



Evaluating the Efficacy of Concern-Driven Metrics: A Comparative Study

Cláudio Sant'Anna - Federal University of Bahia (UFBA), Brazil

Alessandro Garcia - Lancaster University, UK

Carlos Lucena - PUC-Rio, Brazil

(santanna@dcc.ufba.br)

Concern Modularization

- Achieving modular designs is far from being trivial
 - multiple widely-scoped concerns need to be simultaneously modularized
- Inadequate concern modularizations may lead to multiple design flaws
 - E.g. low coupling and narrow interfaces

Design Modularity Assessment

- Software metrics are key mechanisms for identifying modularity-related design flaws
- Although typical design flaws are related to the inadequate modularization of concerns
 - Most of the current design metrics are not calibrated to the concerns that drive the design
 - Design assessment is mostly rooted at module-based metrics
 - Traditional measures of modularity
 - class coupling, class cohesion and interface size

Concern-driven Assessment

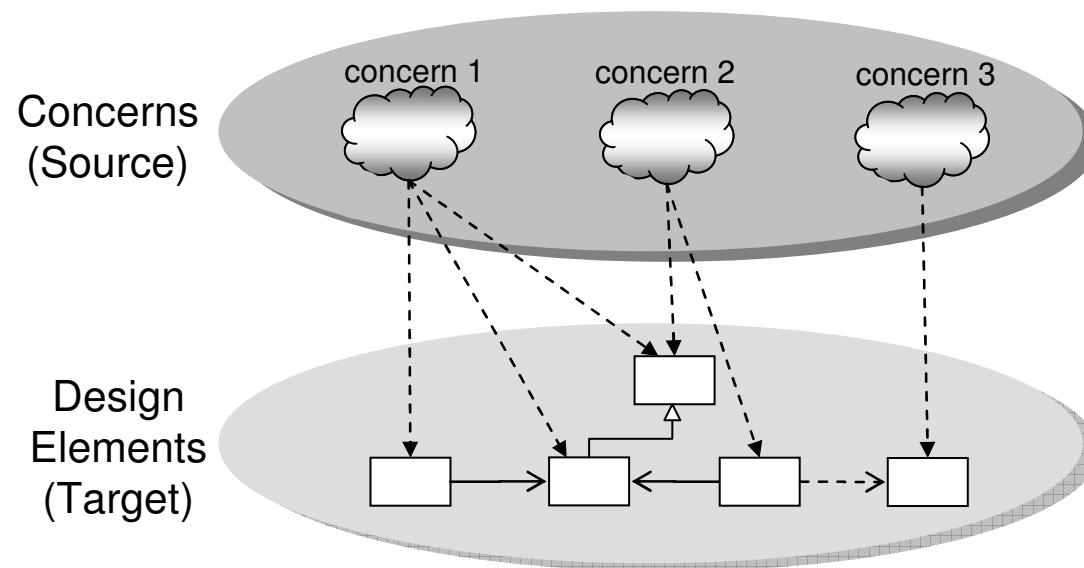
- Growing body of relevant work on concern analysis techniques
 - Emergence of new design decomposition approaches
- Concern-driven metrics
 - There is not much knowledge and empirical evidence on the efficacy of concern-driven metrics
- Are concern-driven metrics efficacious?
 - for identifying design flaws
 - compared with conventional module-based metrics

Our Study

- First exploratory study
- Compares the efficacy of concern-driven and conventional metrics to identify modularity-related design flaws
 - Number of classes correctly identified as suspects of exhibiting specific design flaws

Concern-Driven Metrics

- Capture modularity properties associated with the realization of concerns in software artefacts
- Concern-driven metrics are based on a concern-to-design (or concern-to-code) mapping.



Concern-Driven Metrics

- Identification of design flaws caused by the poor modularization of concerns.
 - E.g. Scattering and Tangling
- Metrics evaluated in our empirical study
 - Concern Diffusion over Components (CDC)
 - Concern Diffusion over Operations (CDO)
 - Lack of Concern-based Cohesion (LCC)

Study Settings

- Goal
 - Compare concern-driven and conventional metrics
 - detection of two specific bad smells
 - *Shotgun Surgery*
 - *Divergent Change*
- Target System
 - Health Watcher
 - Concerns: graphical user interface (GUI), business rules, concurrency, distribution, exception handling, and persistence

Study Settings

- Subjects
 - Eight master students grouped in pairs
 - Each pair worked with different metrics
 - Conventional metrics (two groups)
 - Concern-driven metrics (one group)
 - Both conventional and concern-driven metrics (one group, called hybrid metrics group)

