     aggregate(c_s_f_ratio ~ idschool, data=data, mean),
     data|>distinct(idschool,c_gender_type)
)

Reduce(function(x,y) merge(x,y, all=TRUE),
        list_school) |>

  rename("Average Digital Literacy"=s_pv1cil,
         "Gender Composition Type (Categorical)" = c_gender_type) |>

    tbl_summary(include = c("Average Digital Literacy",
         "Gender Composition Type (Categorical)"),
              statistic = list(
              all_continuous() ~ "{mean} ({sd}) [{median}]",
              all_categorical() ~ "{n} / {N} ({p}%)"
              )) |>
  modify_header(label="**Variables**") |>
  as_gt()|>
  tab_options(
    table.border.top.color = "white",
    heading.title.font.size = px(16),
    column_labels.border.top.width = 3,
    column_labels.border.top.color = "black",
    column_labels.border.bottom.width = 3,
    column_labels.border.bottom.color = "black",
    table_body.border.bottom.color = "black",
    table.border.bottom.color = "white",
    table.width = pct(100),
```

```
    table.background.color = "white",
    table.font.size = 15
  ) |>
  # center column text
  cols_align(align="center")|>
  tab_style(
    style = list(
      # remove horizontal lines
      cell_borders(
        sides = c("top", "bottom"),
        color = "white",
        weight = px(1)
      ),
      #center text
      cell_text(
        align="center"
      ),
    # remove row striping in Markdown documents
    cell_fill(color = "white", alpha = NULL)
      ),
      #do this for all columns and rows
    locations = cells_body(
      columns = everything(),
      rows = everything()
    ))
```

Items Basic Self-Efficacy

```
geneff_plot <- data |>
  mutate(s_sex=to_label(s_sex),
         is2g27a=to_label(is2g27a),
         is2g27c=to_label(is2g27c),
         is2g27d=to_label(is2g27d),
         is2g27i=to_label(is2g27i),
         is2g27j=to_label(is2g27j),
         is2g27k=to_label(is2g27k),
         is2g27l=to_label(is2g27l),
         is2g27m=to_label(is2g27m))

rbind(
  geneff_plot|>count(variable=is2g27a)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Edit graphic images"),
  geneff_plot|>count(variable=is2g27c)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Write or edit text"),
  geneff_plot|>count(variable=is2g27d)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Search and find information on the Internet"),
  geneff_plot|>count(variable=is2g27i)|>drop_na()|>
```

```
    mutate(prop=n/sum(n),
           name="Create a multimedia presentation"),
  geneff_plot|>count(variable=is2g27j)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Upload media to an online profile"),
  geneff_plot|>count(variable=is2g27k)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Insert an image into a document/message"),
  geneff_plot|>count(variable=is2g27l)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Install a program/application"),
  geneff_plot|>count(variable=is2g27m)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Judging the veracity of information on the Internet")
  ) |>
  mutate(name=factor(name,
                     levels=c(
                       "Write or edit text",
                       "Search and find information on the Internet",
                      "Insert an image into a document/message",
                       "Install a program/application",
                       "Upload media to an online profile",
                       "Edit graphic images",
                       "Judging the veracity of information on the
Internet",
                       "Create a multimedia presentation")),
         variable=factor(variable, levels = c(
  "I don't think I could do this.",
  "I have never done this but I could work out how to do this.",
  "I know how to do this."
  )
         )
         )|>
  mutate(prop=round(prop,2)) |>
  ggplot(aes(x=name,y=prop,fill=variable))+
  geom_bar(position = "fill",stat="identity",color="black")+
  coord_flip()+
  scale_x_discrete(labels = function(x) str_wrap(x, width = 25))+
  labs(x="",y="",fill="",
       caption="Note: The percentages are rounded with 2 decimals. If
values don't sum 100% is because the rounding")+
  theme_minimal()+
  theme(legend.position = "top")+
  scale_fill_manual(
    values=c("#ff3057","#5f5758","white"),
    limits=c(
  "I know how to do this.",
  "I have never done this but I could work out how to do this.",
  "I don't think I could do this."
  ),
    labels = ~ stringr::str_wrap(.x, width = 30))+
  scale_y_continuous(labels = scales::percent)+
  geom_text(aes(label=paste0(prop*100,"%")),position =
```

