Evaluating Experiments

Idaho State University | Computer Science

Isaac Griffith

CS 6620
Department of Informatics and Computer Science
Idaho State University

ROAR

Evaluating SMSs

Petersen et al. presents a set of evaluation criteria and a means by which an SMS study can be scored based on the actions taken during the following phases of the study:

- Need for review

- Study Identification
  - Choosing the search strategy
  - Developing the search
  - Evaluating the search
  - Inclusion/Exclusion criteria

- Data Extraction and Classification
  - Extraction process
  - Classification Scheme

- Validity Discussion

ROAR

Need for Review

- Actions:
  - Motivated the need and relevance of the study
  - Defined objectives and questions
  - Consulted with target audience to define questions

Evaluation	Description	Score
No description	The study is not motivated and the goal is not stated	0
Partial evaluation	Motivations and questions are provided	1
Full evaluation	Motivations and questions are provided, and have been defined in correspondence with target audience	2

Choosing Search Strategy

- Actions:
  - Conducted snowballing (forward or backward)
  - Conducted a manual search
  - Conducted a database search

Evaluation	Description	Score
No description	Only one type of search has been conducted	0
Minimal evaluation	Two search strategies have been used	1
Full evaluation	All three search strategies have been used	2

ROAR

Developing the Search

- Actions:
  - Improving the search string
    - Used PICO(C) to identify search terms
    - Consulted with librarians or experts
    - Iteratively tried finding more relevant papers
  - Selecting relevant terms
    - Used keywords from known papers
    - Used standards, encyclopedias, and thesaurus

ROAR

Developing the Search

Evaluation	Description	Score
No description	No actions have been taken to further develop the search	0
Minimal evaluation	At least one action was taken to improve search string xor an action to identify relevant terms was taken	1
Partial evaluation	At least one action was taken to improve search string and one action to identify relevant terms was taken	2
Full evaluation	All actions identified have been taken	3

ROAR

Search Evaluation

- Actions:
  - Evaluate the Search
    - Test-set of known papers
    - Expert evaluates results
    - Search web-pages of key authors
    - Test-retest
  - Inclusion and Exclusion
    - Identify objective criteria for decision
    - Add additional reviewer, resolve disagreements between them when needed
    - Decision rules

Search Evaluation

Evaluation	Description	Score
No description	No actions have been reported to improve the reliability of the search and inclusion/exclusion	0
Minimal evaluation	At least one action has been taken to improve the reliability of the search xor the reliability of the inclusion/exclusion	1
Partial evaluation	At least one action has been taken to improve the reliability of the search and the inclusion/exclusion	2
Full evaluation	All actions identified have been taken	3

ROAR

Extraction and Classification

- Actions:
  - Extraction Process
    - Identify objective criteria for decision
    - Obscuring information that could bias
    - Add additional reviewer, resolve disagreements between them when needed
    - Test-retest
  - Classification Scheme
    - Research type
    - Research method
    - Venue type

ROAR

Extraction and Classification

Evaluation	Description	Score
No description	No actions have been reported to improve on the extraction process or enable comparability between studies through the use of existing classifications	0
Minimal evaluation	At least one action has been taken to increase the reliability of the extraction process	1
Partial evaluation	At least one action has been taken to increase the reliability of the extraction process, and research type and method have been classified	2
Full evaluation	All actions identified have been taken	3

ROAR

Study Validity

- Actions:
  - Validity discussion/limitations provided

Evaluation	Description	Score
No description	No threats or limitations are described	0
Full evaluation	Threats and limitations are described	1

ROAR

Are there any questions?

ROAR