



# Evaluating Methodology

# What we will cover:

- Selection and demographics of subjects
- Concepts of validity and reliability
- Demonstrate validity
- Demonstrating reliability





## 1. Subject Selection

1. Subject selection
2. Presentation
3. Sampling bias
4. Example

- How were the subjects selected?
- What population do they represent?
- Can the sample be generalised?
- Can the sample be compared?

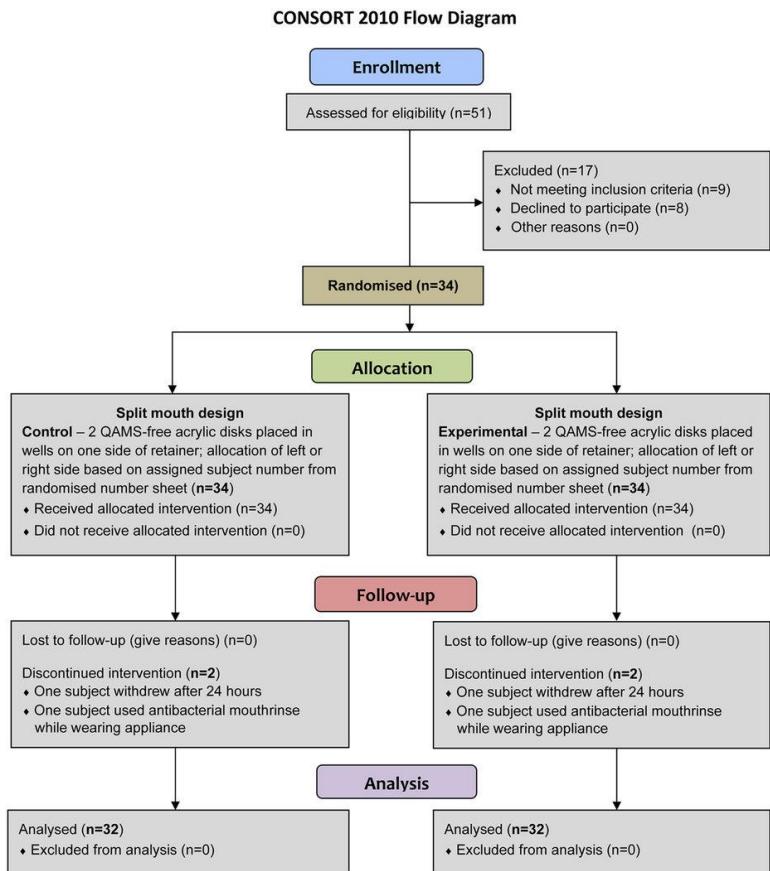


1. Subject selection
2. **Presentation**
3. Sampling bias
4. Example

- Every subject should be accounted for!
- Are there statements of which samples are included and excluded?
- Is there a sampling flow chart or a table?



1. Subject selection
2. Presentation
3. Sampling bias
4. Example



- 51 subjects
- 17 excluded
- 2 experimental groups
- 2 lost to follow up
- 32 completed

Liu, S.-y. et al. (2016)

Antimicrobial activity of a quaternary ammonium methacryloxy silicate-containing acrylic resin: a randomised clinical trial. *Sci. Rep.* **6**, 21882; doi: [10.1038/srep21882](https://doi.org/10.1038/srep21882)



1. Subject selection
2. Presentation
- 3. Sampling bias**
4. Example

- Opportunity sampling
- Volunteer sampling
- Stratified sampling
- Random sampling



1. Subject selection
2. Presentation
- 3. Sampling bias**
4. Example

- Opportunity sampling
  - Most convenient participants available



1. Subject selection
2. Presentation
- 3. Sampling bias**
4. Example

- Volunteer sampling
  - May share personality characteristics



1. Subject selection
2. Presentation
- 3. Sampling bias**
4. Example

- Stratified sampling
  - Potential selection bias



1. Subject selection
2. Presentation
- 3. Sampling bias**
4. Example

- Random sampling
  - Unbalanced clusters of demographics



1. Subject selection
2. Presentation
3. Sampling bias
4. Example



1. Subject selection
2. Presentation
3. Sampling bias
4. Example





## 2. Validity and Reliability

*“The principles of validity and reliability are fundamental cornerstones of the scientific method. Together they are at the core of what is accepted as scientific proof”*

Martyn Shuttleworth, 2008



1. Validity
2. Reliability
3. Reproducibility
4. Example

- The ability to measure what you intended to measure
  - Internal validity
  - External validity



1. Validity
2. Reliability
3. Reproducibility
4. Example

- Internal validity
  - Does the method measure what it intends?



1. Validity
2. Reliability
3. Reproducibility
4. Example

- External validity
  - Can the results be generalized?



