ABOUT

This notebook illustrates how to run mediation analyses in R. It is used as a companion for a lecture in the CS-411 course "Digital Education". Patrick Jermann, CEDE, EPFL

MEDIATION

- explains how or why an intervention works
- mediator explains all or part of the treatment's impact on an intended outcome
- is an intermediate outcome that is measured or observed after the onset of the intervention. E.g. fidelity of application, how many questions were asked ?

MODERATION

- explains who the intervention benefits or what conditions must exist for the intervention to be effective.
- a factor that reflects who is most affected by the treatment
- a factor that exists prior to the introduction of an intervention

Eg. student characteristics, such as special education status, gender, ...

REFERENCES

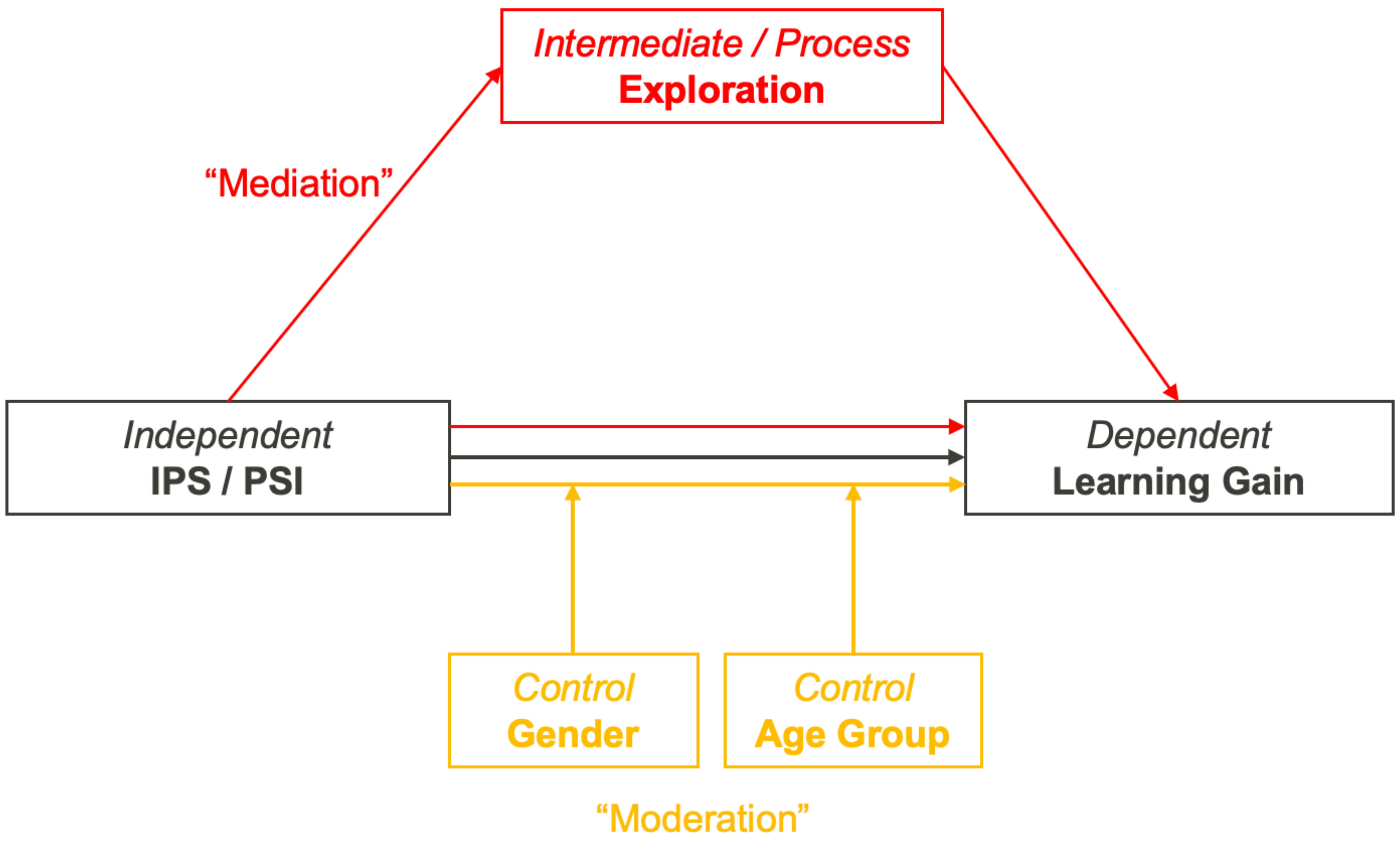
Seltman, H. J. (2012). Experimental design and analysis.

http://www.stat.cmu.edu/~hseltman/309/Book/

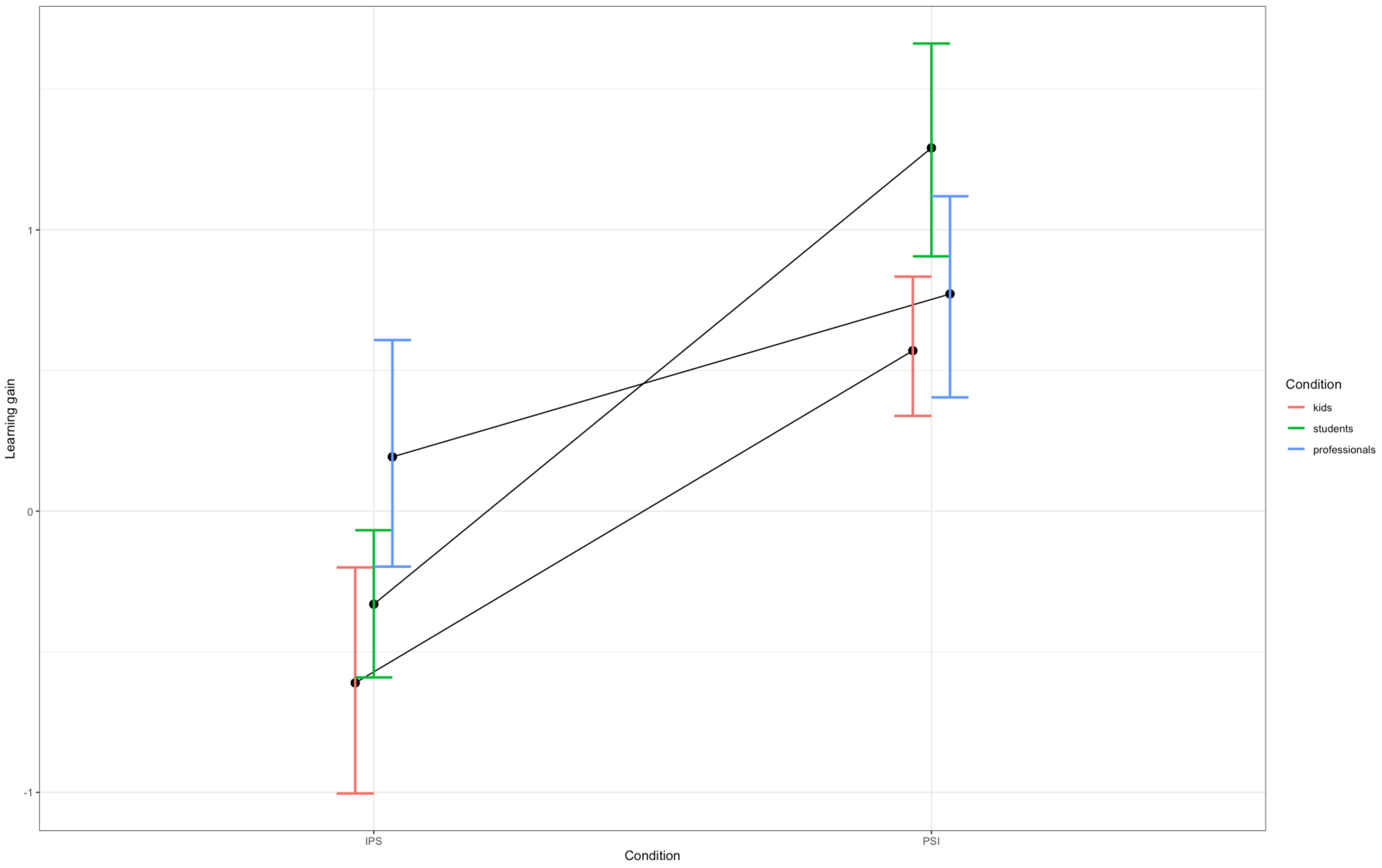
- t-test: chapter 6
- ANOVA: chapter 7
- Regression: chapter 10
- Chi-square: chapter 16

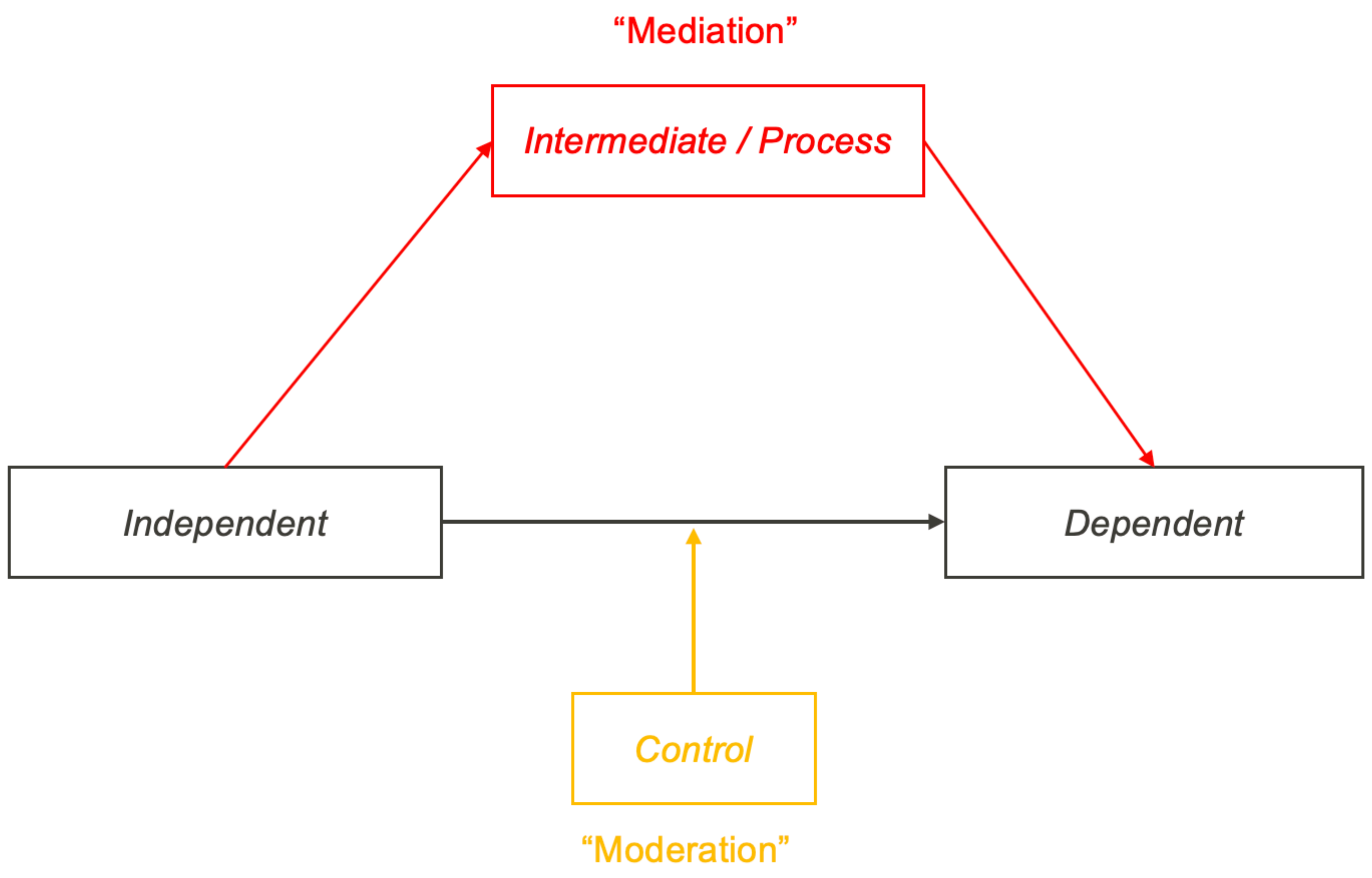
Jose, P. E. (2013). Doing statistical mediation and moderation. Guilford Press.
https://books.google.ch/books?id=aJFcO81Ro-0C&printsec=copyright&redir_esc=y#v=onepage&q&f=false

