THE IMPACT OF

DOMAIN KNOWLEDGE

ON THE EFFECTIVENESS OF

REQUIREMENTS ENGINEERING

ACTIVITIES

ALI NIKNAFS

aniknafs@uwaterloo.ca
OUTLINE

• Introduction

• Controlled Experiments
  • E1
    • E1+E2
  • E1+E2

• Case Study

• Conclusions
The process of arriving at a specification of a set of features that need to be developed is referred to as requirements engineering (RE).
ROLE OF PEOPLE

• Boehm observed that the quality of the development personnel is the most powerful factor in determining an organization’s software productivity.

• Currently, most decisions about staffing development teams arise from anecdotes and folklore, not from scientific studies.
THE RE GAP

• One issue in RE is the gap between what the customer wants and what the analyst thinks the customer wants.

• To bridge this gap, many believe that an analyst needs to know the customer’s problem domain well to do RE well for a system in the domain.

• However, deep knowledge of the problem domain can lead to falling into the tacit assumption tarpit.
BENEFITS OF DOMAIN IGNORANCE

A domain ignorant has:

1. the ability to **think out of the domain’s box**, leading to ideas that are independent of the domain assumptions,

2. the ability to ask questions that expose the domain’s **tacit assumptions**, leading to a common explicit understanding.
IGNORANT
NOT
STUPID!
GOAL

To form the most effective teams of requirements engineers.

Requires answering the research question:

- Does a mix of DIs and DAs perform an RE activity more effectively than only DAs?
CONTROLLED EXPERIMENTS
A team consisting of a mix of DIs and DAs is more effective in a requirements idea generation activity than is a team consisting of only DAs.
EXPERIMENT CONTEXT

- Participants perform the requirement idea generation for some system.
- The units generated are requirements ideas.
- The system is situated in some domain.
- Each participant has a different amount of knowledge about the domain. Each is either:
  - a domain ignorant (DI), or
  - a domain aware (DA).
DOMAIN SELECTION

- **BiDirectional Word Processing** (BDWP)
- Participants were drawn from School of CS;
  - those from the Middle East are DAs.
  - those from elsewhere are DIs.
- Clearly divides the population more so than other domains I tried.
MIX OF DOMAIN FAMILIARITIES

3I: a team consisting of 3 DIs and 0 DAs,

2I: a team consisting of 2 DIs and 1 DA,

1I: a team consisting of 1 DI and 2 DAs, and

0I: a team consisting of 0 DIs and 3 DAs.
PROCEDURE

Part 1
- Read the information letter
- Fill out the general info form
- Sign the consent form
- Take the creativity test

Part 2
- Team assignment
  - Tutorial on Bidirectional Word Processing
  - Brainstorming Session
- Collect the results
ANALYSIS METRICS

- **Quantitative:**
  - Number of generated ideas

- **Qualitative:**
  - Relevancy
  - Feasibility
  - Innovation
EVALUATION OF QUALITY

• To eliminate any bias in classifying an idea that might arise from the evaluator’s knowing the domain familiarity mix of the team from which the idea came,

  • a list of all ideas generated by all teams was produced, and
  • sorted using the first letters of each idea.

• Each evaluator classifies the ideas in the full list.

• After evaluations were done, the each evaluator’s classifications of each idea are transferred to the idea’s occurrences in the individual team lists.

• Berry and I are experts in BDWP and did independent evaluations.
EXPERIMENT 1 (E1)
# INDEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>NAME</th>
<th>VARIABLE</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0I, 1I, 2I, 3I</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>REXP</td>
<td>Average RE experience</td>
<td>None, Some</td>
</tr>
<tr>
<td>IEXP</td>
<td>Average industrial experience</td>
<td>None, 1-2 years, More than 2 years</td>
</tr>
</tbody>
</table>
# DEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>NAME</th>
<th>VARIABLE</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
<td>Raw number of ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_R</td>
<td>Average number of relevant ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_F</td>
<td>Average number of feasible ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_I</td>
<td>Average number of innovative ideas</td>
<td>Numeric</td>
</tr>
</tbody>
</table>
FINE-GRAINED HYPOTHESES

$H_{\text{MIX}}$: The effectiveness of a team in requirements idea generation is affected by the team’s MIX.

$H_{\text{CR}}$: The effectiveness of a team in requirements idea generation is affected by the team’s CR.

$H_{\text{REXP}}$: The effectiveness of a team in requirements idea generation is affected by the team’s REXP.

$H_{\text{IEXP}}$: The effectiveness of a team in requirements idea generation is affected by the team’s IEXP.
CONCLUSIONS

After ANOVA on **RAW, AVG_R, and AVG_F**, and non-parametric test on **AVG_I**,

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

- **$H_{CR}$** is **rejected**: The effectiveness of a team in requirements idea generation **is not** affected by the team’s **CR**.
CONCLUSIONS

• $H_{REXP}$ is rejected:
The effectiveness of a team in requirements idea generation is not affected by the team’s REXP.

• $H_{IEXP}$ is accepted:
The effectiveness of a team in requirements idea generation is affected by the team’s IEXP.
THREATS TO VALIDITY

• **Low Statistical Power**: 20 teams would be enough to achieve statistical power of 0.80, but

  • the **unequal number** of teams in the mixes **reduces** statistical power.