1. Validity
2. **Reliability**
3. Reproducibility
4. Example

- Ability to measure consistently
- Expect to see the same results
- Indicator of reproducibility



1. Validity
2. Reliability
- 3. Reproducibility**
4. Example

- Ability of the experiment to be duplicated
- How validity and reproducibility are checked
- External verification
- Suggest the author fill in the blanks



1. Validity
2. Reliability
3. Reproducibility
4. Example



**Reliable  
Not Valid**



**Low Validity  
Low Reliability**



**Not Reliable  
Not Valid**



**Both Reliable  
and Valid**

by Experiment-Resources.com





### 3. Demonstrating Validity

1. Face validity
2. Construct validity
3. Content validity

- Face Validity
- Construct Validity
- Content Validity



Measuring physical health, based on number of visits to the doctor in the past 12 months



1. Face validity
2. Construct validity
3. Content validity

- Whether the measure appears appropriate for the variable it is setting out to assess

### *Example*

- Doctor visits are a proxy measure for 'good health'



1. Face validity
2. **Construct validity**
3. Content validity

- Whether the specific measure measures the theoretical concept adequately

### *Example*

- Visits to the doctor may measure psychological factors instead of physical



1. Face validity
2. **Construct validity**
3. Content validity

- Intentional desirability bias
  - Report what they think the researcher wants to hear



1. Face validity
2. **Construct validity**
3. Content validity

- Unintentional behavioural bias
  - Alter behaviour in response to the hypothesis



1. Face validity
2. Construct validity
- 3. Content validity**

- Whether the measure fully represents the theoretical concept

### *Example*

- Visits to the doctor do not only correlate to physical health





## 4. Demonstrating Reliability

1. Double coding
2. Re-testing

- Obtain opinions from two people on subjective decisions

*Example*

- If two radiologists disagree over an x-ray, a third radiologist may be required



1. Double coding
2. Re-testing

- The same people perform a test twice

*Example*

- Differences in test results indicate an error with the test measures
- Large differences mean the test is not valid



## Methods

Data for the 20 year period from 1995 to 2014 were obtained from the Darwin Awards (<http://darwinawards.com>). Nominations for a Darwin Award are evaluated according to five rigorous selection criteria: death, style, veracity, capability, and self selection.<sup>20</sup>

- The candidate must be eliminated from the gene pool
- The candidate must show an astounding misapplication of common sense
- The event must be verified
- The candidate must be capable of sound judgment
- The candidate must be the cause of his or her own demise.

The Darwin Awards are open to all ethnic groups, cultures, and socioeconomic groups.

We reviewed all Darwin Award nominations, noting the sex of the winner. Our analysis included only confirmed accounts verified by the Darwin Awards Committee. Urban legends and unverified accounts were excluded. Honourable

mentions—worthy examples of idiotic behaviour not resulting in elimination from the gene pool—were also excluded from the analysis. Examples include the man who slipped when using a belt sander as an auto-erotic device and lost a testicle.

Repairing his scrotum with a staple gun, he was able to salvage his remaining testicle thus failing to eliminate himself completely from the gene pool.<sup>17-21</sup>



**publons**

Search, or import by DOI/arXiv/PMID

HOME BROWSE COMMUNITY FAQ

1

Publications ▶ Publication Details

why it is important to understand consumer perceptions of and attitudes towards geovisualization technology – simply stating that the study has not been done is a weak justification for study. It is important to first highlight to what extent existing (or closely-related) studies in the area has helped us to understand the investigated phenomenon, followed by clearly stating the extant gap in understanding, the corresponding research objectives/questions that will address the identified gap, and the contribution of the study to both theory and practice.

Second, the arguments supporting the hypothesis do not provide adequate answers as to "why" the hypotheses were proposed (e.g. what the underlying reasons that suggest GRA will have a positively influence on PE?) – simply stating that other studies have found a positive relationship between similar constructs do not provide answers to the "why" question (e.g. "Win et al. [63] highlight the influence of spatial ability on enjoyment and presence within virtual environment. Ho [35] found a significant relationship between the use of location-based services and perceived enjoyment. ... etc." in page 6 – this does not provide the answers to "why", thus readers cannot comprehend why you are proposing such hypotheses).

Third, I am not convinced about the current method of investigating the effects of a user's ability on their perceptions (i.e. as a latent variable), simply because "ability" is a user characteristic and thus should be measured as a moderator rather than a latent variable. I believe that the GR ability will not be the same for everyone – that is, some people might score high on GR ability measures and thus indicating that they have high GR ability, while others might score low on GR ability measures and thus indicating that they have low GR ability (I believe you will understand my point here because you talked about low GR ability somewhere in your implications). Thus, I believe it would be better to create an index to identify users with high and low GR ability and see whether there is a difference in the important of PE, PEOU, and PU in influencing their attitudes and intentions to use geovisualization technology.

Fourth, the discussion section is poorly written – it is unclear what exactly are the new theoretical contributions of the current study to the existing literature (i.e. implications to scholarship). I believe this can be addressed by articulating how the findings of the current study is similar or is different

Lim WM. [Pre-pub review, Jul 2014](#) for: Erskine MA, Gregg DG & Karimi J (2016) Perceptions and Attitudes toward Online Mapping Services, Journal of Computer Information Systems, 56:2, 175-184, DOI: [10.1080/08874417.2016.1117836](https://doi.org/10.1080/08874417.2016.1117836)





p

## 5. Summary

# Summary

- The importance of subject selection
- Subject retention through studies
- Validity - measuring what is intended
- Reliability - providing consistent measurements
- Demonstrated validity and reliability