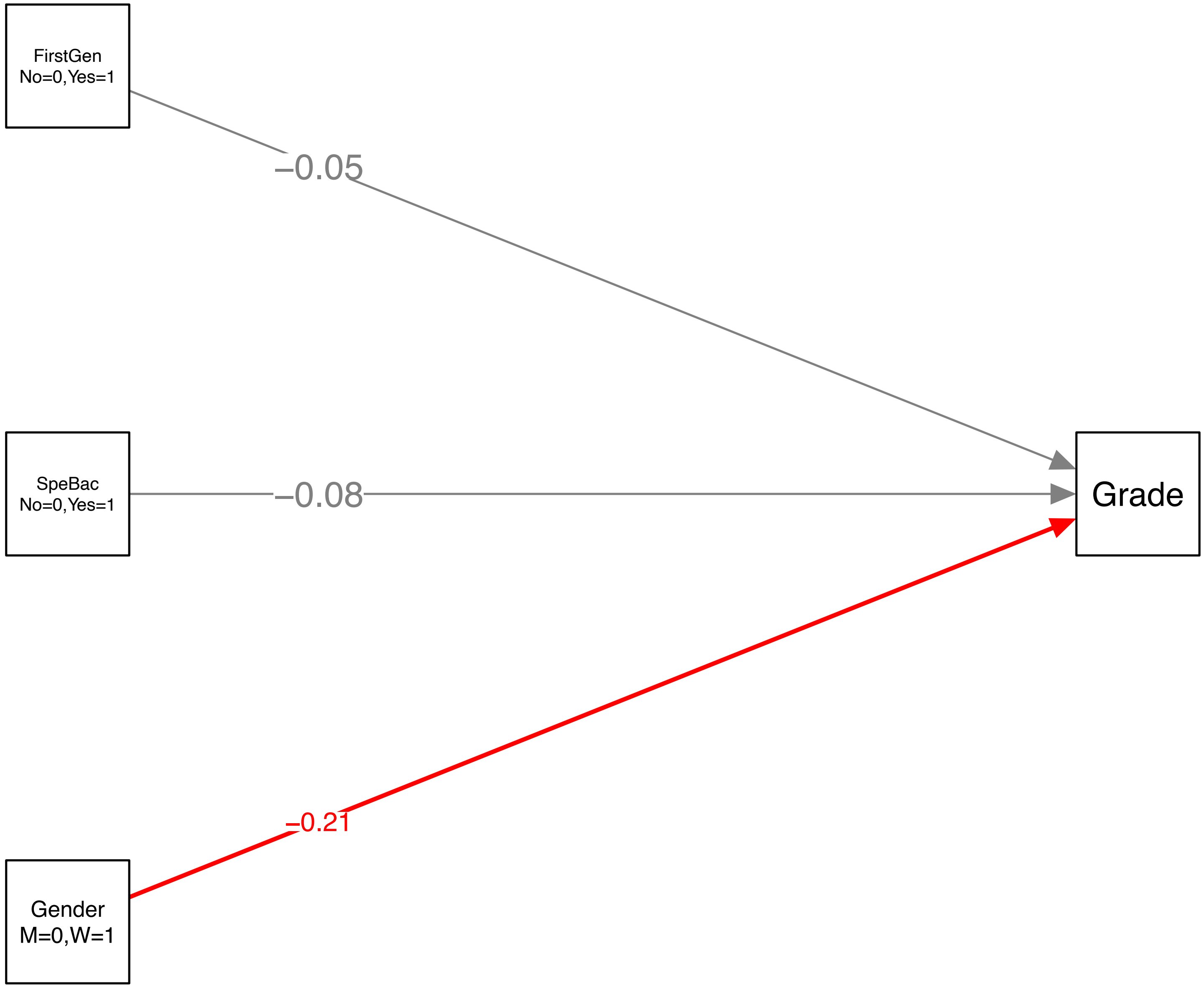
- Basic Mediation: chapter 3
- Basic Moderation: chapter 5

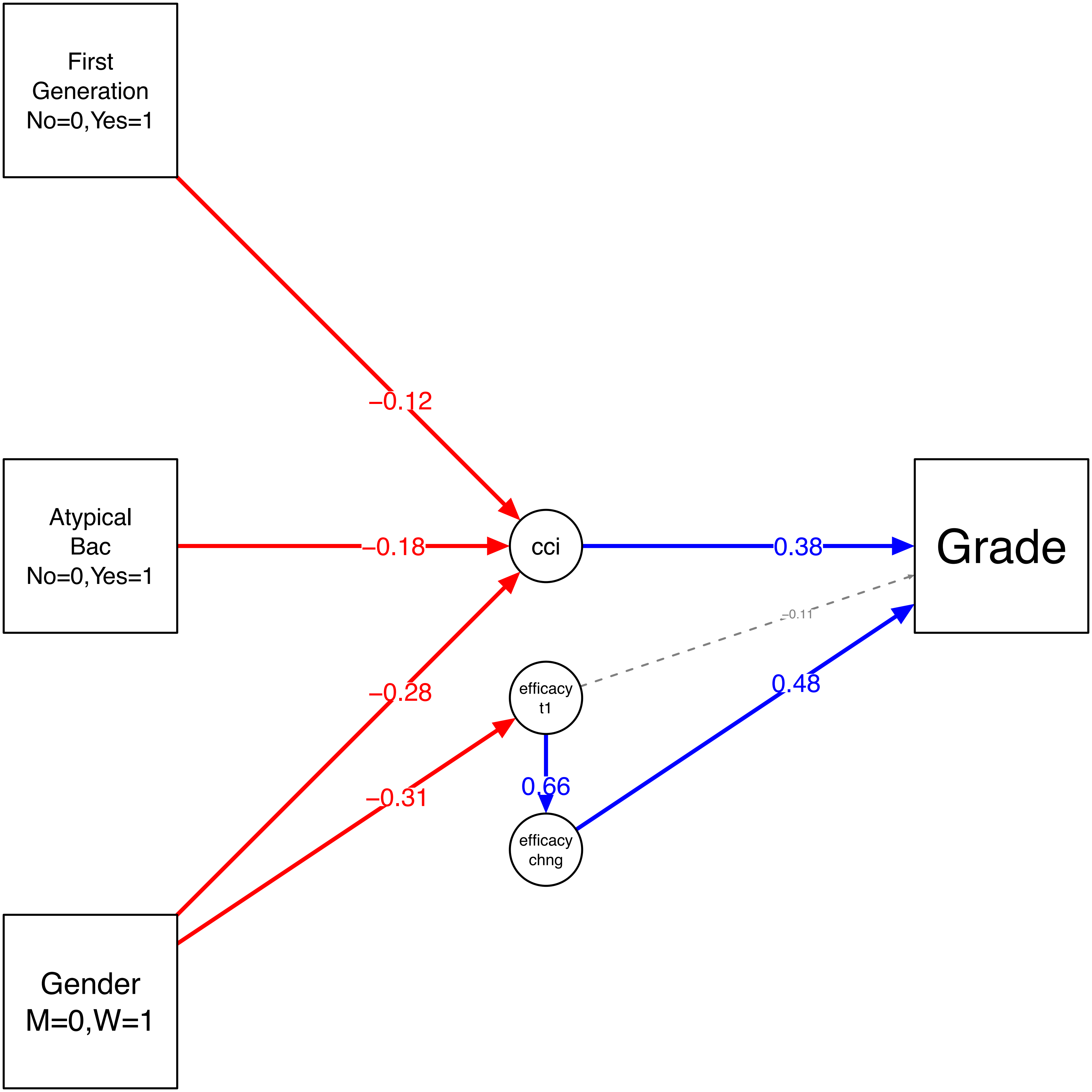

Intermediate / Process
Exploration
"Mediation"
Independent
IPS / PSI
Dependent
Learning Gain
Control
Gender
Control
Age Group
"Moderation"

```
Warning message:
"Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead."
```




“Mediation”
Intermediate / Process
Independent
Dependent
Control
“Moderation”


