How to evaluate an evaluation study: Comparing and contrasting evaluative practices in VIS to those of other disciplines

Anamaria Crisan (@amcrisan) and Madison Elliott
University of British Columbia

Presented at:
BELIV - an IEEE VIS 2018 Affiliated Workshop
October, 20th, 2019

https://bit.ly/2q1QrnC
Why did we write this position paper?

• We are vis researchers with substantive backgrounds in other disciplines

• We found vis evaluation to be at odds with our other disciplines

Ana Crisan
Epidemiology, Statistics
(On the job market this year!)

Madison Elliott
Cognitive Psychology
What does our paper try to do?

- Improve reporting standards in VIS **evaluative user studies**
What does our paper try to do?

• Improve reporting standards in VIS **evaluative user studies**

• Layout the basic structure of a study design
  • For quantitative, qualitative, and mixed methods studies
  • Compare and contrast vis use to other disciplines
What does our paper try to do?

• Improve reporting standards in VIS **evaluative user studies**

• Layout the basic structure of a study design
  • For quantitative, qualitative, and mixed methods studies
  • Compare and contrast vis use to other disciplines

• Provide commentary and background readings
What does our paper try to do?

• Improve reporting standards in VIS **evaluative user studies**

• Layout the basic structure of a study design
  • For quantitative, qualitative, and mixed methods studies
  • Compare and contrast vis use to other disciplines

• Provide commentary and background readings

• Propose a checklist criteria for reviewing evaluative user studies
  • Checklist items for study components
  • Four levels of rigor (unacceptable, baseline, good, and gold-standard)
A snapshot of our paper
You don’t always need an evaluation

• Sometimes an interesting engineering project is a good contribution
  • Novel, innovative ideas are worth publishing
You don’t always need an evaluation

Sad truth: Most "mad scientists" are actually just mad engineers.
You don’t always need an evaluation

• Sometimes an interesting engineering project is a good contribution
  • Novel, innovative ideas are worth publishing

• Techniques, idioms, typologies, and other contributions are huge undertakings
  • Evaluations could be road blocks to success / publication
You don’t always need an evaluation

• Sometimes an interesting engineering project is a good contribution
  • Novel, innovative ideas are worth publishing

• Techniques, idioms, typologies, and other contributions are huge undertakings
  • Evaluations could be road blocks to success / publication

• Contrived evaluations are not useful
You don’t always need an evaluation

• Sometimes an interesting engineering project is a good contribution
  • Novel, innovative ideas are worth publishing

• Techniques, idioms, typologies, and other contributions are huge undertakings
  • Evaluations could be road blocks to success / publication

• Contrived evaluations are not useful

• Claims about the impact on users must be evaluated
  • If you claim it, you should test it!
  • Over speculation of impact should be avoided
You might want to rethink research methods

- Rethinking qualitative/quantitative and objective/subject studies
  - An example formulating research questions about this presentation:

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>Qualitative</td>
<td></td>
</tr>
</tbody>
</table>
You might want to rethink research methods

• Rethinking qualitative/quantitative and objective/subject studies
  • An example formulating research questions about this presentation:

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>How many people are giving this presentation?</td>
<td></td>
</tr>
<tr>
<td>Qualitative</td>
<td></td>
</tr>
</tbody>
</table>
You might want to rethink research methods

- Rethinking qualitative/quantitative and objective/subject studies
  - An example formulating research questions about this presentation:

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantitative</strong></td>
<td>How many people are giving this presentation?</td>
</tr>
<tr>
<td></td>
<td>On a scale of 1 to 5, how much do you think you learned from this presentation?</td>
</tr>
<tr>
<td><strong>Qualitative</strong></td>
<td></td>
</tr>
</tbody>
</table>
You might want to rethink research methods

- Rethinking qualitative/quantitative and objective/subject studies
  - An example formulating research questions about this presentation:

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantitative</strong></td>
<td></td>
</tr>
<tr>
<td>How many people are giving this presentation?</td>
<td>On a scale of 1 to 5, how much do you think you learned from this presentation?</td>
</tr>
<tr>
<td><strong>Qualitative</strong></td>
<td></td>
</tr>
<tr>
<td>What institutions are the presenters from?</td>
<td></td>
</tr>
</tbody>
</table>
You might want to rethink research methods

- Rethinking qualitative/quantitative and objective/subject studies
  - An example formulating research questions about this presentation:

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantitative</strong></td>
<td></td>
</tr>
<tr>
<td>How many people are giving this</td>
<td>On a scale of 1 to 5, how much do you think you learned from this presentation?</td>
</tr>
<tr>
<td>presentation?</td>
<td></td>
</tr>
<tr>
<td><strong>Qualitative</strong></td>
<td></td>
</tr>
<tr>
<td>What institutions are the presenters from?</td>
<td>What did you learn from this presentation?</td>
</tr>
</tbody>
</table>
We should have a checklist for paper reviews

• We identify the following components that should be evaluated upon:
  o Research Motivation and research methodologies
  o Data Collection
  o Data Analysis
  o Validity of findings
  o Generalizability

• Provide general and specific criteria for quantitative, qualitative, and mixed methods designs
Checklist criteria presented in question format

- Examples from **data analysis** checklist components
Checklist criteria presented in question format

- Examples from **data analysis** checklist components

  - Is there a clearly articulated data analysis plan? Is it appropriate?
  - *Within reason and without violating ethics and the study participants’ privacy*, do researchers make analysis artifacts available so that others can independently verify the results?
  - Is it clear whether the researchers are conducting a confirmatory or exploratory study analysis?
  - **Quantitative Methods** | Is it clear what the researcher is measuring?
  - **Quantitative Methods** | Is analysis code made available?
  - **Mixed Methods** | Is is clear when results were analyzed? In sequential study designs, separately and different points in times, in convergent study designs simultaneously.
  - **Qualitative Methods** | Are artifacts of the data coding processing made available?

- See our paper for full checklist – we’ve got a poster too!
Four reporting levels based upon checklist

- Insufficient Reporting
- Bare Minimum Reporting
- See our paper of definitions of each reporting level

- Good Reporting
- Gold Standard Reporting
Checklist goals

• Improve reporting and evaluation standards in VIS research
Checklist goals

• Improve reporting and evaluation standards in VIS research

• Increase replicability and quality of studies
Checklist goals

• Improve reporting and evaluation standards in VIS research
• Increase replicability and quality of studies
• Encourage new discussion about identifying standards for all three methodological approaches: quant, qual, mixed.
Checklist goals

• Improve reporting and evaluation standards in VIS research

• Increase replicability and quality of studies

• Encourage new discussion about identifying standards for all three methodological approaches: quant, qual, mixed.

• Engage VIS researchers in broader, pan-disciplinary methodological discourse
Discussion points:

• What do we want?
  • Clear reporting in VIS evaluative user studies!

• When do we want it?
  • NOW!

• We encourage activate dialogue on refining our checklist criteria

• We recommend checklists be used as standard reviewing practice

Check out our paper!

https://bit.ly/2q1QrnC