```
  position_fill(vjust=0.5))
```

Items Advanced Self-Efficacy

```
speceff_plot <- data |>
  mutate(s_sex=to_label(s_sex),
         is2g27b=to_label(is2g27b),
         is2g27e=to_label(is2g27e),
         is2g27g=to_label(is2g27g),
         is2g27h=to_label(is2g27h))

rbind(
  speceff_plot|>count(variable=is2g27b)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Create a database"),
  speceff_plot|>count(variable=is2g27e)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Build a web page"),
  speceff_plot|>count(variable=is2g27g)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Create a program/application"),
  speceff_plot|>count(variable=is2g27h)|>drop_na()|>
    mutate(prop=n/sum(n),
           name="Setting up a local network")
) |>
  mutate(prop=round(prop,2)) |>
  mutate(name=factor(name,levels=c(
    "Create a database",
    "Setting up a local network",
    "Build a website",
    "Create a program/application"
  )),
  variable=factor(variable, levels = c(
  "I don't think I could do this.",
  "I have never done this but I could work out how to do this.",
  "I know how to do this."
  )
  )
  )|>
  ggplot(aes(x=name,y=prop,fill=variable))+
  geom_bar(position = "fill",stat="identity",color="black",width=0.6)+
  scale_fill_manual(
    values=c("#ff3057","#5f5758","white"),
    limits=c(
  "I know how to do this.",
  "I have never done this but I could work out how to do this.",
  "I don't think I could do this."
  ),
    labels = ~ stringr::str_wrap(.x, width = 30))+
  coord_flip()+
```

```
  scale_x_discrete(labels = function(x) str_wrap(x, width = 25))+
  labs(x="",y="",fill="",
       caption="Note: The percentages are rounded with 2 decimals. If
values don't sum 100% is because the rounding")+
  theme_minimal()+
  theme(legend.position = "top")+
  scale_y_continuous(labels = scales::percent)+
  geom_text(aes(label=paste0(prop*100,"%")),position =
position_fill(vjust=0.5))
```

Self-efficacy items by gender

```
data |>
  mutate(s_sex=to_label(s_sex))|>
  mutate_at(vars(starts_with("is2g")), ~
            ifelse(.==1,1,0)) |> #Dummy
  rename(gender=s_sex)|>
  group_by(gender) |>
  summarise(across(starts_with("is2g"),
                   mean, na.rm = TRUE)) |>
  pivot_longer(cols = starts_with(c("is2g")),
               names_to = "items",
               values_to = "puntaje") |>
  mutate(type=ifelse(items%in%c("is2g27b","is2g27e",
                                "is2g27g","is2g27h"),
                                "Specialized ICT self-efficacy",
                                "General ICT self-efficacy")) |>

  mutate(items=
  factor(items,
  levels=c(
  "is2g27b",
  "is2g27h",
  "is2g27e",
  "is2g27g",
  "is2g27c",
  "is2g27d",
  "is2g27k",
  "is2g27l",
  "is2g27j",
  "is2g27a",
  "is2g27m",
  "is2g27i"),
  labels=c(
  "Crear una base de datos",
  "Configurar una red de área local***",
  "Construir una página web",
  "Crear un programa/aplicación informática***",
  "Escribir o editar texto***",
  "Buscar y encontrar información en internet**",
```

```
  "Insertar una imagen en un documento/mensaje***",
  "Instalar un programa/aplicación*",
  "Subir multimedia a un perfil en línea**",
  "Editar imágenes gráficas***",
  "Evaluar la veracidad de la información en internet",
  "Crear una presentación multimedia**"
))|>

  ggplot(aes(x=to_label(puntaje),y=to_label(items)))+
    geom_line(aes(group=to_label(items)),color="#E7E7E7",linewidth=2.0)+
    geom_point(aes(shape=gender,color=gender),size=4)+
    theme_minimal()+
    labs(x="Proporción de 'Sé como hacerlo'",y="",color="Gender",
         caption= "Signif. codes: <0.001 = ***, <0.01 = **, <0.05 = *")+
    scale_color_manual(values = c("#5f5758","#ff3057"))+
    theme(legend.position = "top")+
    scale_y_discrete(labels=function(x) str_wrap(x,width = 50))+
    facet_wrap(~type,ncol=1,scales="free_y")+
    guides(shape=guide_legend(title="Gender"))
```

Multilevel Regression Models (Forest Plot) {.smaller}

```
m1_1_geneff <- lmer(s_geneff ~
                  s_sex + s_pv1cil +
                  c_gender_type + c_pv1cil +
                  s_nisb +
                  (1 | idschool),  data=data)

sjPlot::plot_model(m1_1_geneff,
                   wrap.labels = 20,
                    show.p = T,
                    show.values = T,
                    show.intercept = T,
                    colors = c("#ff3057","#ff3057")) +
  theme_minimal()
```

Multilevel Regression Models (Forest Plot) {.smaller}

```
m1_1_geneff <- lmer(s_geneff ~
                  s_sex + s_pv1cil +
                  c_gender_type + c_pv1cil +
                  s_nisb +
                  (1 | idschool),  data=data)

m1_1_speceff <- lmer(s_speceff ~
                  s_sex + s_pv1cil +
                  c_gender_type + c_pv1cil +
                  s_nisb +
                  (1 | idschool),  data=data)
```

