

CONTROLLED

EXPERIMENT 1 (E1) +
EXPERIMENT 2 (E2)

NEW INDEPENDENT VARIABLES

<i>NAME</i>	<i>VARIABLE</i>	<i>VALUES</i>
MIX	Mix of domain familiarities	0,1,2,3
CR	Average creativity score level	Low, Medium, High
REXP	Average RE experience	None, Low, Medium, High
IEXP	Average industrial experience	None, Low, Medium, High
IREXP	Average industrial RE experience	None, Low, Medium, High
NCS	Number of participants with CS background	0,1,2,3
NSE	Number of participants studying SE	0,1,2,3
NGRAD	Number of graduate student participants	0,1,2,3

NEW DEPENDENT VARIABLES

<i>NAME</i>	<i>VARIABLE</i>	<i>VALUES</i>
RAW	Raw number of ideas	Numeric
NRAW	Normalized RAW	Numeric
AVG_R	Average number of relevant ideas	Numeric
NR	Normalized AVG_R	Numeric
AVG_F	Average number of feasible ideas	Numeric
NF	Normalized AVG_F	Numeric
AVG_I	Average number of innovative ideas	Numeric
NI	Normalized AVG_I	Numeric

FACTOR ANALYSIS

NAME	VARIABLE	VALUES
MIX	Mix of domain familiarities	0,1,2,3
MIX	Mix of domain familiarities	0,1,2,3
CR	Average creativity score level	Low, Medium, High
CR	Average creativity score level	Low, Medium, High
REXP	Average RE experience	None, Low, Medium, High
REXP	Average RE experience	None, Low, Medium, High
IREXP	Average industrial RE experience	None, Low, Medium, High
IREXP	Average industrial RE experience	None, Low, Medium, High
EDU	Sum of NCS and NSE	Low, High
NCS	Number of participants with CS background	0,1,2,3
NSE	Number of participants studying SE	0,1,2,3
NGRAD	Number of graduate student participants	0,1,2,3
NGRAD	Number of graduate student participants	0,1,2,3

HYPOTHESES

H_{MIX} : The effectiveness of a team in requirements idea generation is affected by the team's **MIX**.

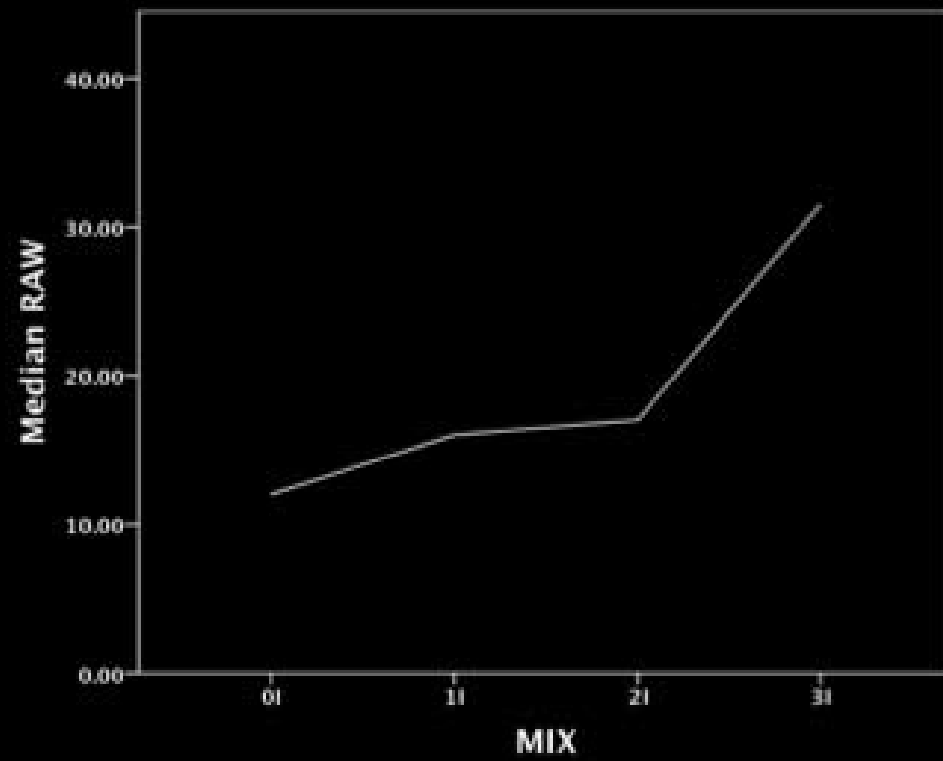
H_{CR} : The effectiveness of a team in requirements idea generation is affected by the team's **CR**.

H_{EXP} : The effectiveness of a team in requirements idea generation is affected by the team's **EXP**.

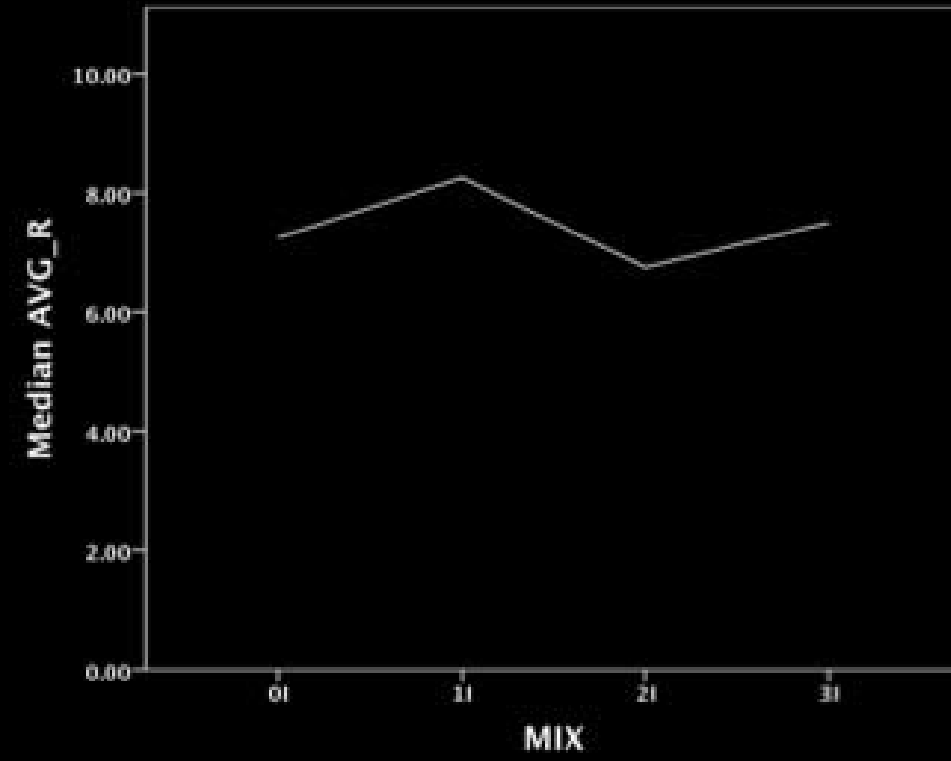
H_{EDU} : The effectiveness of a team in requirements idea generation is affected by the team's **EDU**.