FirstGen
No=0,Yes=1
SpeBac
No=0,Yes=1
Gender
M=0,W=1
Grade
−0.05
−0.08
−0.21

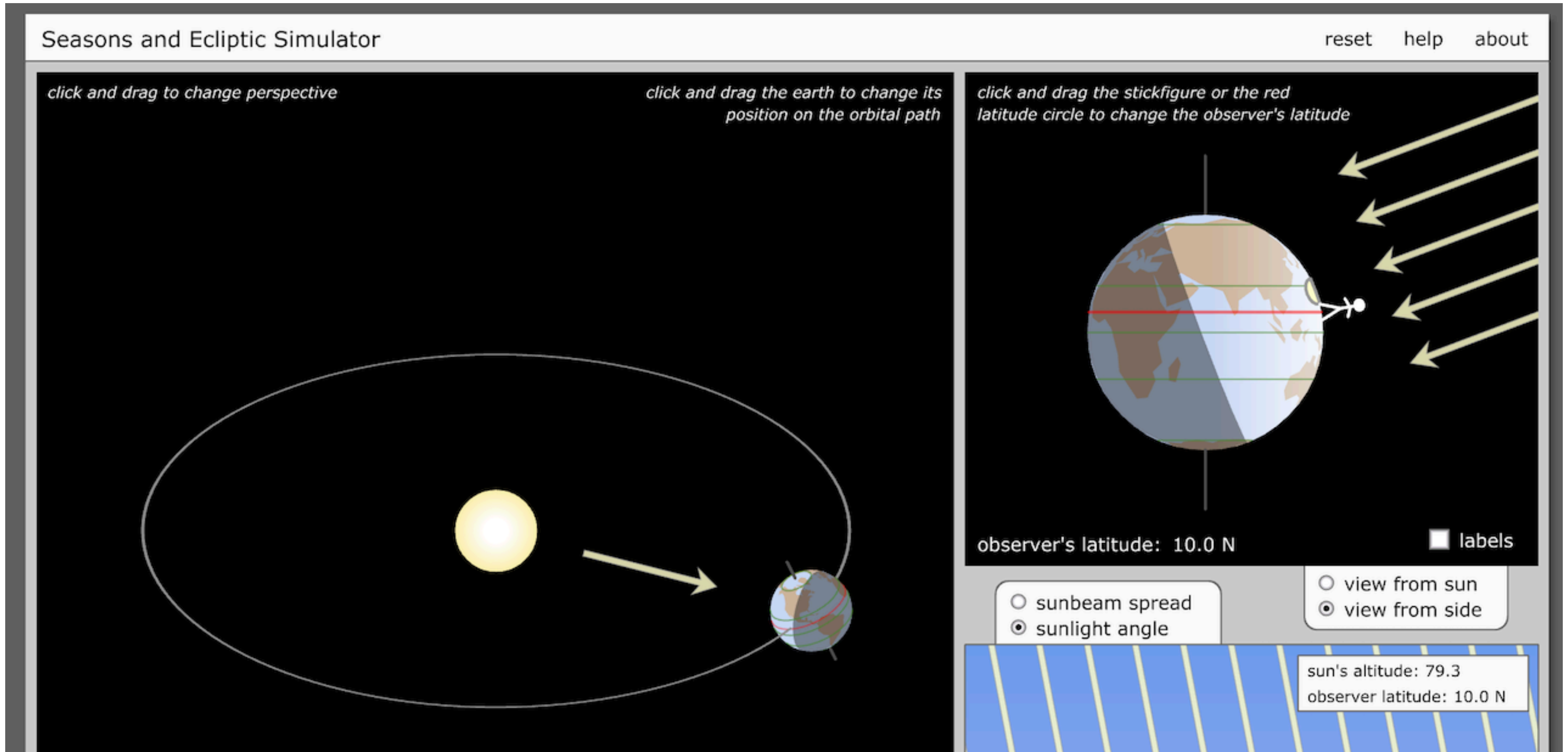


First Generation No=0,Yes=1
Atypical Bac No=0,Yes=1
Gender M=0,W=1
cci
efficacy t1
efficacy chng
Grade
−0.12
−0.18
−0.28
−0.31
0.38
0.48
0.66
−0.11

EXPERIMENT (IPS VS PSI)

In this **imaginary** experiment, we are studying the effect of the order of instruction and problem-solving (independent variable) on learning (dependent variable) how the position of the earth relative to the sun influences seasons.

Participants used a simulation (https://astro.unl.edu/classaction/animations/coordsmotion/eclipticsimulator.html) during the problem-solving phase and watched a video during the instruction phase.



PARTICIPANTS

The sample consisted of N=200 participants.

INDEPENDENT VARIABLE

Order of instruction The independent variable has two modalities (also called conditions):

- I-PS : instruction followed by problem-solving
- PS-I : problem-solving followed by instruction

Participants were *randomly* assigned to one of the experimental conditions.

DEPENDENT VARIABLE

Learning gain. Participants completed a 10 question *pre-test* before starting the experiment. The pre-test was a series of questions about their understanding of the sun-earth relative positions. After the experiment, particpants completed a 10 question *post-test* with similar questions as the pre-test. The learning gain was computed as :

$$learning.gain = post.test - pre.test$$

another possibilty would be the relative learning gain

$$rel.gain = \frac{post.test - pre.test}{max - pre.test}$$

CONTROL VARIABLES

Age group. Participants were recruited among highschool students who are interested in following studies at EPFL (kids), students doing their bachelor as well as alumni who are active professionally (professionals).

Young learners (e.g., second to fifth graders) may have insufficient prior knowledge about cognitive and metacognitive learning strategies to generate multiple solutions during initial problem solving

Gender. Experimenters also asked for the gender of the participants, either Male (M) or Female (M).

Self-regulation skills. Participants also filled in a questionnaire about their self.regulation skills by using the Learning Companion (https://companion.epfl.ch)

INTERMEDIATE / PROCESS VARIABLES

Solutions. The simulation system logged every simulation run and counted how often students used the simulation to generate a potential solution.

DATASET

This dataset was generated to illustrate basic statistical techniques like ANOVA and regression as well lightly more advanced techniques like mediation and moderation. However, we tried as much as possible to implement variations compatible with insights found in the literature about Productive Failure:

Sinha, T., & Kapur, M. (2021). When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. Review of Educational Research, 91(5), 761–798. https://doi.org/10.3102/00346543211019105

ANALYSIS

LOADING DATA

In [126]:

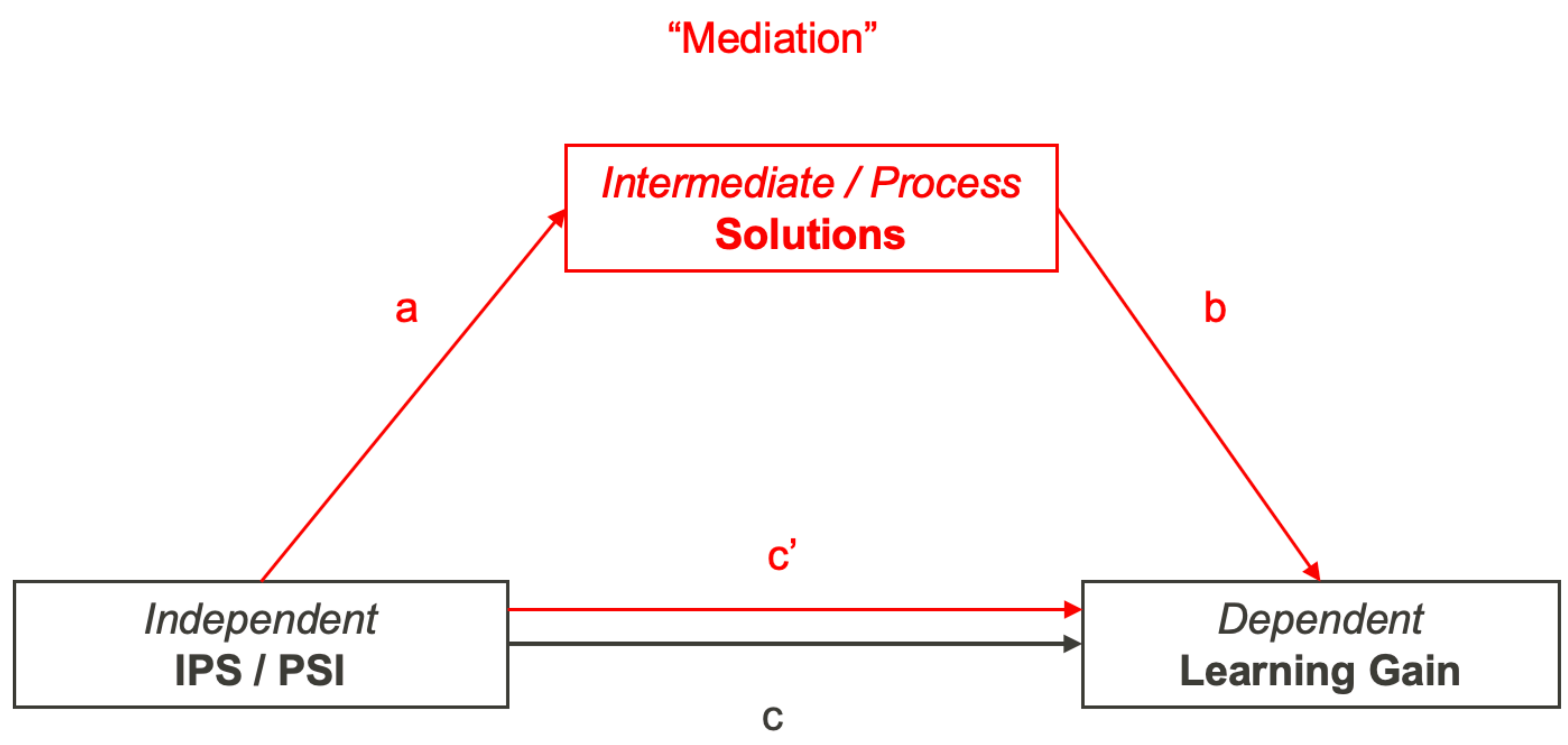
```
library(tidyverse) # Give ggplot, read_delim, tidyr, etc.

df <- suppressMessages(read_delim(file = "dataset.csv", delim = ",") %>%
  mutate(
    condition = factor(condition, labels = c("IPS", "PSI")),
    gender = factor(gender, labels = c("M", "F")),
    age.group = factor(age.group,
      labels = c("kids", "students", "professionals")
    )
  ))

head(df)
```