• **Population Validity**: The experiment used student subjects instead of professional analysts, although the students are mostly co-op.
CONTROLLED

EXPERIMENT 1 (E1) + EXPERIMENT 2 (E2)
# INDEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>NAME</th>
<th>VARIABLE</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>REXP</td>
<td>Average RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IEXP</td>
<td>Average industrial experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>NAME</td>
<td>VARIABLE</td>
<td>VALUES</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>REXP</td>
<td>Average RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IEXP</td>
<td>Average industrial experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IREXP</td>
<td>Average industrial RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>NCS</td>
<td>Number of participants with CS background</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>NSE</td>
<td>Number of participants studying SE</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>NGRAD</td>
<td>Number of graduate student participants</td>
<td>0,1,2,3</td>
</tr>
</tbody>
</table>
### DEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>NAME</th>
<th>VARIABLE</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
<td>Raw number of ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>NRAW</td>
<td>Normalized RAW</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_R</td>
<td>Average number of relevant ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>NR</td>
<td>Normalized AVG_R</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_F</td>
<td>Average number of feasible ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>NF</td>
<td>Normalized AVG_F</td>
<td>Numeric</td>
</tr>
<tr>
<td>AVG_I</td>
<td>Average number of innovative ideas</td>
<td>Numeric</td>
</tr>
<tr>
<td>NI</td>
<td>Normalized AVG_I</td>
<td>Numeric</td>
</tr>
</tbody>
</table>
# Factor Analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>REXP</td>
<td>Average RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IEXP</td>
<td>Average industrial experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IREXP</td>
<td>Average industrial RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>NCS</td>
<td>Number of participants with CS background</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>NSE</td>
<td>Number of participants studying SE</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>NGRAD</td>
<td>Number of graduate student participants</td>
<td>0, 1, 2, 3</td>
</tr>
</tbody>
</table>
## FACTOR ANALYSIS

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>REXP</td>
<td>Average RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IEXP</td>
<td>Average industrial experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>IREXP</td>
<td>Average industrial RE experience</td>
<td>None, Low, Medium, High</td>
</tr>
<tr>
<td>NCS</td>
<td>Number of participants with CS background</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>NSE</td>
<td>Number of participants studying SE</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>NGRAD</td>
<td>Number of graduate student participants</td>
<td>0, 1, 2, 3</td>
</tr>
</tbody>
</table>
## FACTOR ANALYSIS

<table>
<thead>
<tr>
<th>NAME</th>
<th>VARIABLE</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIX</td>
<td>Mix of domain familiarities</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>CR</td>
<td>Average creativity score level</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>EXP</td>
<td>Sum of REXP, IREXP, and IEXP</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>EDU</td>
<td>Sum of NCS and NSE</td>
<td>Low, High</td>
</tr>
<tr>
<td>NGRAD</td>
<td>Number of graduate student participants</td>
<td>0,1,2,3</td>
</tr>
</tbody>
</table>
HYPOTHESES

\( H_{\text{MIX}} \): The effectiveness of a team in requirements idea generation is affected by the team’s \( \text{MIX} \).

\( H_{\text{CR}} \): The effectiveness of a team in requirements idea generation is affected by the team’s \( \text{CR} \).

\( H_{\text{EXP}} \): The effectiveness of a team in requirements idea generation is affected by the team’s \( \text{EXP} \).

\( H_{\text{EDU}} \): The effectiveness of a team in requirements idea generation is affected by the team’s \( \text{EDU} \).

\( H_{\text{NGRAD}} \): The effectiveness of a team in requirements idea generation is affected by the team’s \( \text{NGRAD} \).
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 a team.
CONCLUSIONS

• A team’s IEXP was positively correlated with the effectiveness of a team.

• Considering educational background,

  • teams with NCS of 2 were generally most effective,
  • teams with NSE of 2 were generally most effective.
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

- $H_{\text{EXP}}$: 
  - 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, $H_{\text{EXP}}$ 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**.
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
       were
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.
AN INDUSTRIAL 
CASE STUDY
GOAL OF THE STUDY

• To corroborate the conclusions of the controlled experiments, by:

  • getting one group with a **mix** of DAs and DIs to carry out the **idea generation** part of a requirements idea brainstorming session, and

  • then asking the DA members of the group to **compare** the case study session with previous DA-only sessions.
PARTICIPANTS

• **Eight** participants
• **Four** C developers (DAs)
• **Four** UW affiliates (DIs)
PROCEDURE

1. The session started by a brief description of the system given by the supervisor among the DAs.

2. During the session, I monitored generated ideas only to analyze the relation between ideas.

• For each idea, I noted

  1. **who** generated it,
  2. was it **new** (relative to the session), and
  3. which idea, if any, it was **built on**.
OBSERVATIONS

• The DAs were less active than the DIs in the beginning of the session.

• The DAs became more active after DIs threw out some ideas.

• Many ideas offered by DIs appeared to be from outside D’s box.

• DAs built on many of these apparent out-of-the-box ideas.
CONCLUSIONS

• The DIs were generating out-of-the-box ideas.

• The DAs were interested in technical details, as they were seeking only implementable ideas.

• DAs are tied to solutions that they are already familiar with.

• There were indications that the DIs may have generated some ideas that were innovative to C.

• Finally, the experience suggest that, brainstorming groups should be composed of domain experts and new employees.
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.
• 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 did not provide any support for accepting $H_{MIX}$. 
WHY E1 AND E1+E2 RESULTS ARE DIFFERENT?

1. 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. 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.
Why E1 and E1+E2 results are different?

- Maybe differences between the educational background of the participants affected the results.
FUTURE WORK

• Replication of the controlled experiment to
  • increase data points,
  • improve external validity,
  • improve internal validity.

• Apply the study to other disciplines, esp. those that need tacit assumptions to be surfaced. e.g. knowledge management.

• Replication within industry, surveys and examination of project histories.
FUTURE WORK

• Testing level of domain familiarity.

• Investigate the impact of participants’ knowledge of domains different from the domain of the system under study.
THANKS!