H_{NGRAD} : The effectiveness of a team in requirements idea generation is affected by the team's **NGRAD**.

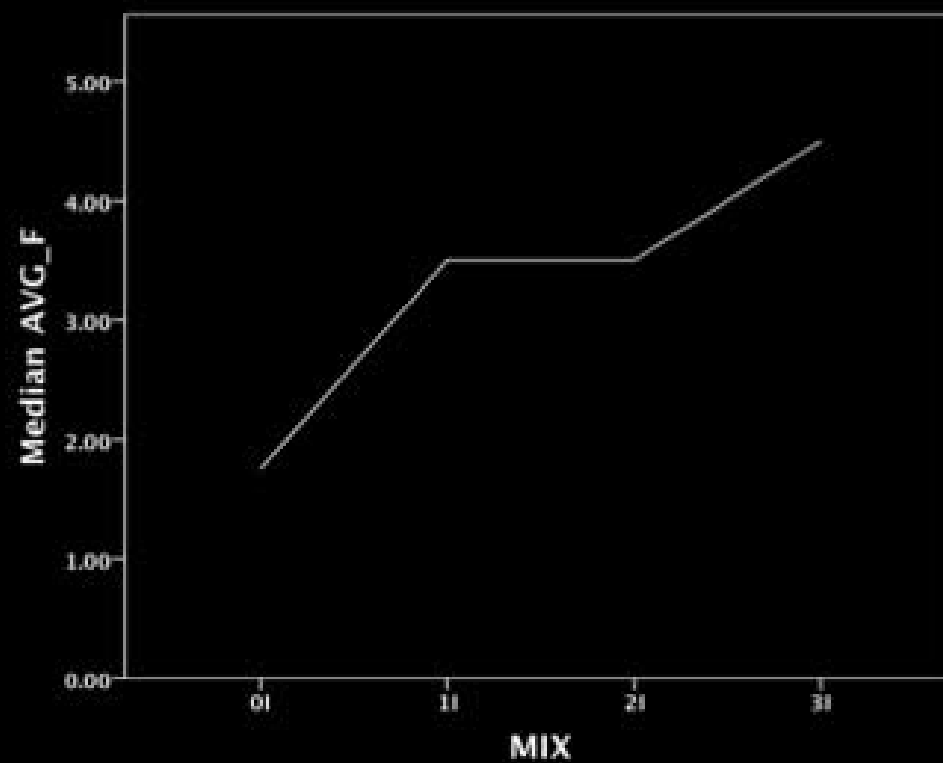
IMPACT OF MIX



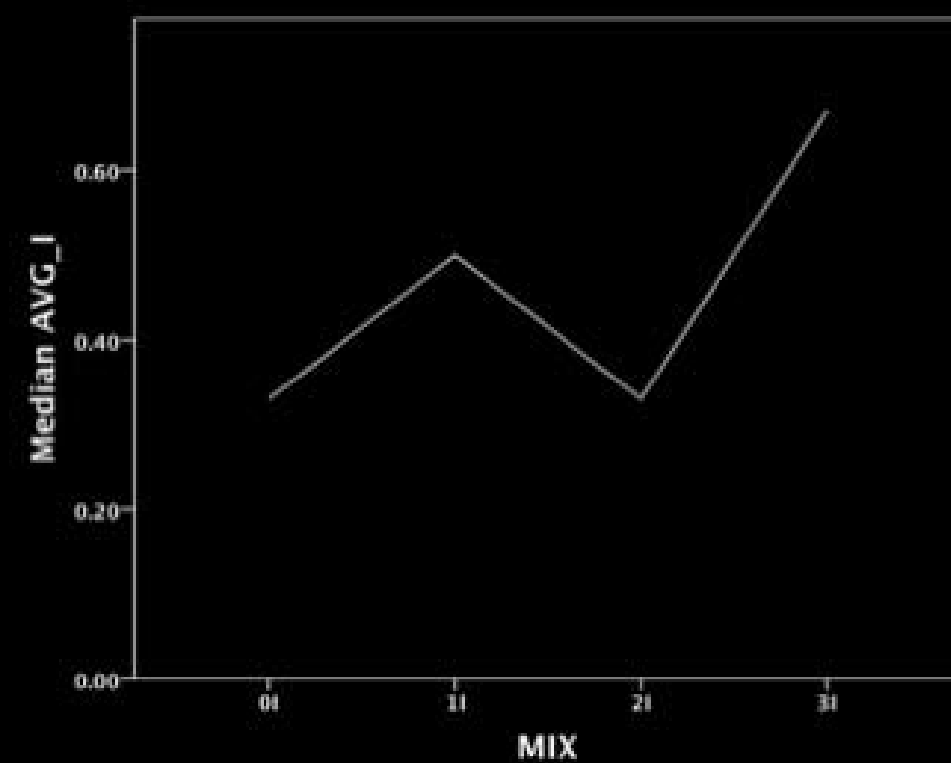
(a) RAW



(b) AVG_R

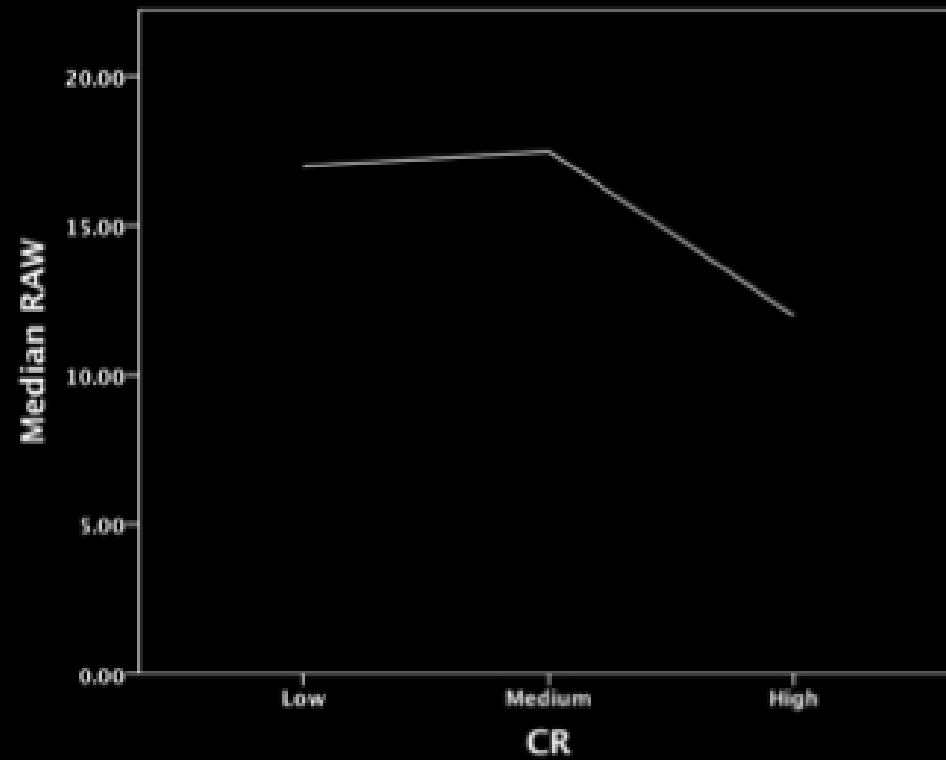


(c) AVG_F

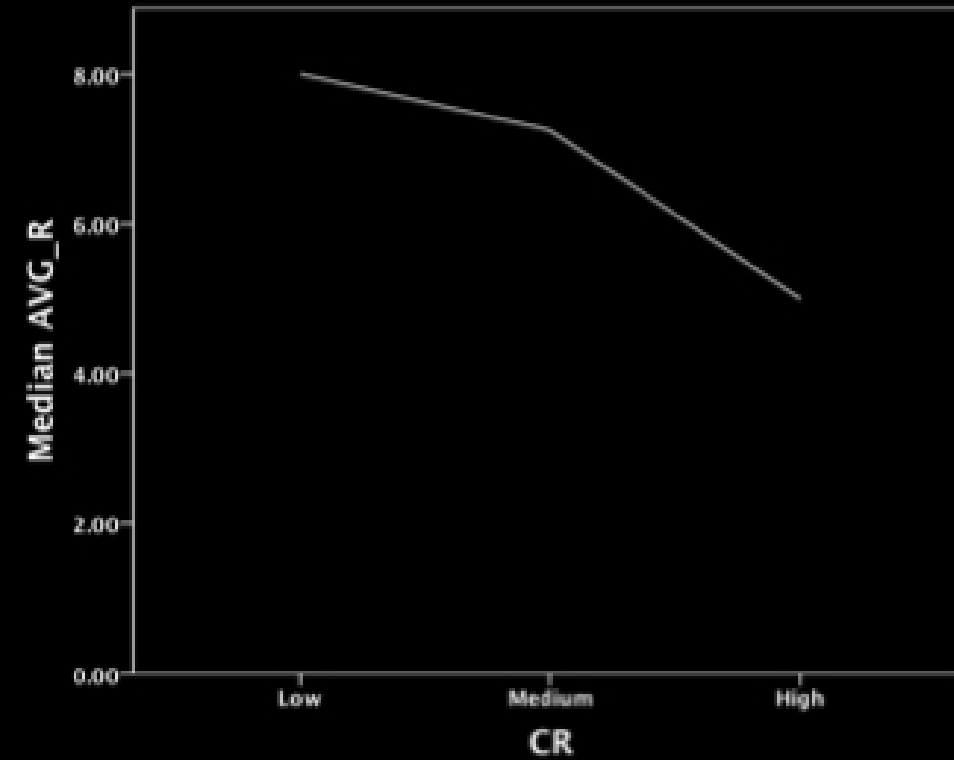


(d) AVG_I

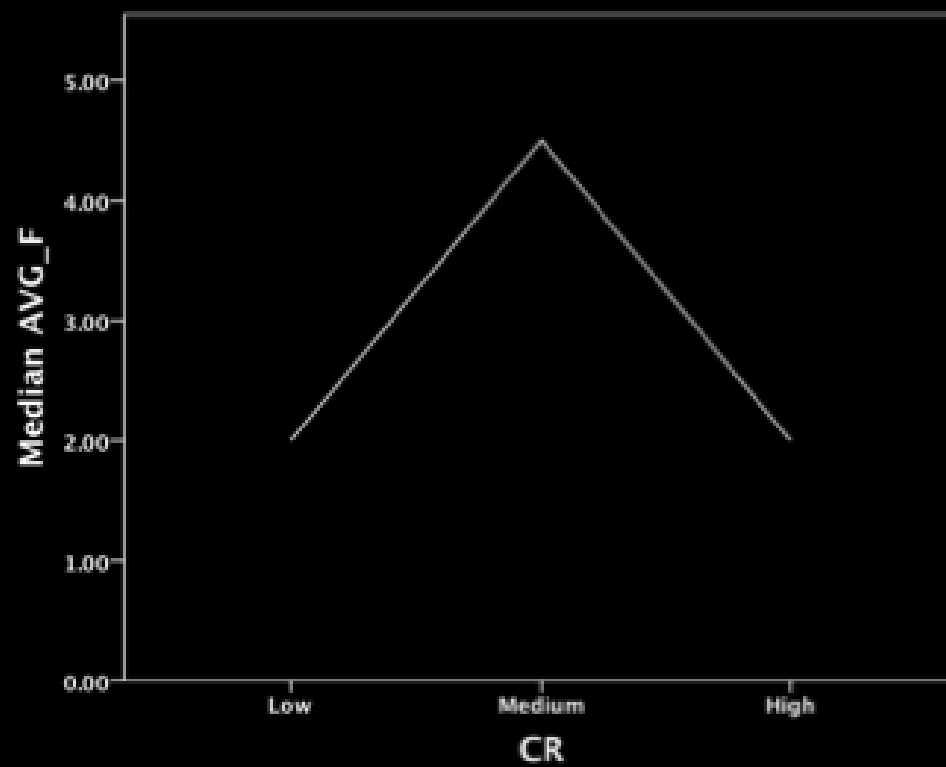
IMPACT OF CR



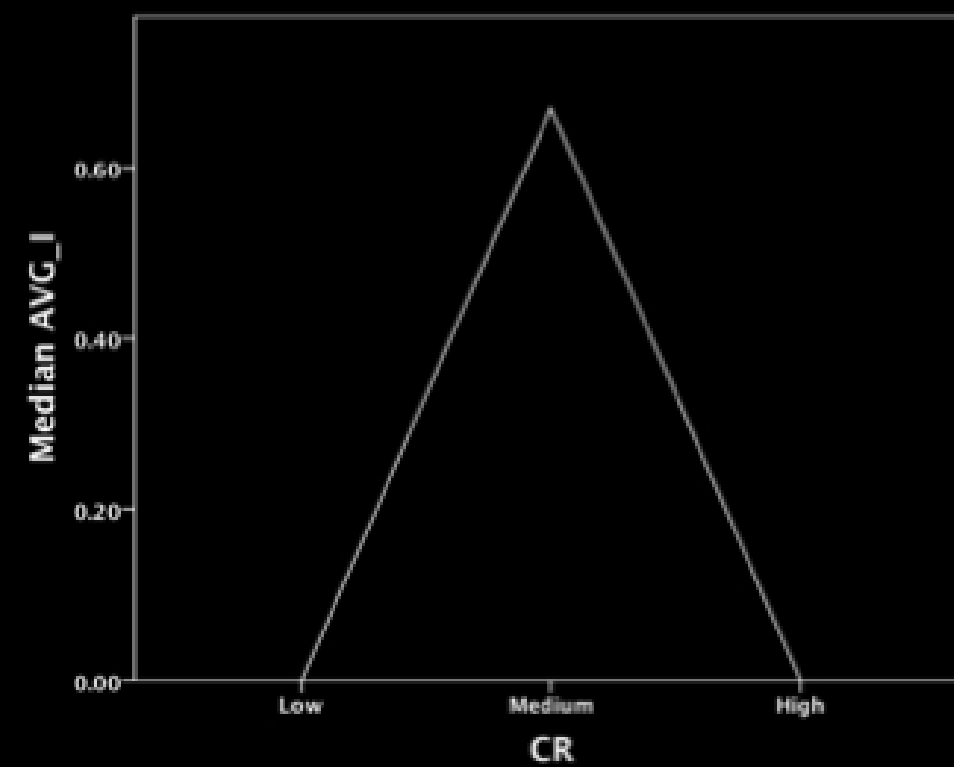
(a) RAW



(b) AVG_R

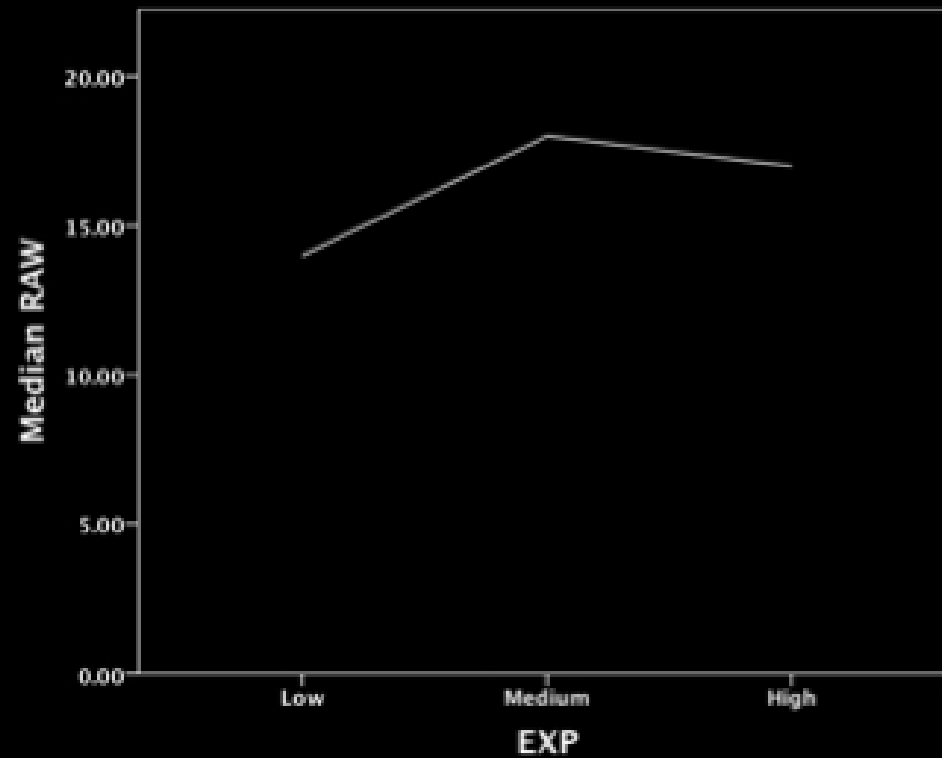


(c) AVG_F

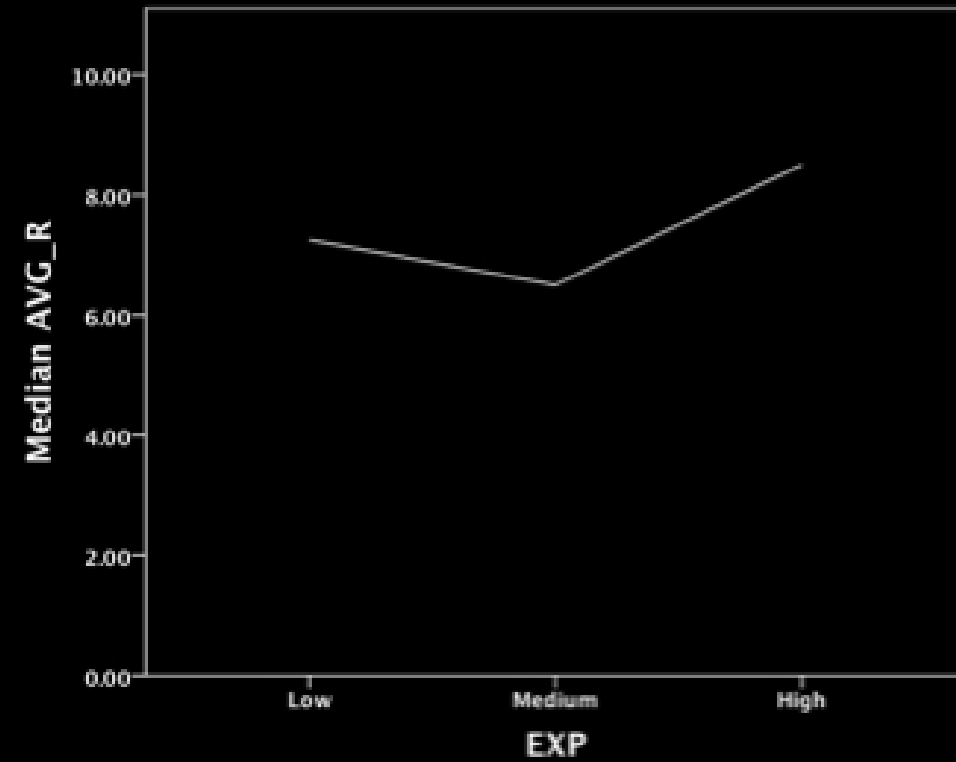


(d) AVG_I

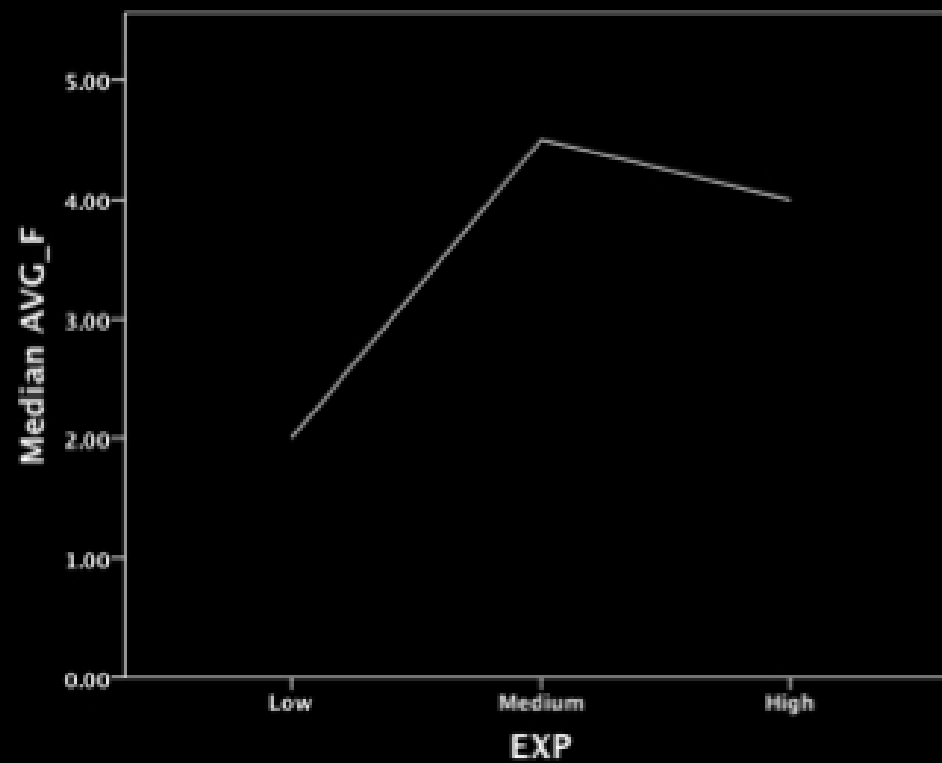
IMPACT OF EXP



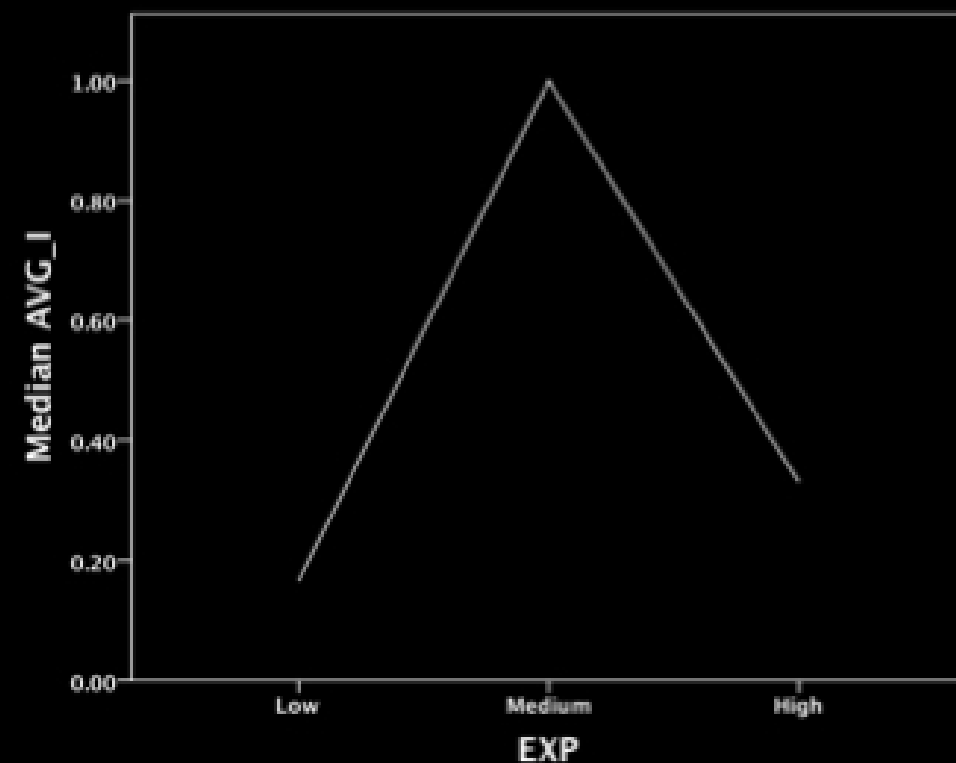
(a) RAW



(b) AVG_R

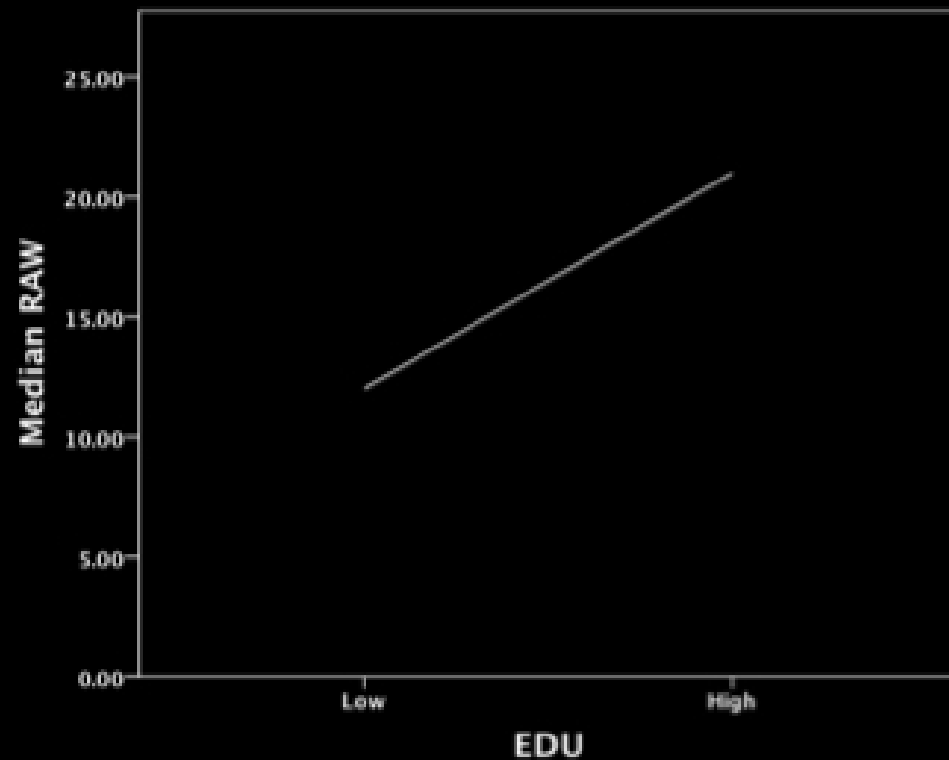


(c) AVG_F

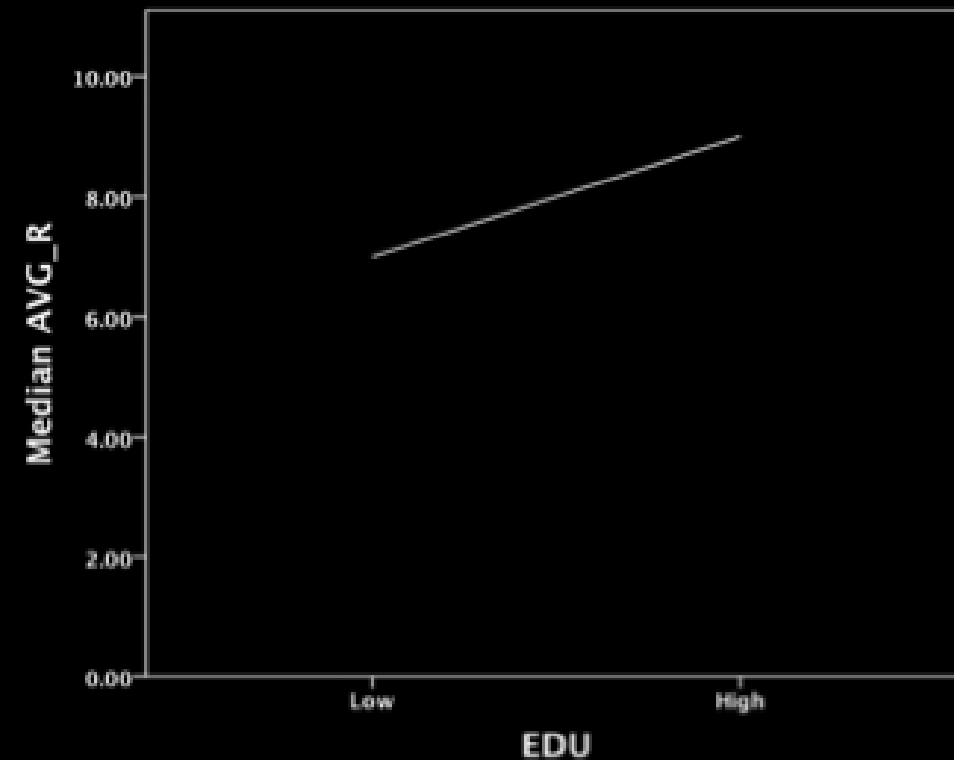


(d) AVG_I

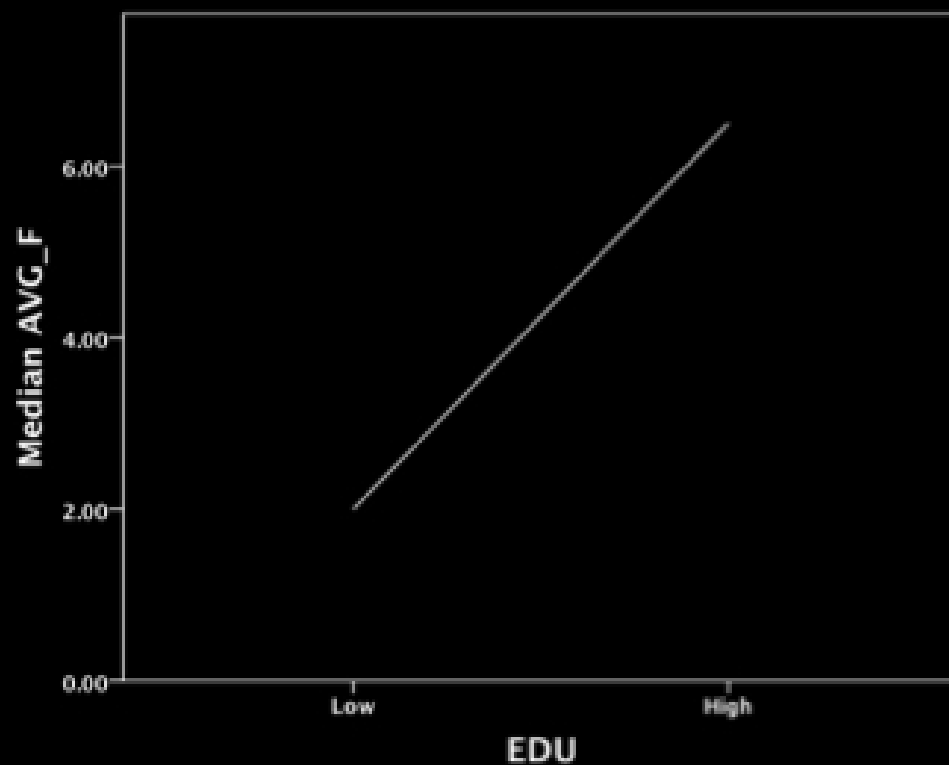
IMPACT OF EDU



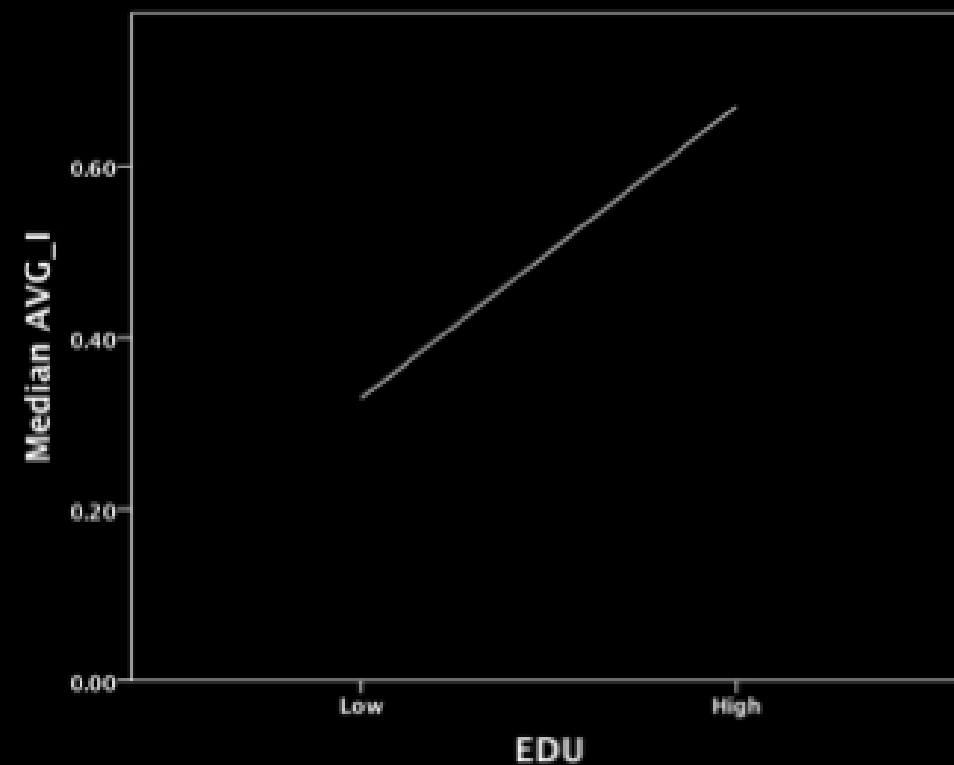
(a) RAW



(b) AVG_R

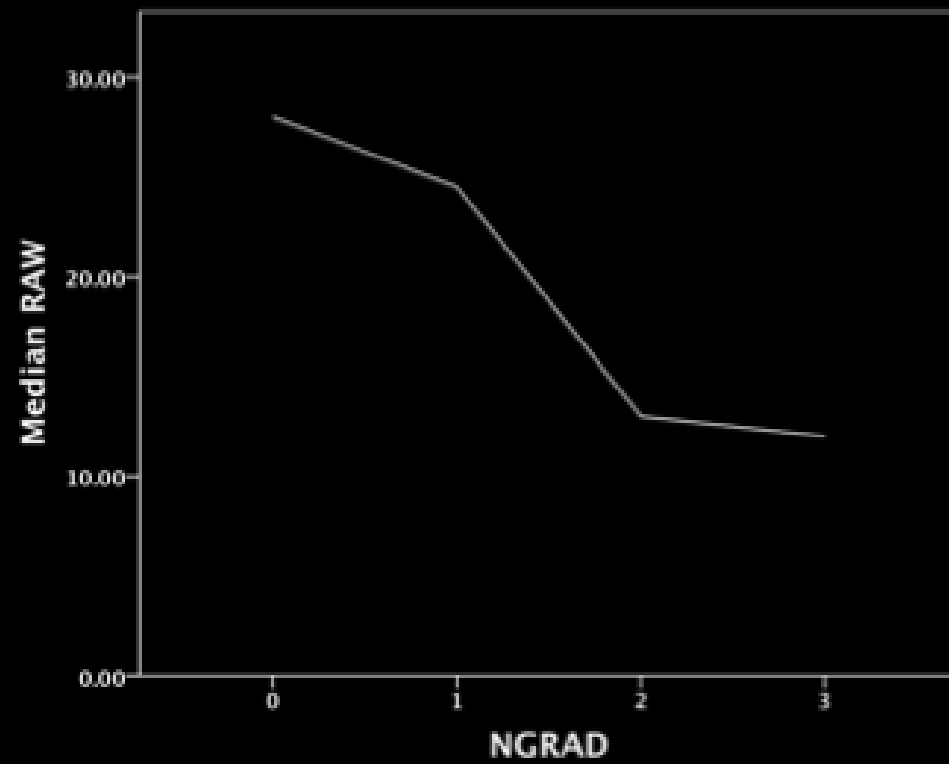


(c) AVG_F

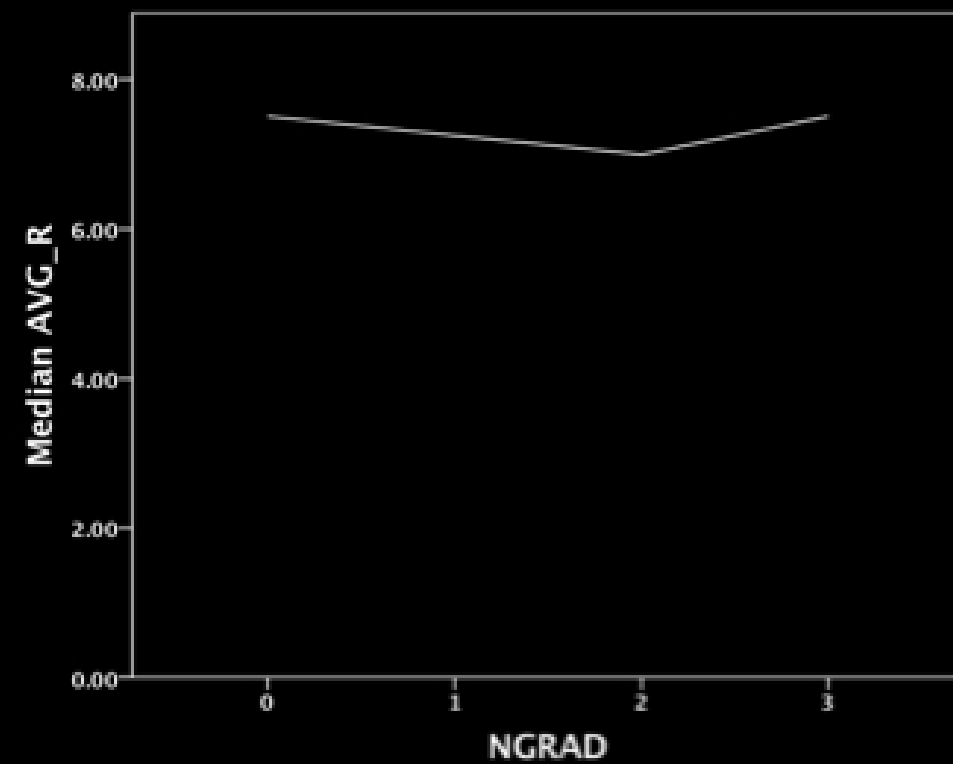


(d) AVG_I

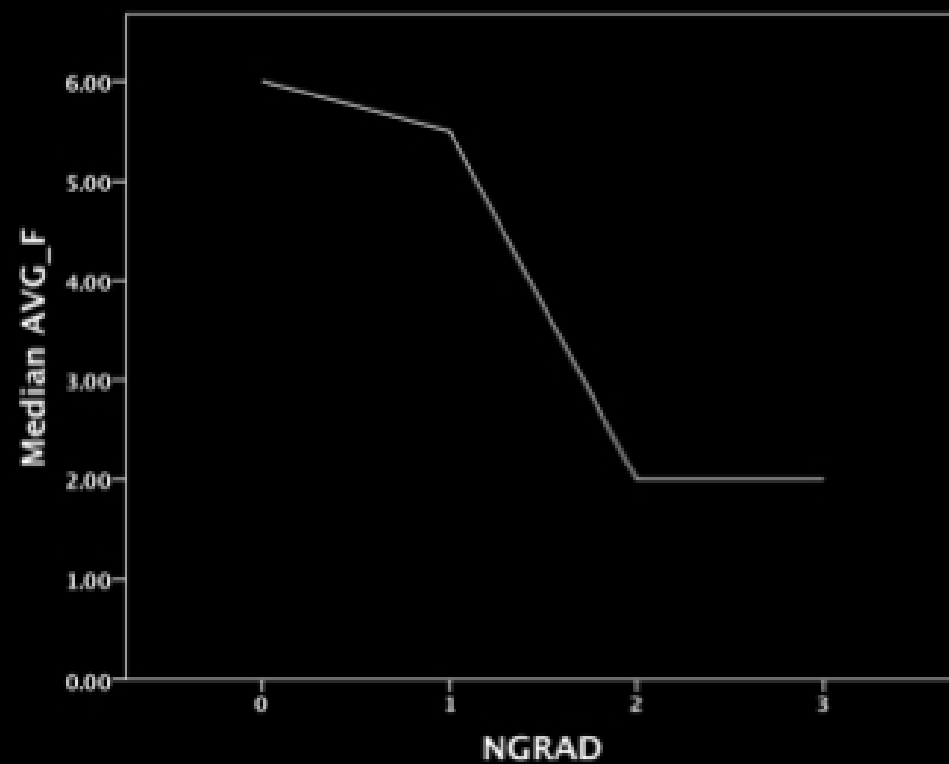