A tibble: 6 x 6

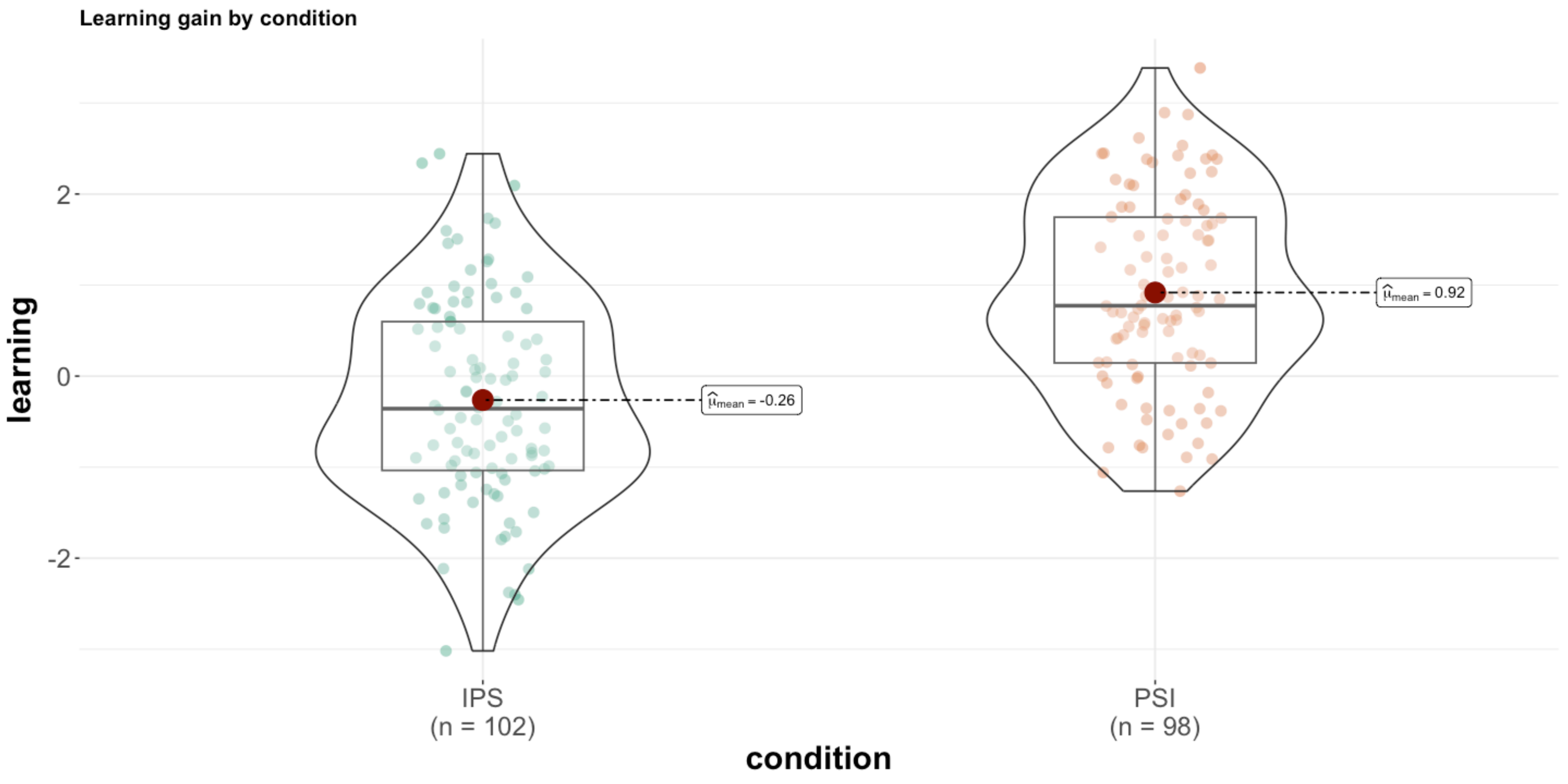
condition	gender	age.group	solutions	self.regulation	learning
<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>
PSI	F	kids	20	6.4129659	1.1447316
PSI	F	students	20	5.4942910	2.3840504
PSI	F	professionals	24	10.1754505	-0.3823993
PSI	M	kids	12	7.6805230	0.2294454
PSI	M	kids	9	0.4995889	0.4921680
IPS	M	kids	5	6.8204271	-1.3863676

MEDIATION

- explains how or why an intervention works
- mediator explains all or part of the treatment's impact on an intended outcome
- is an intermediate outcome that is measured or observed **after** the onset of the intervention. E.g. fidelity of application, how many questions were asked ?
- there is a plausible causality relation between the experimental treatment and the mediating variable.


"Mediation"
Intermediate / Process
Solutions
a
b
c'
Independent
IPS / PSI
Dependent
Learning Gain
c

CONDITION AFFECTS THE LEARNING GAIN (C PATH)



STEP 1

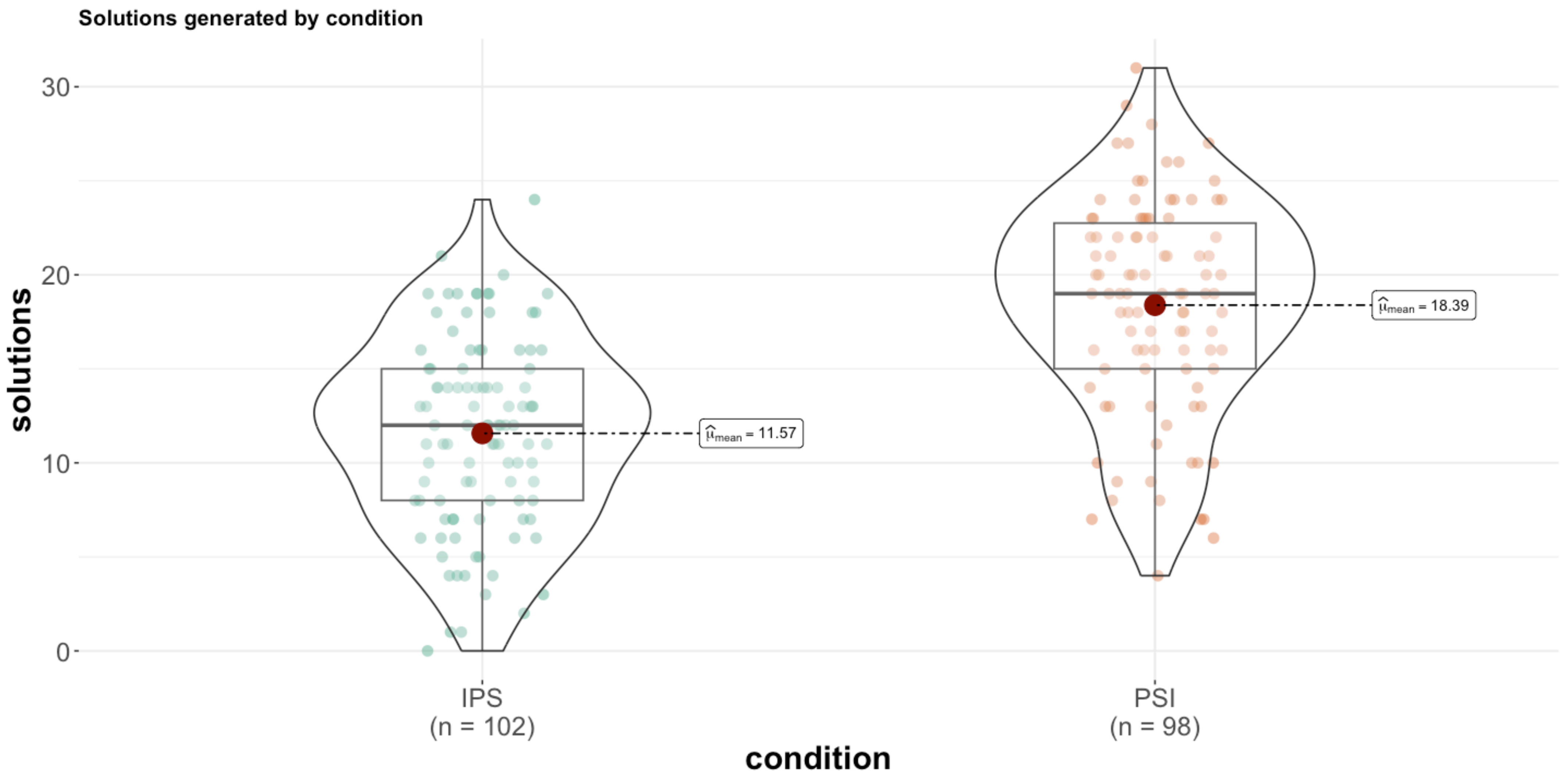
First we make sure the experimental treatment affects the dependent variable.

In [131]:

```
c.path <- lm(learning ~ condition, data=df)
summary(c.path)
```

```
Call:
lm(formula = learning ~ condition, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7555 -0.7754 -0.1243  0.8605  2.7054 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -0.2626     0.1092  -2.405   0.0171 *  
conditionPSI   1.1806     0.1560   7.567 1.41e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.103 on 198 degrees of freedom
Multiple R-squared:  0.2243,	Adjusted R-squared:  0.2204 
F-statistic: 57.26 on 1 and 198 DF,  p-value: 1.412e-12
```

CONDITION AFFECTS THE NUMBER OF SOLUTIONS GENERATED (A PATH)



STEP 2

This model checks whether the experimental treatment affects the mediating variable

In [132]:

```
a.path <- lm(solutions ~ condition, data=df)
summary(a.path)
```

```
Call:
lm(formula = solutions ~ condition, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-14.3878  -3.5686   0.5218   3.6122  12.6122 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   11.5686     0.5344  21.647  < 2e-16 ***
conditionPSI   6.8191     0.7635   8.932  2.9e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.397 on 198 degrees of freedom
Multiple R-squared:  0.2872,	Adjusted R-squared:  0.2836 
F-statistic: 79.78 on 1 and 198 DF,  p-value: 2.903e-16
```

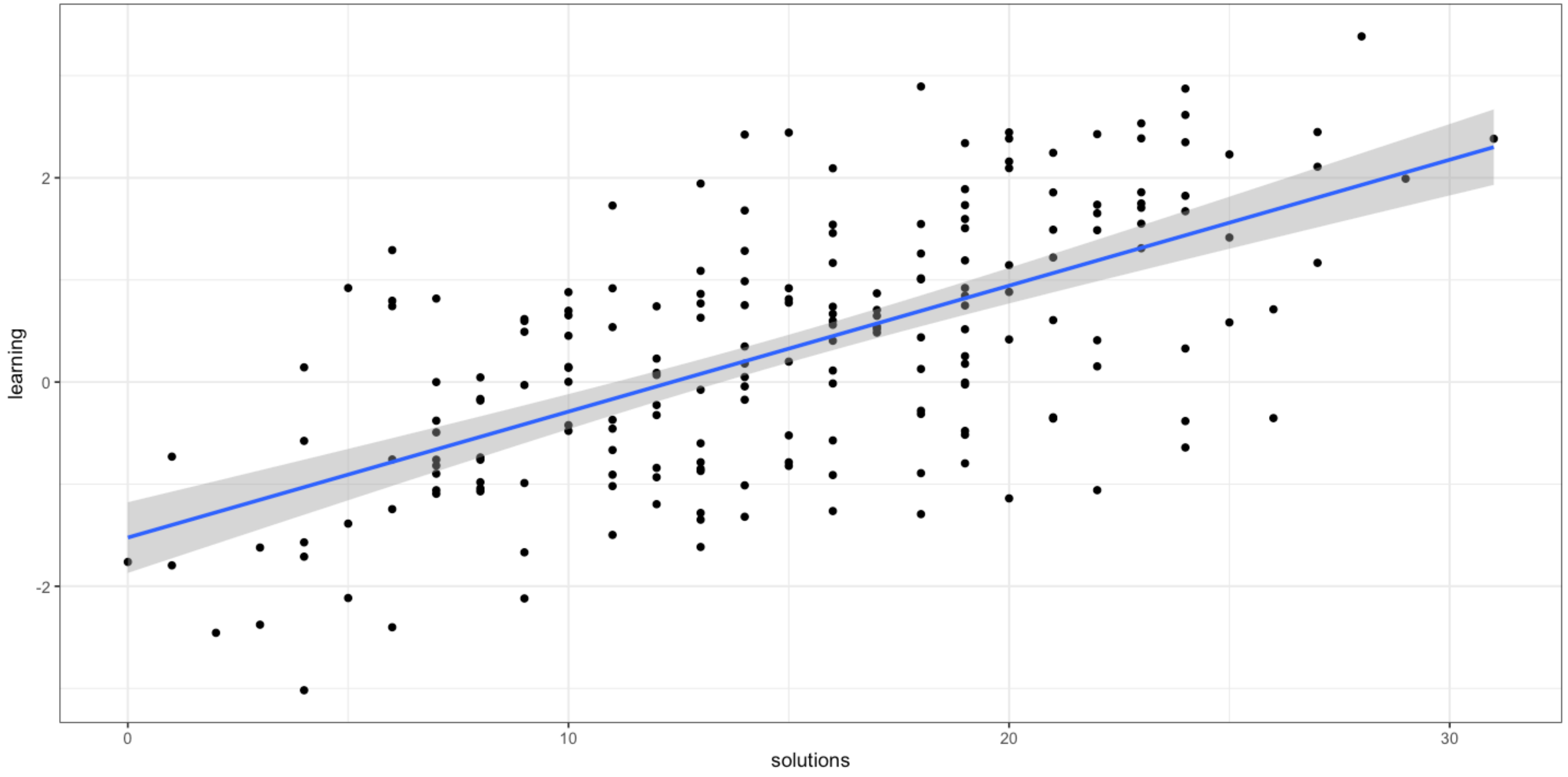
SOLUTIONS AND LEARNING GAIN ARE CORRELATED (B PATH)