```
sjPlot::plot_models(m1_1_geneff,m1_1_speceff,
                    wrap.labels = 20,
                    show.p = T,
                    show.values = T,
                    show.intercept = T,
                    colors = c("#5f5758","#ff3057")) +
  theme_minimal()
```

- ICC: General A. (0.05)/ Specific A. (0.06).

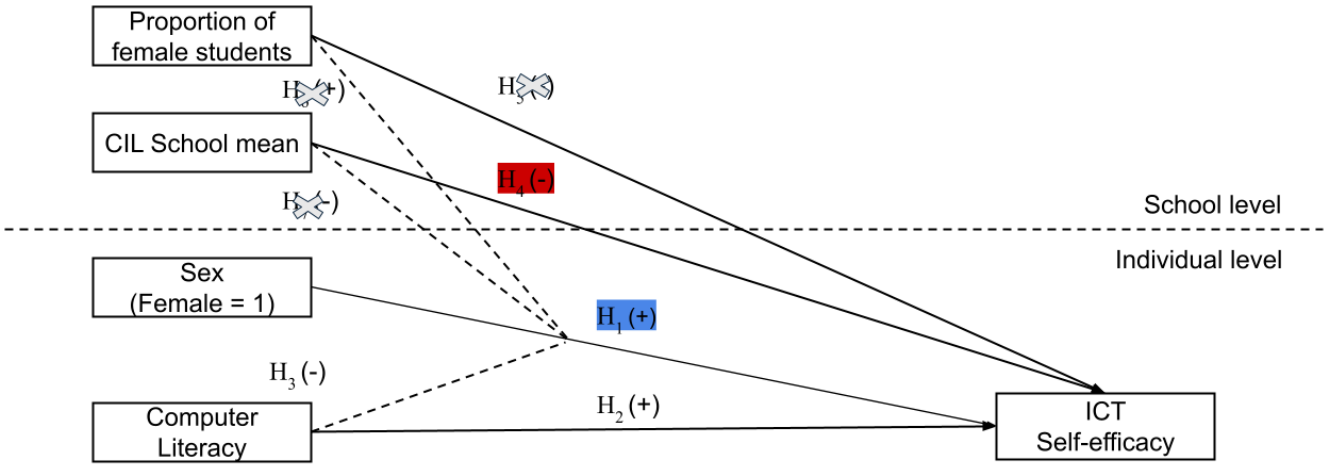
Sex-CIL Interaction Basic Self-Efficacy {.smaller}

```
m2_geneff <- lm(s_geneff ~
                  s_sex + s_pv1cil +
                  s_sex*s_pv1cil +
                  s_nisb, data=data)

plot_model(m2_geneff, type = "pred", terms = c("s_pv1cil", "s_sex"))+
  theme_minimal()+
  labs(title="")
```

Sex-CIL Interaction Advanced Self-Efficacy {.smaller}

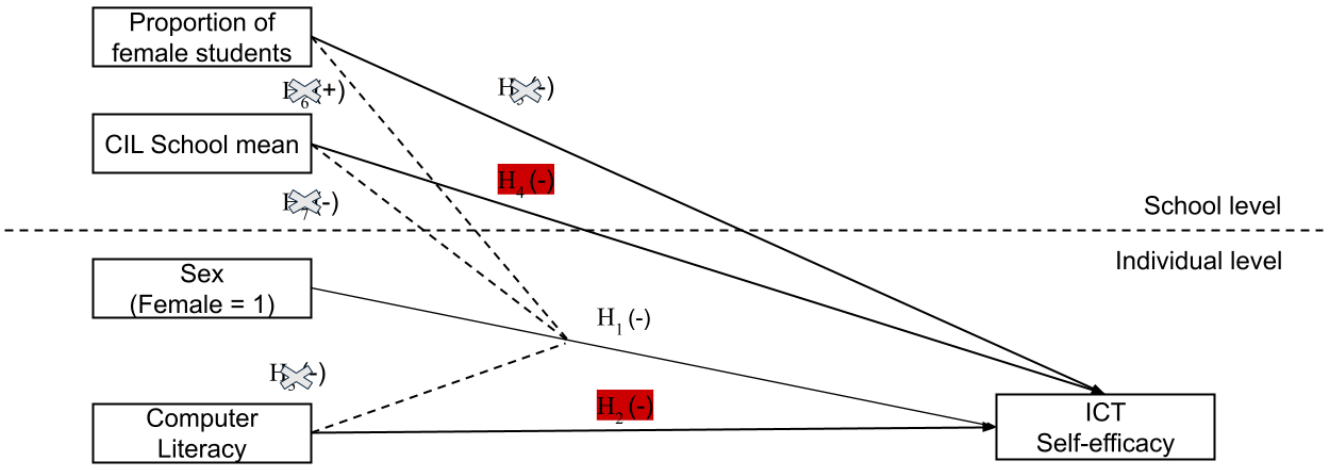
```
m2_speceff <- lm(s_speceff ~
                  s_sex + s_pv1cil +
                  s_sex*s_pv1cil +
                  s_nisb, data=data)