IMPACT OF NGRAD



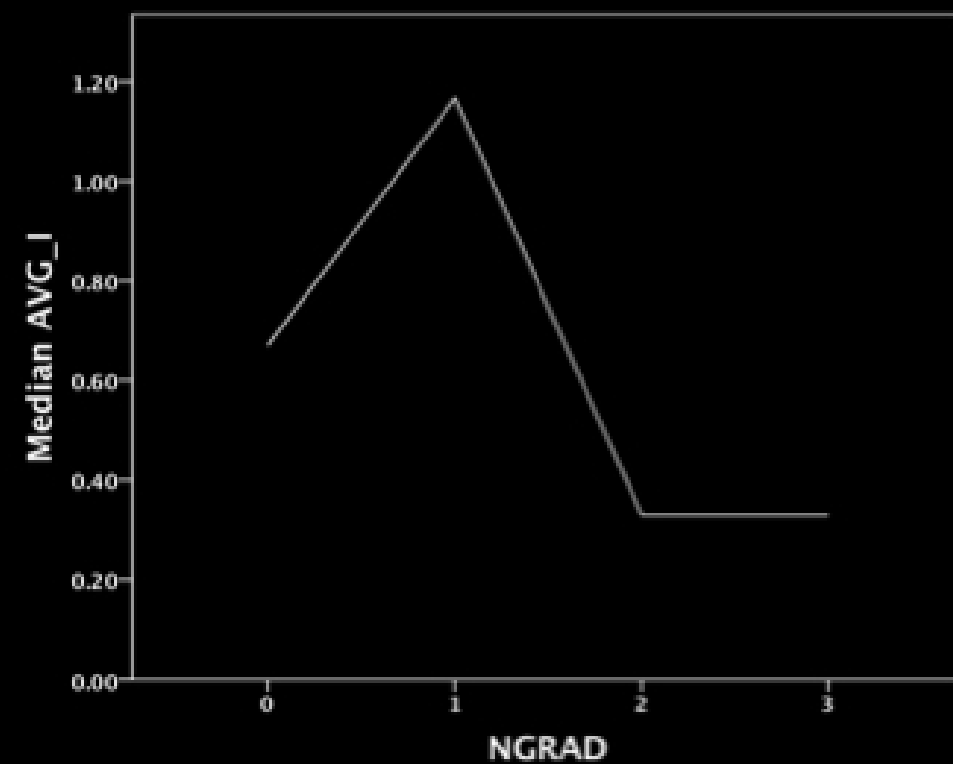
(a) RAW



(b) AVG_R



(c) AVG_F



(d) AVG_I

STATISTICAL ANALYSIS RESULTS

MIX: **no** significant effect on any dependent variable.

CR: **no** significant effect on any dependent variable.

EXP: a significant effect on **only one** dependent variable, **NI**.

EDU: a significant effect **on three** dependent variables, **NRAW, NF** and **NI**.

NGRAD: a significant effect **on three** dependent variables, **NRAW, NF**, and **NI**.

CONCLUSIONS

- In general, teams with at least **one DI** were more effective than teams with **no DIs**.
- Teams with a **medium** level of **CR** were more effective than the others.
- Teams with **no REXP** were at least as effective as teams with **some REXP**.
- A team's **IREXP** was **positively correlated** with the effectiveness of the team.

CONCLUSIONS

- A team's **IEXP** was **positively correlated** with the effectiveness of the team.
- Considering **educational background**,
 - teams with **NCS of 2** were generally most effective,