In [129]:

```
cor(df$learning, df$solutions)
```

0.629564052503678

```
`geom_smooth()` using formula = 'y ~ x'
```



STEP 3 & 4

Finally we check whether a) the mediating variable affects the dependent variable and b) that the effect of the independent variable **decreases** (partial mediation) or even **dissapears** (full mediation).

In [133]:

```
c.dash.path <- lm(learning  ~ condition + solutions, data=df)
summary(c.dash.path)
```

```
Call:
lm(formula = learning ~ condition + solutions, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.34975 -0.64637 -0.00054  0.65168  2.35102 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.45730    0.17336  -8.406 8.41e-15 ***
conditionPSI  0.47639    0.15987   2.980  0.00325 ** 
solutions     0.10327    0.01256   8.219 2.70e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9542 on 197 degrees of freedom
Multiple R-squared:  0.4224,	Adjusted R-squared:  0.4165 
F-statistic: 72.03 on 2 and 197 DF,  p-value: < 2.2e-16
```

THE "OLD" WAY

Baron and Kenny's (1986) steps for mediation analysis See https://davidakenny.net/cm/mediate.htm

Step 1: Show that the causal variable is correlated with the outcome. Use Y as the criterion variable in a regression equation and X as a predictor (estimate and test path c in the above figure). This step establishes that there is an effect that may be mediated.

Step 2: Show that the causal variable is correlated with the mediator. Use M as the criterion variable in the regression equation and X as a predictor (estimate and test path a). This step essentially involves treating the mediator as if it were an outcome variable.

Step 3: Show that the mediator affects the outcome variable. Use Y as the criterion variable in a regression equation and X and M as predictors (estimate and test path b). It is not sufficient just to correlate the mediator with the outcome because the mediator and the outcome may be correlated because they are both caused by the causal variable X. Thus, the causal variable must be controlled in establishing the effect of the mediator on the outcome.

Step 4: To establish that M completely mediates the X-Y relationship, the effect of X on Y controlling for M (path c') should be zero (see discussion below on significance testing). The effects in both Steps 3 and 4 are estimated in the same equation.

MEDIATION - SOBEL TEST.

Following the Baron and Kenny method, in our case: a) the Estimate for the mediator (solutions) is statistically significant and b) the Estimate for conditionPSI went down from 1.1806 to 0.4764, but is still significant.

Sobel has developped a method to test whether this mediation effect is significant. See http://www.quantpsy.org/sobel/sobel.htm

$$z - value = \frac{a * b}{\sqrt{b^2 * S_a^2 + a^2 + S_b^2}}$$

where :

- a is the unstandardised Coefficient of the independent variable
- S_a is the standard error of the independent variable
- b is the unstandardised Coefficient of the mediation variable
- S_b is the standard error of the mediation variable

Downsides, the distribution of a*b is only normal for large samples, therefore people use a bootstrap method to estimate the confidence interval of ab. If it comprises 0 it is not significant, else it is.

In [134]:

```
# a path
coef(summary(a.path))
# a is the unstandardised coefficient for the path from X to M
a = 6.819128
Sa = 0.7634712

# c_dash path
coef(summary(c.dash.path))
# b refers to the unstadardised coefficient for the path from M to Y
b=0.1032689
Sb=0.01256431

a*b
z <- (a*b) / sqrt(b^2 * Sa^2 + b^2 * Sa^2)
z
```

A matrix: 2 x 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.568627	0.5344298	21.646673	4.337682e-54
conditionPSI	6.819128	0.7634712	8.931742	2.902753e-16

A matrix: 3 x 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.4573040	0.17336240	-8.406113	8.413584e-15
conditionPSI	0.4763879	0.15987428	2.979766	3.247729e-03
solutions	0.1032689	0.01256431	8.219223	2.700434e-14

0.7042038475192

6.31569553714542

DOING THE Z-TEST

We now determine whether this mediation effect is statistically significant (not due to sampling error) with a two-tailed z-test of the hypothesis that the mediated effect equals zero in the population.

H_0: the mediation effect is zero

H_1: the mediation effect is not zero

In [135]:

```
# The critical value (1.96 for a two tailed test with alpha = 0.05)
critical.value <- qnorm(0.025, mean = 0, sd = 1, lower.tail = FALSE)
critical.value
```

1.95996398454005

The z-value we obtained (6.33) is much higher than the critical value of 1.96.

We can compute the p-value associated with it. Since it is much lower than alpha (0.05) we can reject H_0 and conclude that the mediation effect is significant.

In [136]:

```
p.value <- pnorm(q = z, mean = 0, sd = 1, lower.tail = FALSE)
p.value
```

1.34474541849271e-10

THE MEDIATION PACKAGE

ESTIMATING THE PROPORTION OF MEDIATION

To alleviate the downsides of the z test (not normally distributed for small samples), we can use a bootstrapping method and simulate 1000 samples to estimage the confidence intervals for the indirect and direct effects.

In [137]:

```
library(mediation) # Gives mediate()

results <- mediate(a.path, c.dash.path,
    treat = "condition",
    mediator = "solutions",
    boot = FALSE, sims = 1000, # change to TRUE to get bootstrapped results
    control.value = "IPS",
    treat.value = "PSI"
)
```

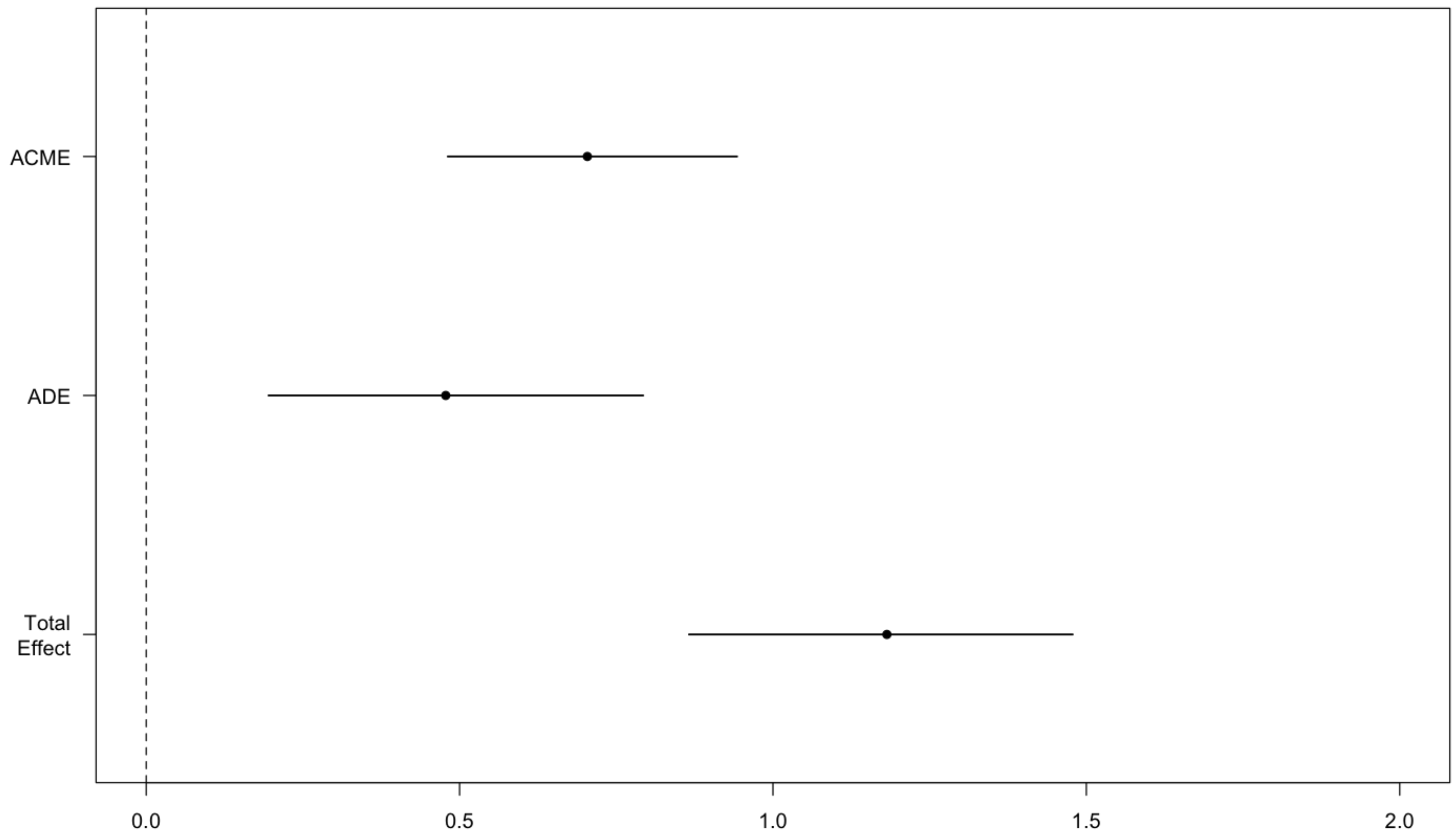
The direct effect (c path) is listed as ADE (average direct effect), the mediation path (ab path) is listed as ACME (average causal mediation effects, ACME). The total effect is ACE + ACME.

