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)

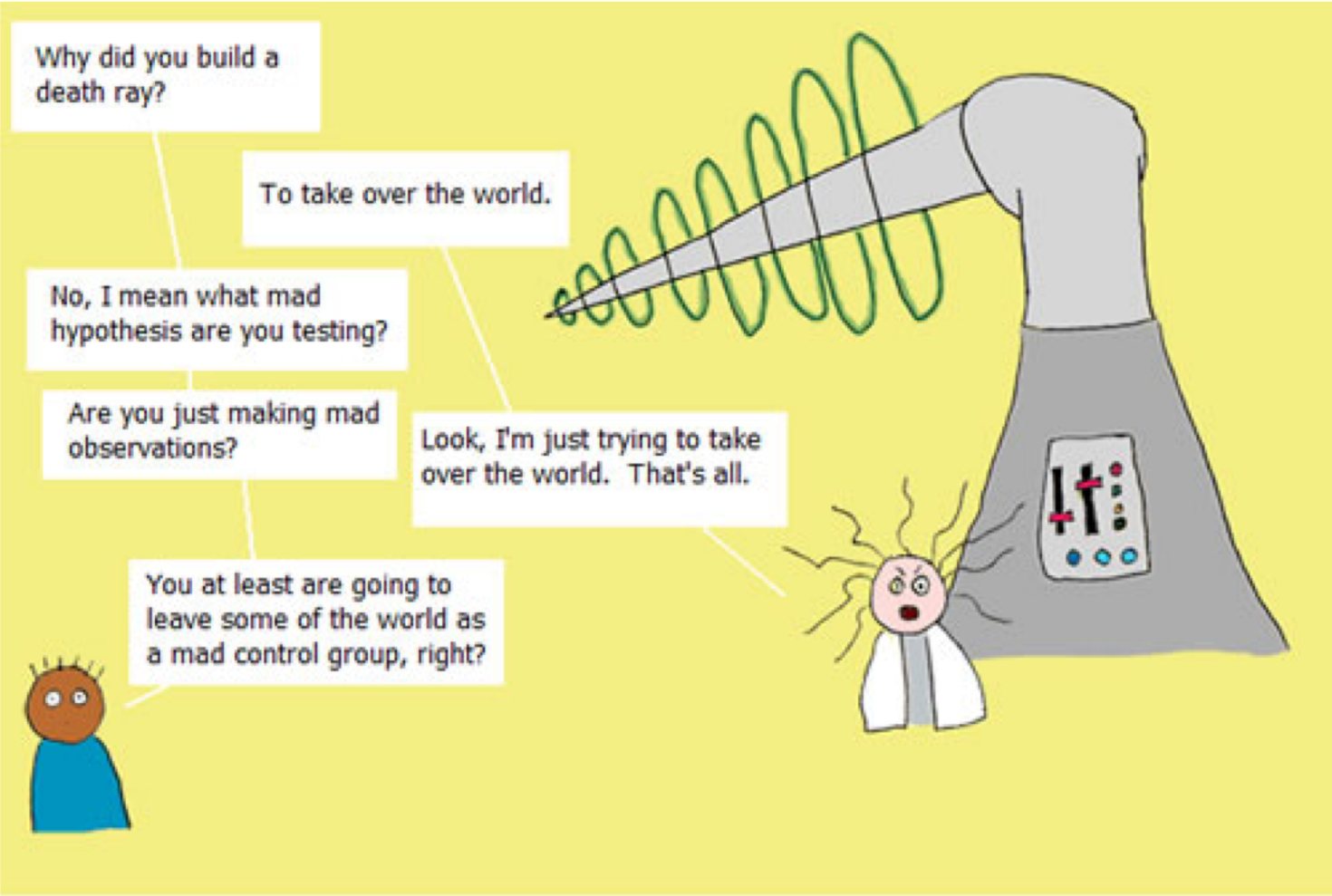
A snapshot of our paper

Photo by Tirachard Kumtanom from Pexels

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:

	Objective	**Subjective**
Quantitative		
Qualitative		

You might want to rethink research methods

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

	Objective	**Subjective**
Quantitative	How many people are giving this presentation?	
Qualitative		

You might want to rethink research methods

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

	Objective	**Subjective**
Quantitative	How many people are giving this presentation?	On a scale of 1 to 5, how much do you think you learned from this presentation?
Qualitative		

You might want to rethink research methods

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

	Objective	**Subjective**
Quantitative	How many people are giving this presentation?	On a scale of 1 to 5, how much do you think you learned from this presentation?
Qualitative	What institutions are the presenters from?	

You might want to rethink research methods

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

	Objective	**Subjective**
Quantitative	How many people are giving this presentation?	On a scale of 1 to 5, how much do you think you learned from this presentation?
Qualitative	What institutions are the presenters from?	What did you learn from this presentation?

We should have a checklist for paper reviews

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

Reject / Revise	Accept
• Insufficient Reporting • Bare Minimum Reporting	• Good Reporting • Gold Standard Reporting

- See our paper of definitions of each reporting level

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