 - teams with **NSE of 2** were generally most effective.
- A team's **NGRAD** was **negatively correlated** with the effectiveness of the team.

CONCLUSIONS

IMPACT OF THE RESULTS ON THE HYPOTHESES

- H_{MIX} :
 - The initial observations revealed that the effectiveness of a team is **positively affected** by the team's **MIX**.
 - The statistical analysis showed that it **is statistically significant** only **in conjunction** with **EXP** and **EDU**.
 - Therefore, H_{MIX} is **weakly rejected**.

CONCLUSIONS

IMPACT OF THE RESULTS ON THE HYPOTHESES

- H_{CR} :
 - The initial observations revealed that the effectiveness of a team is **positively affected** by the team's **CR**.
 - The statistical analysis showed **no significant effect** of this variable.
 - Therefore, H_{CR} is **rejected**.

CONCLUSIONS

IMPACT OF THE RESULTS ON THE HYPOTHESES

- H_{EDU} :
 - The initial observations revealed that the effectiveness of a team is **positively affected** by the team's **NCS** and **NSE**.
 - The statistical analysis showed that the effect of **NCS** and **NSE** is **statistically significant**.
 - Therefore, H_{EDU} is **strongly accepted**.

CONCLUSIONS

IMPACT OF THE RESULTS ON THE HYPOTHESES

- H_{EXP} :
 - The initial observations revealed that the effectiveness of a team is
 - **positively affected** by the team's **IEXP** and **IREXP**, and
 - **negatively affected** by the team's **REXP**.
 - The statistical analysis showed **no significant effect** of **IEXP** and **IREXP**, and **REXP** showed a **small effect**.
 - Therefore, H_{EXP} is **rejected**.

CONCLUSIONS

IMPACT OF THE RESULTS ON THE HYPOTHESES

- **H_{NGRAD} :**
 - The initial observations revealed that the effectiveness of a team is **negatively affected** by the team's **NGRAD**.
 - The statistical analysis showed that the effect of this variable is **statistically significant**.
 - Therefore, **H_{NGRAD}** is **strongly accepted**.

COMPARING E1 AND E1+E2

- In E1, all of the participants were computer science or software engineering students.
- The results suggest that those RE teams with a **mix** of domain familiarities are more effective than teams composed of **only one domain familiarity**.