In [138]:
```
summary(results)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value
ACME              0.704        0.481         0.94  <2e-16 ***
ADE               0.478        0.195         0.79   0.002 **
Total Effect      1.182        0.866         1.48  <2e-16 ***
Prop. Mediated    0.599        0.403         0.81  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

In [139]:

```
options(repr.plot.width=12, repr.plot.height=8)
plot(results, xlim=c(0,2))
```



MODERATED MEDIATION

Remember we found out in the previous ANOVA analyses that the age group was a moderator for the effect of the experimental treatment.

We now look into whether the mediation is also moderated by this variable.

MODEL C

Predicting the dependent variable with the condition

NB: this was our model.2 when doing 2-factor ANOVA earlier. In this context we use the default contrasts ("treatment") when building the lm model.

In [140]:

```
model.c <- lm(learning ~ condition + age.group +
                        condition:age.group, data = df)
coef(summary(model.c))
```

A matrix: 6 x 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.6108569	0.1802479	-3.388983	8.496561e-04
conditionPSI	1.1814799	0.2731392	4.325559	2.432415e-05
age.groupstudents	0.2800331	0.2549090	1.098561	2.733210e-01
age.groupprofessionals	0.8037029	0.2608150	3.081506	2.359029e-03
conditionPSI:age.groupstudents	0.4406413	0.3701604	1.190406	2.353413e-01
conditionPSI:age.groupprofessionals	-0.6020656	0.3802612	-1.583295	1.149831e-01

MODEL A

Predicting the mediator variable with the condition

In [141]:

```
model.a <- lm(solutions ~ condition + age.group +
                          condition:age.group,
                          data=df)
coef(summary(model.a))
```

A matrix: 6 x 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.428571	0.6949723	12.127925	1.466060e-25
conditionPSI	3.164021	1.0531287	3.004401	3.012129e-03
age.groupstudents	2.971429	0.9828392	3.023311	2.838132e-03
age.groupprofessionals	6.758929	1.0056107	6.721218	1.957578e-10
conditionPSI:age.groupstudents	6.593874	1.4272079	4.620121	6.978416e-06
conditionPSI:age.groupprofessionals	2.406055	1.4661531	1.641066	1.024039e-01

MODEL C.DASH

Same as model c but we add the mediator (solutions).

In [142]:
```
model.c.dash <- lm(learning ~ condition +
                            age.group + # moderator
                            solutions + # mediation
                            solutions:age.group + # moderation of the mediator
                            condition:age.group, # moderation of the contidion
                data=df)
coef(summary(model.c.dash))
```

A matrix: 9 x 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.754133256	0.28595013	-6.1344028	4.821609e-09
conditionPSI	0.752302689	0.25588710	2.9399790	3.687580e-03
age.groupstudents	-0.215817874	0.45784024	-0.4713825	6.379056e-01
age.groupprofessionals	0.469272526	0.54511758	0.8608648	3.903919e-01
solutions	0.135642952	0.02825983	4.7998509	3.195754e-06
age.groupstudents:solutions	0.008140145	0.03987183	0.2041578	8.384476e-01
age.groupprofessionals:solutions	-0.038345393	0.04017082	-0.9545583	3.410076e-01
conditionPSI:age.groupstudents	-0.533201889	0.43462399	-1.2268119	2.214034e-01
conditionPSI:age.groupprofessionals	-0.714843156	0.38039008	-1.8792371	6.173514e-02

WITH MEDIATE() PACKAGE

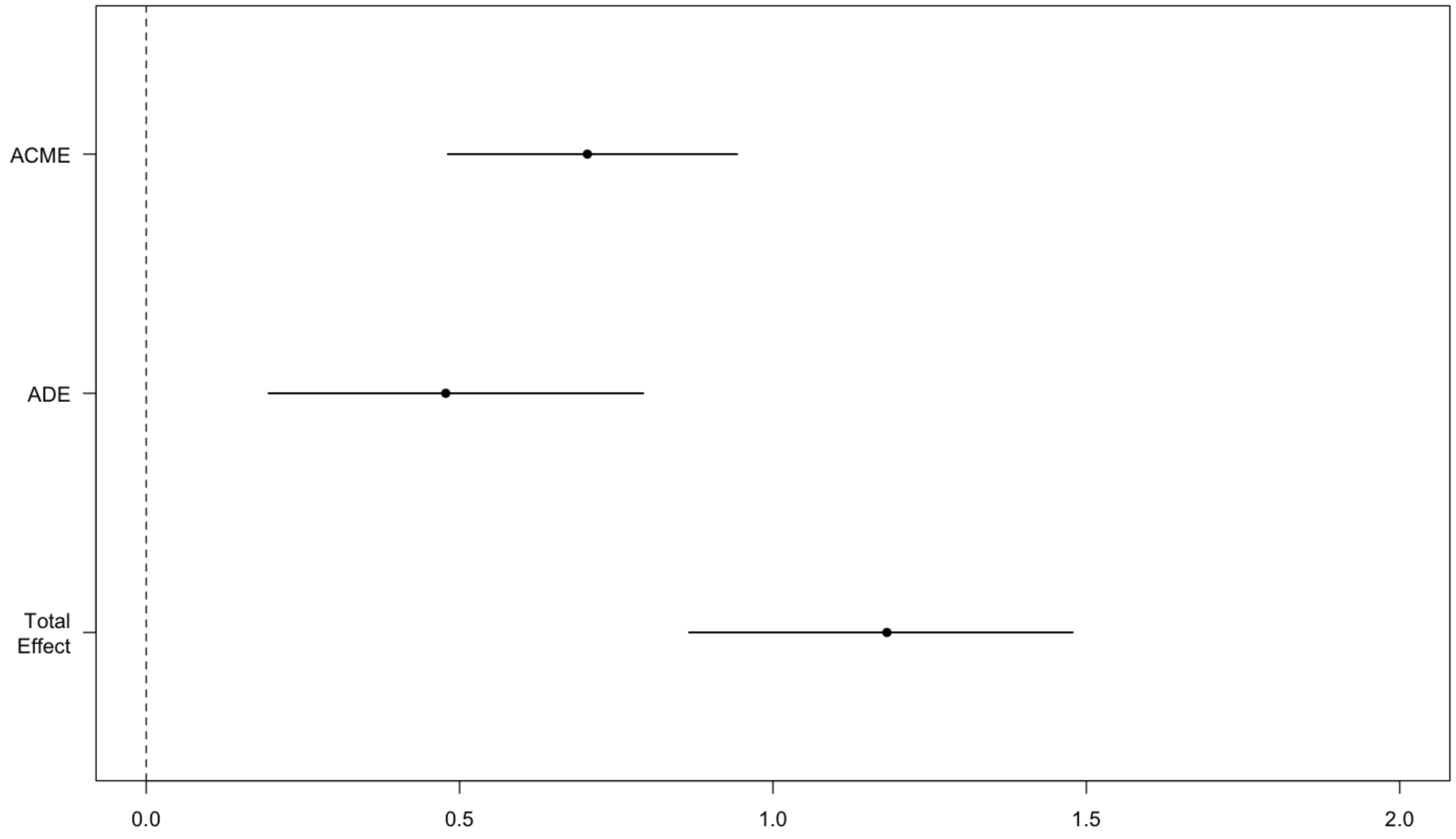
In [143]:

```
results <- mediate(model.a, # predicts mediator with condition
                   model.c.dash, # predicts learning with condition and solutions
                   treat='condition',
                   mediator='solutions',
                   boot = FALSE, sims = 1000, # change to TRUE to get bootstrapped results
                   control.value = "IPS",
                   treat.value = "PSI")
```

The results indicate a **partial** mediation (ACME **and** ADE are significant).

In [144]:
```
summary(results)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME             0.8240       0.5618         1.10  <2e-16 ***
ADE              0.3210      -0.0151         0.66    0.07 .  
Total Effect     1.1450       0.8512         1.44  <2e-16 ***
Prop. Mediated   0.7207       0.4916         1.02  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```



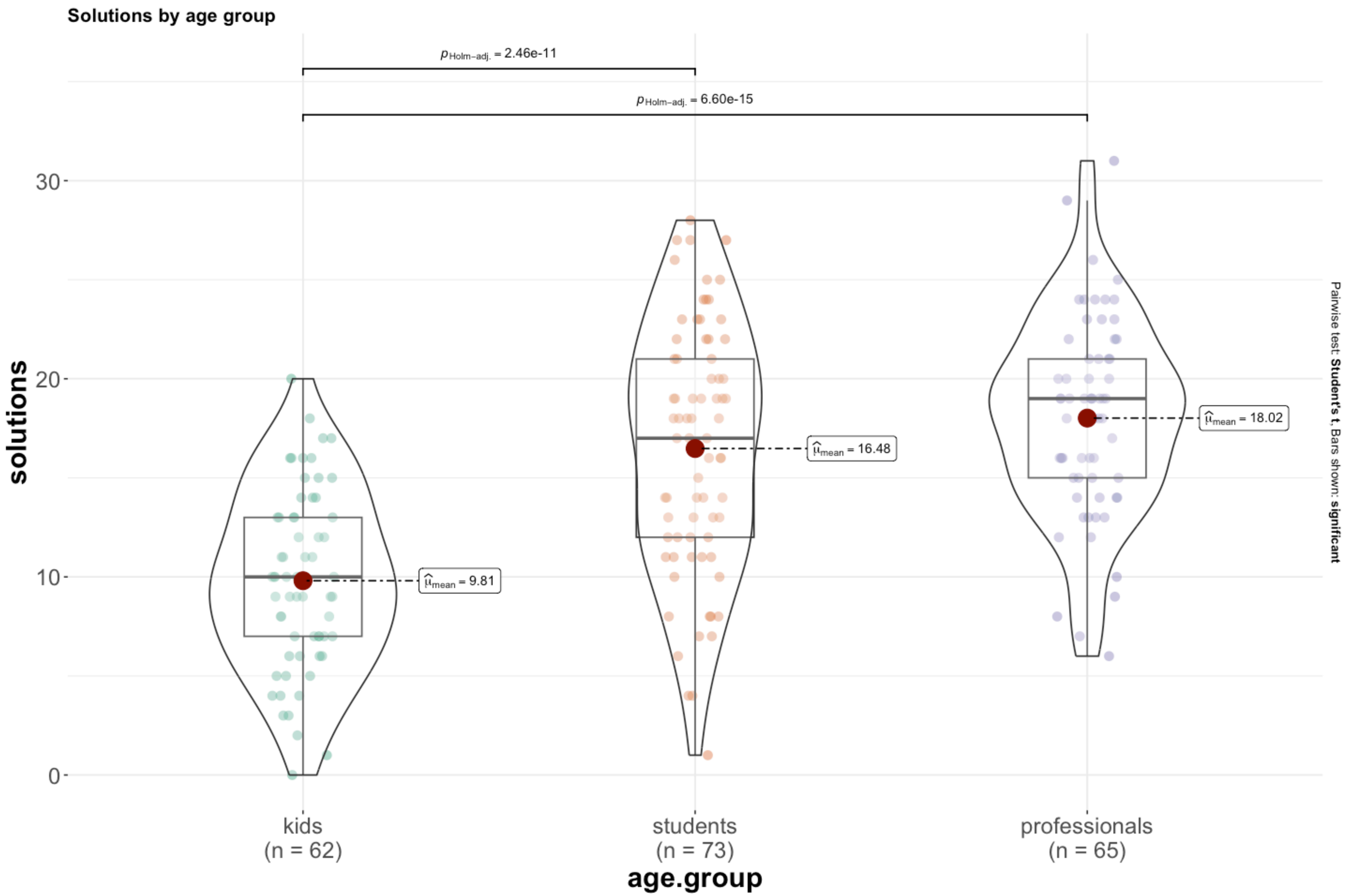
ACME
ADE
Total Effect
0.0
0.5
1.0
1.5
2.0