plot_model(m2_speceff, type = "pred", terms = c("s_pv1cil", "s_sex"))+
  theme_minimal()+
  labs(title="")
```

Discussion Basic Self-Efficacy



{width="90%" fig-align="center"}

Discussion Advanced Self-Efficacy



{width="90%" fig-align="center"}

Conclusions {data-background-color="black"}

- Limitations:

    - Causality.
    - Estimated proportion of women in the sample.

- Projections

    - Time comparison 2013-2018-2023.
    - International comparison.
    - SIMCE data cross-referencing.

Thank you so much!

- **NUDOS website:** www.nudos.cl
- **Github repository of the project:** https://github.com/milenio-nudos/multilevel_icils_gender_chile

Appendix

Correlations: Dunning-Kruger effect?

- As students become more knowledgeable, they tend to underestimate their advanced abilities. This effect is even stronger for women.

```
#| label: fig-corr
#| fig-cap: "Gráficos de correlaciones"
#| fig-subcap:
#| - "Correlaciones generales"
#| - "Refuerzo de correlaciones por género"
#| layout-ncol: 2

data_corr <- data |> select(s_sex,
                            s_geneff,s_speceff,
                            s_pv1cil
                            )

data_corr <- rcorr(as.matrix(data_corr))

rownames(data_corr$r)<-c(
  "A. Género (1=Girl)",
  "B. Autoeficacia básica",
  "C. Autoeficacia avanzada",
  "D. Alfabetización computacional"
)

colnames(data_corr$r) <-c("1=Girl", "(B)","(C)","(D)")

corrplot(data_corr$r, p.mat = data_corr$P, method = 'color', type =
'lower', insig='blank',
         tl.col = "black",bg="white",na.label="-",
         addCoef.col ='black', number.cex = 0.8, diag=FALSE,
         sig.level = 0.05)
```

```
rbind(

data |>
  group_by(s_sex) |>
  dplyr::summarize(corr=cor(s_geneff,s_pv1cil,use="complete.obs")|>
                     round(2))|>
  mutate(group="Autoeficacia Básica - Alfabetización")|>
  to_label(s_sex),

data |>
  group_by(s_sex) |>
  dplyr::summarize(corr=cor(s_speceff,s_pv1cil,use="complete.obs")|>
                     round(2))|>
  mutate(group="Autoeficacia Avanzada - Alfabetización")|>
  to_label(s_sex)
) |>
  ggplot(aes(x=group,y=corr,color=s_sex))+
  geom_point()+
  geom_text(aes(label=corr, vjust=1.2),size=4.5)+
  theme_minimal()+
  scale_color_manual(values = c("#5f5758","#ff3057"))+
  scale_y_continuous(limits = c(-0.2,0.4))+
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank())+
  labs(color="Gender",
       title= "Correlación Autoeficacia - Alfabetización por género")
```

Mean differences in indices

- There are significant differences in literacy (CIL) and advanced self-efficacy.
- Girls tend to score higher on the test, but boys report greater skills in specialized tasks

```
data |>
  mutate(s_sex=to_label(s_sex)) |>
  tbl_summary(include = c(s_pv1cil,s_geneff,s_speceff),
              by = s_sex,
              missing = "no",
              statistic = list(
                all_continuous() ~ "{mean} ({sd})",
                all_categorical() ~ "{n} / {N} ({p}%)")
              ) |>
  add_p() |>
  add_n() |>
  modify_header(label="**Variables**") |>
  add_overall()|>
  as_gt()|>
  tab_options(
    table.border.top.color = "white",
    heading.title.font.size = px(16),
    column_labels.border.top.width = 3,
```

```
    column_labels.border.top.color = "black",
    column_labels.border.bottom.width = 3,
    column_labels.border.bottom.color = "black",
    table_body.border.bottom.color = "black",
    table.border.bottom.color = "white",
    table.width = pct(100),
    table.background.color = "white",
    table.font.size = 15
  ) |>
  # center column text
  cols_align(align="center")|>
  tab_style(
    style = list(
      # remove horizontal lines
      cell_borders(
        sides = c("top", "bottom"),
        color = "white",
        weight = px(1)
      ),
      #center text
      cell_text(
        align="center"
      ),
    # remove row striping in Markdown documents
    cell_fill(color = "white", alpha = NULL)
      ),
      #do this for all columns and rows
    locations = cells_body(
      columns = everything(),
      rows = everything()
    ))|>
  tab_header(
    title="Significant differences in means by gender")
```

References