- E1 suffered from **unequal** numbers of teams with different mixes of domain familiarities, and therefore, the statistical test results were weak.

COMPARING E1 AND E1+E2

- E2, was conducted using the same plan used for E1 with the goal of having an **equal** number of teams of all mixes of domain familiarity.
- It was necessary to include participants other than computer science and software engineering students in E2.
- After combining the data of E1 and E2, there were an equal number of teams with the different mixes of domain familiarities, and therefore the statistical tests would be more reliable.

COMPARING E1 AND E1+E2

- The initial observations of the results of E1+E2 are not very different from those of E1.
- But the statistical analysis results shows some differences with the statistical analysis of E1.
- E1 data showed some support for accepting H_{MIX} .
- E1+E2 data did not provide any support for accepting H_{MIX} .

WHY E1 AND E1+E2 RESULTS ARE DIFFERENT?

1. Maybe a Type I error occurred during E1:
 - the **null hypothesis** is in fact **true** and there is really **no effect** of the mix of domain familiarities.
2. Maybe a Type II error occurred during E1+E2:
 - the **null hypothesis** is really **false** and the effectiveness of a team is really **affected** by the team's mix of domain familiarities.
3. Maybe differences between the **educational background** of the participants affected the results.

If you have time for case study go to defense slides page 36.

THREATS TO VALIDITY

- The ratios of the ideas in E1 and E2 are different.
 - The differences might be due to the changes in the classifiers.
- To find the cause:
 1. Data were adjusted.
 2. Graphs of
 - the correlations between the **original data** and the dependent variables

were compared with

 - the correlations between the **adjusted data** and the dependent variables.

THREATS TO VALIDITY

- The correlation graphs did not show any significant difference or have a slight difference in strength but the same direction with the corresponding graphs of the unadjusted data.
- Naturally, DAs are better in generating relevant and feasible ideas. Since E2 had significantly more DAs, it is anticipated that the data of E2 had more relevant and feasible ideas.
- The difference between the ratios of the ideas in E1 and E2 is due to the changes in the **participants** not the **classifiers**.

FUTURE WORK

- Replication of the controlled experiment to
 - increase **data points**,
 - improve **external validity**
 - by replicating in different domains and also within industry,
 - improve **internal validity**
 - by controlling more independent variables esp. educational background.

FUTURE WORK

- Apply the study to other disciplines, esp. those that need **tacit assumptions** to be surfaced. e.g. knowledge management.
- Testing level of **domain familiarity**.
- Investigate the impact of participants' knowledge of domains **different** from the domain of the system under study.
- Apply other research methods, e.g. surveys and examination of project histories.
- Try running E2 again with only CS and SE students