MODERATED MEDIATION WITH AGE GROUP

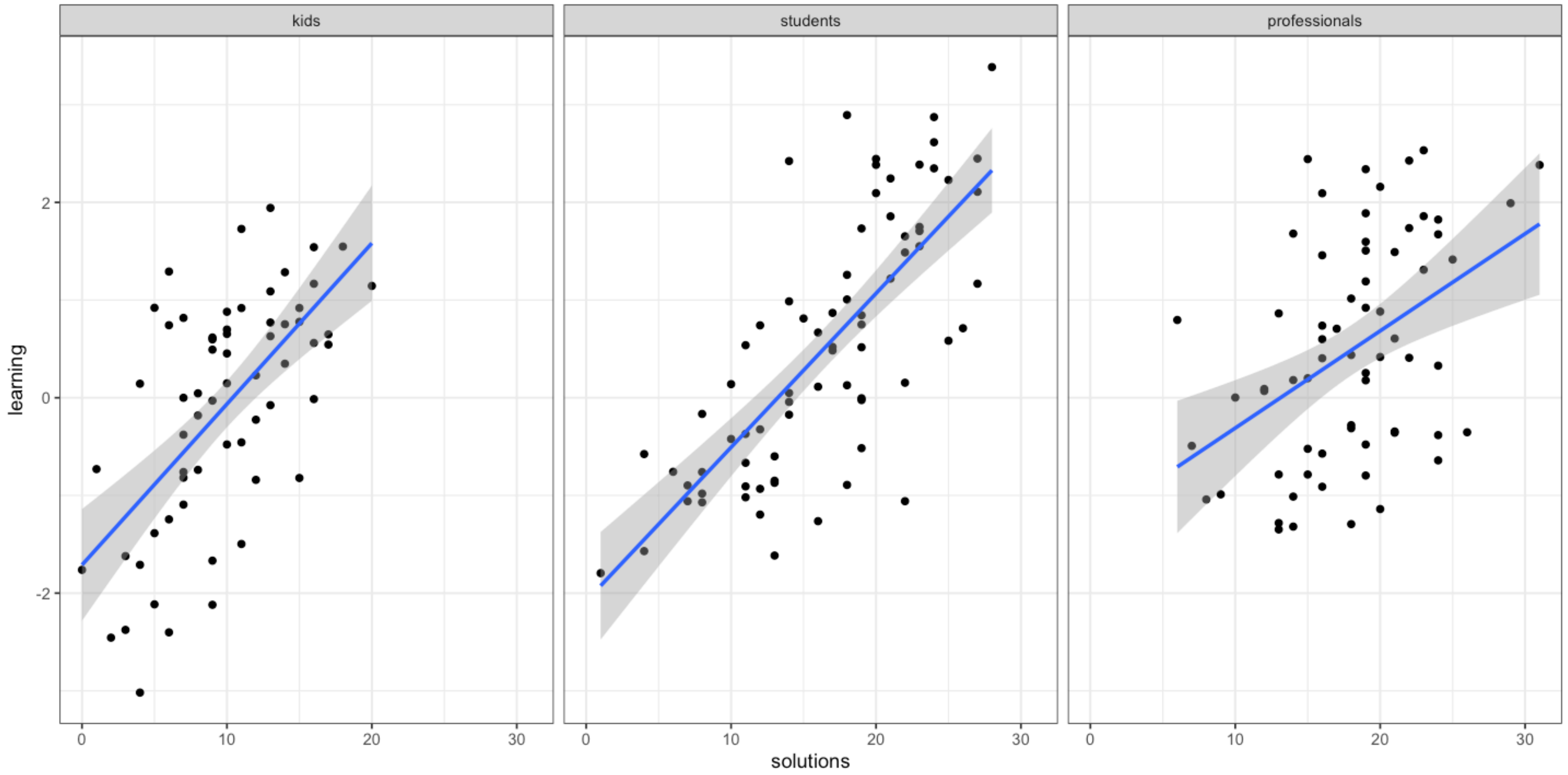
- It appears that kids have created less solutions than students and professionals.
- It appears that the relation between solutions and learning is different for different age groups.

Does the mediation exist for all age groups ?

KIDS HAVE CREATED LESS SOLUTIONS THAN STUDENTS AND PROFESSIONALS



THE RELATION BETWEEN SOLUTIONS AND LEARNING IS DIFFERENT FOR DIFFERENT AGE GROUPS



ADDING COVARIATES TO THE MEDIATE FUNCTION

Approach: We compute the mediation for each age subgroup. This is done by adding a covariates argument to the mediate function.

In [148]:

```
results.kids <- mediate(model.a, model.c.dash, treat='condition', mediator='solutions',
                        boot=FALSE, sims=1000, control.value = "IPS", treat.value = "PSI",
                        covariates = list(age.group="kids")) # change boot to TRUE to get bootstrapped results
summary(results.kids)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

               Estimate 95% CI Lower 95% CI Upper p-value
ACME              0.427        0.130         0.76   0.004 **
ADE               0.762        0.290         1.25  <2e-16 ***
Total Effect      1.188        0.630         1.74  <2e-16 ***
Prop. Mediated    0.361        0.121         0.64   0.004 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

In [149]:

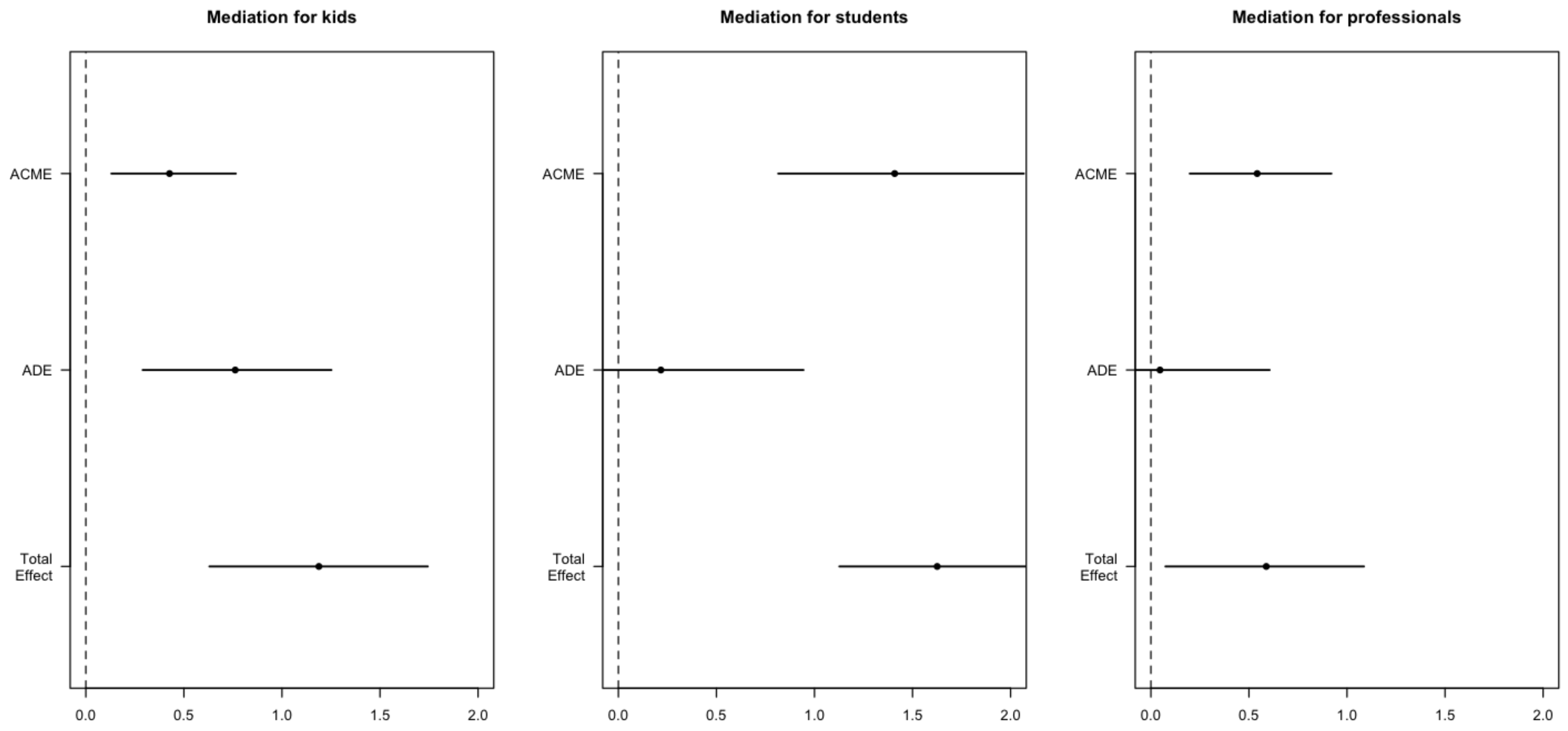
```
results.students <-mediate(model.a, model.c.dash, treat='condition', mediator='solutions',
                           boot=FALSE,sims=1000, control.value = "IPS", treat.value = "PSI",
                           covariates = list(age.group="students")) # change boot to TRUE to get bootstrapped results
summary(results.students)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

               Estimate 95% CI Lower 95% CI Upper p-value
ACME              1.409        0.815         2.07  <2e-16 ***
ADE               0.217       -0.434         0.94    0.54
Total Effect      1.626        1.127         2.14  <2e-16 ***
Prop. Mediated    0.867        0.493         1.34  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

In [150]:

```r
results.professionals <- mediate(model.a, model.c.dash, treat='condition', mediator='solutions',
                           boot=FALSE,sims=1000, control.value = "IPS", treat.value = "PSI",
                           covariates = list(age.group="professionals")) # change boot to TRUE to get bootstrapped results
summary(results.professionals)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME             0.5420       0.1984         0.92  <2e-16 ***
ADE              0.0463      -0.5190         0.61    0.88    
Total Effect     0.5883       0.0742         1.09    0.03 *  
Prop. Mediated   0.8954       0.2388         3.99    0.03 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200 


Simulations: 1000 
```



Mediation for kids
Mediation for students
Mediation for professionals
ACME
ADE
Total Effect
0.0
0.5
1.0
1.5
2.0

PUSHING THE ENVELOPE ...

Testing whether the direct and indirect effects are different across groups ?

In [152]:

```
test.modmed(results,
            covariates.1 = list(age.group = "kids"),
            covariates.2 = list(age.group = "students"), sims = 100)
```

```
        Test of ACME(covariates.1) - ACME(covariates.2) = 0

data:  estimates from results
ACME(covariates.1) - ACME(covariates.2) = -0.97806, p-value = 0.02
alternative hypothesis: true ACME(covariates.1) - ACME(covariates.2) is not equ
95 percent confidence interval:
 -1.9220014 -0.2656049


        Test of ADE(covariates.1) - ADE(covariates.2) = 0

data:  estimates from results
ADE(covariates.1) - ADE(covariates.2) = 0.52652, p-value = 0.32
alternative hypothesis: true ADE(covariates.1) - ADE(covariates.2) is not equal
95 percent confidence interval:
 -0.4760505  1.3969293
```

In [153]:

```
test.modmed(results,
            covariates.1 = list(age.group = "kids"),
            covariates.2 = list(age.group = "professionals"), sims = 1000)
```

```
	Test of ACME(covariates.1) - ACME(covariates.2) = 0

data:  estimates from results
ACME(covariates.1) - ACME(covariates.2) = -0.10127, p-value = 0.686
alternative hypothesis: true ACME(covariates.1) - ACME(covariates.2) is not equ
95 percent confidence interval:
 -0.6136053  0.3795400


	Test of ADE(covariates.1) - ADE(covariates.2) = 0

data:  estimates from results
ADE(covariates.1) - ADE(covariates.2) = 0.71843, p-value = 0.056
alternative hypothesis: true ADE(covariates.1) - ADE(covariates.2) is not equal
95 percent confidence interval:
 -0.003499324  1.473899382
```

In [154]:

```
test.modmed(results,
            covariates.1 = list(age.group = "students"),
            covariates.2 = list(age.group = "professionals"), sims = 1000)
```

```
        Test of ACME(covariates.1) - ACME(covariates.2) = 0

data:  estimates from results
ACME(covariates.1) - ACME(covariates.2) = 0.84938, p-value = 0.014
alternative hypothesis: true ACME(covariates.1) - ACME(covariates.2) is not equ
95 percent confidence interval:
 0.1762777 1.5470583


        Test of ADE(covariates.1) - ADE(covariates.2) = 0

data:  estimates from results
ADE(covariates.1) - ADE(covariates.2) = 0.20071, p-value = 0.664
alternative hypothesis: true ADE(covariates.1) - ADE(covariates.2) is not equal
95 percent confidence interval:
 -0.6614683  1.1035723
```

ROLE OF A PRE-TREATMENT MODERATOR ?

We now look at the potential influence of a pre-treatment confounder, the `self.regulation` skills of the participants.

We wonder whether the potential positive effect of testing many solutions is conditioned on the level of self-regulation skills. The rationale for this could be: when you run experiments (generate many `solutions`), you learn more if you can accurately revise your hypotheses (`self-regulation`) about the phenomenon. This would mean that the moderation effect that we discovered for the number of solutions would be present mainyl for subjects with a high level of self-regulation skills.

The analysis we conduct is similar to the one we did for `age.group` except that this time, the moderator is a continuous variable.

In [155]:
```
model.a2 <- lm(solutions ~ condition + condition:self.regulation + self.regulation, data=df)
model.c2.dash <- lm(learning ~ condition + self.regulation + condition:self.regulation + solutions + self.regulation, data=df)
```

In [156]:
```
results.self.regulation <- mediate(model.a2, model.c2.dash,
                                   treat='condition',
                                   mediator='solutions',
                                   boot=FALSE,sims=1000, # change boot to TRUE to get bootstrapped results
                                   control.value = "IPS", treat.value = "PSI")

summary(results.self.regulation)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME              0.484        0.300         0.70  <2e-16 ***
ADE               0.464        0.144         0.78   0.002 ** 
Total Effect      0.948        0.658         1.25  <2e-16 ***
Prop. Mediated    0.510        0.319         0.80  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

In [157]:

```r
results.hi.self.regulation <- mediate(model.a2, model.c2.dash,
    treat = "condition", mediator = "solutions",
    boot = FALSE, sims = 1000, control.value = "IPS", treat.value = "PSI",
    covariates = list(self.regulation = quantile(df$self.regulation, .75))
) # change boot to TRUE to get bootstrapped results

summary(results.hi.self.regulation)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

               Estimate 95% CI Lower 95% CI Upper p-value
ACME            0.46127      0.28399         0.69  <2e-16 ***
ADE             0.35471      0.00125         0.71    0.05 *
Total Effect    0.81598      0.44790         1.18  <2e-16 ***
Prop. Mediated  0.57137      0.32768         1.00  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

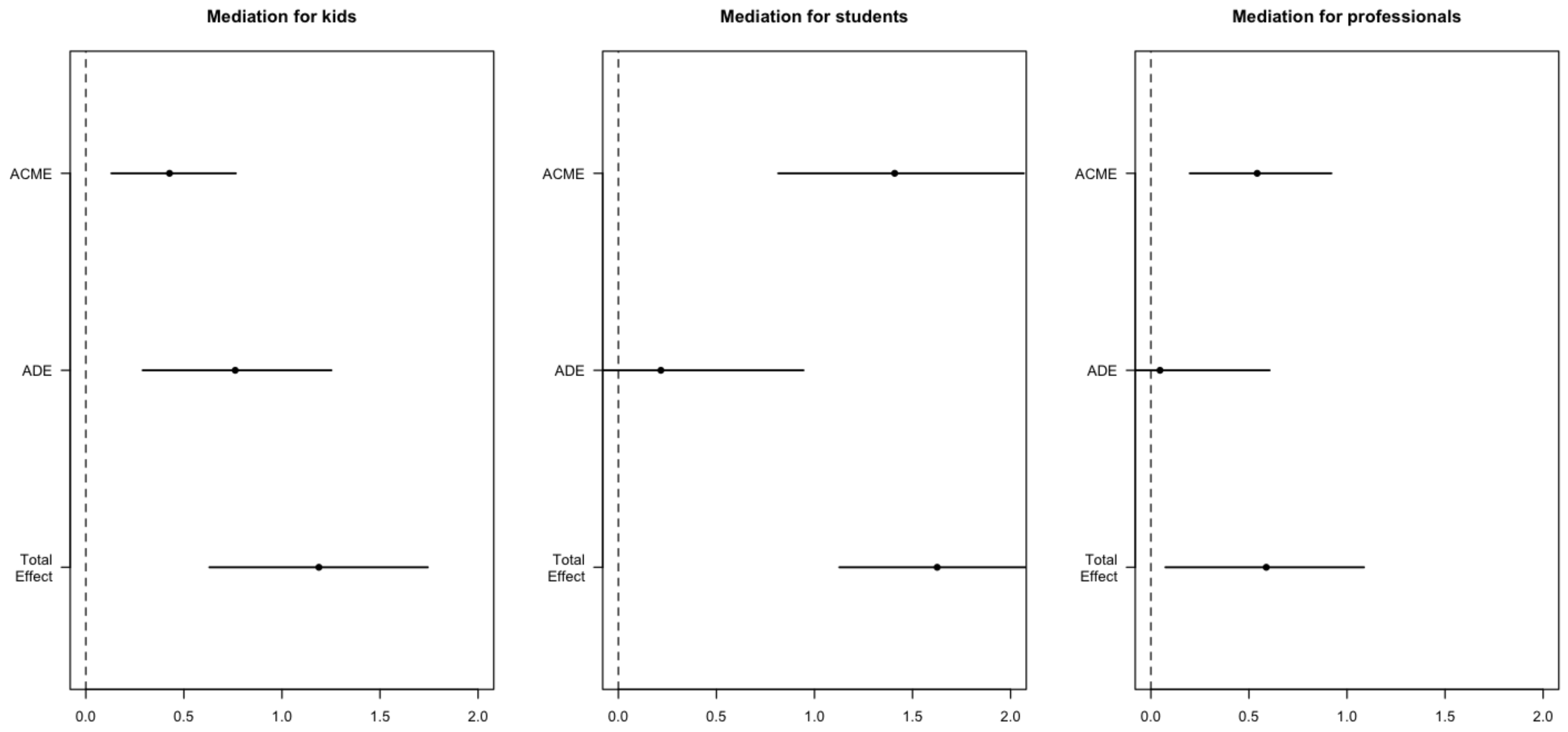
In [158]:

```r
results.lo.self.regulation <- mediate(model.a2, model.c2.dash,
    treat = "condition", mediator = "solutions",
    boot = FALSE, sims = 1000, control.value = "IPS", treat.value = "PSI",
    covariates = list(self.regulation = quantile(df$self.regulation, .25))
) # change boot to TRUE to get bootstrapped results

summary(results.lo.self.regulation)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME              0.485        0.288         0.70  <2e-16 ***
ADE               0.600        0.258         0.97   0.002 ** 
Total Effect      1.085        0.755         1.45  <2e-16 ***
Prop. Mediated    0.451        0.268         0.68  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 200


Simulations: 1000
```

In [159]:

```
par(mfrow=c(1,2))
plot(results.lo.self.regulation, main="Low self-regulation", xlim=c(0,2))
plot(results.hi.self.regulation, main="High self-regulation", xlim=c(0,2))
par(mfrow=c(1,1))
```



In [160]:

```
test.modmed(results.self.regulation,
            covariates.1 = list(self.regulation=quantile(df$self.regulation, .25)),
            covariates.2 = list(self.regulation=quantile(df$self.regulation, .75)), sims = 1000)
```

```
        Test of ACME(covariates.1) - ACME(covariates.2) = 0

data:  estimates from results.self.regulation
ACME(covariates.1) - ACME(covariates.2) = 0.016846, p-value = 0.94
alternative hypothesis: true ACME(covariates.1) - ACME(covariates.2) is not equ
95 percent confidence interval:
 -0.2744128  0.3020783


        Test of ADE(covariates.1) - ADE(covariates.2) = 0

data:  estimates from results.self.regulation
ADE(covariates.1) - ADE(covariates.2) = 0.26845, p-value = 0.296
alternative hypothesis: true ADE(covariates.1) - ADE(covariates.2) is not equal
95 percent confidence interval:
 -0.2528220  0.7586876
```