Study Settings

- Involved Metrics
 - Concern-driven
 - Concern Diffusion over Component (CDC)
 - Concern Diffusion over Operations (CDO)
 - Lack of Concern-based Cohesion (LCC)
 - Conventional
 - Coupling Between Object Classes (CBO)
 - Lack of Cohesion in Methods (LCOM)
 - Weighted Methods per Class (WMC)
 - Number of Attributes (NOA)
 - Number of Operations (NOO)

Study Settings

- Activities
 - Training on the metrics and bad smells
 - The groups were provided with the metrics results
 - Each group
 - Read the description of the Health Watcher design
 - Identify the classes with Divergent Change
 - Identify the classes with Shotgun Surgery
 - Register the time spent on each task
 - Document which metrics they found useful for detecting the bad smell

Hypothesis

- *The hybrid metrics suite is the most effective to support the detection of design bad smells.*
- To test the hypothesis we compared
 - the list of classes identified by each group as having the bad smell
 - the list of classes which actually has the bad smell
- We identified the actual instances of bad smells
 - Divergent Change: 12 classes
 - Shotgun Surgery: 8 classes

Results

	Conventional Metrics	Conventional Metrics	Concern-driven Metrics	Hybrid Metrics
Divergent Change Identification				
Time (minutes)	9	10	21	31
Hits	2 (17%)	2 (17%)	12 (100%)	9 (75%)
False positives	1 (33%)	2 (50%)	7 (36%)	0 (0%)
Shotgun Surgery Identification				
Time (minutes)	6	10	13	35
Hits	1 (12%)	1 (12%)	6 (75%)	1 (12%)
False positives	3 (75%)	2 (33%)	11 (64%)	3 (75%)

Results – Divergent Change

	Conventional Metrics	Conventional Metrics	Concern-driven Metrics	Hybrid Metrics
Divergent Change Identification				
Time (minutes)	9	10	21	31
Hits	2 (17%)	2 (17%)	12 (100%)	9 (75%)
False positives	1 (33%)	2 (50%)	7 (36%)	0 (0%)

- Conventional metrics performed significantly worse than the others
 - Lack of Cohesion in Methods (LCOM) reported as most useful
 - high values for classes with no design anomaly due to “getter” and “setters”
- Concern-driven metrics performed very well (100% of hits)
 - Lack of Concern-based Cohesion (LCC) reported as most useful
 - 36% of false positives due to high number of listed classes
- Hybrid metrics also performed well
 - The presence of LCC minimized the limitations of LCOM

Results – Shotgun Surgery

	Conventional Metrics	Conventional Metrics	Concern-driven Metrics	Hybrid Metrics
Shotgun Surgery Identification				
Time (minutes)	6	10	13	35
Hits	1 (12%)	1 (12%)	6 (75%)	1 (12%)
False positives	3 (75%)	2 (33%)	11 (64%)	3 (75%)

- Again conventional metrics did not perform well
- Hybrid metrics did not perform well either
 - Some conventional metrics (in general, size metrics) might have introduced “noise” in the design assessment.
- Concern-driven metrics performed well again (75% of hits)
 - Concern Diffusion over Components (CDC) reported as most useful
 - 65% of false positives due to high number of listed classes

Conclusions

- Results partially contradict our hypothesis that the hybrid metrics suite would be the most effective.
 - The concern-driven metrics was the most effective
 - Hybrid metrics did not performed well for identifying shotgun surgery
 - The hybrid metrics group took the longest time to finish their tasks

Conclusions

- Study Constraints
 - The results are restricted to the involved metrics, bad smells and the target software system
 - The population was not statistically representative
 - The process of assigning concerns to design elements directly impacts on the measurement results
 - We played a crucial role in deciding which classes were affected by the bad smells

Conclusions

- Despite the limitations, this first study showed that concern-driven metrics for assessing design modularity is worth studying further



Evaluating the Efficacy of Concern-Driven Metrics: A Comparative Study

Cláudio Sant'Anna - Federal University of Bahia (UFBA), Brazil

Alessandro Garcia - Lancaster University, UK

Carlos Lucena - PUC-Rio, Brazil

(santanna@dcc.ufba.br)

Conclusions

- Ongoing work
 - We replicated the study with more students
 - Also included other bad smells
 - The results for divergent change and shotgun surgery were similar to the results of this first study
 - However, for a different bad smell, namely *god class*, the hybrid metrics performed better than the concern-driven metrics

	Conventional Metrics	Concern-driven Metrics	Hybrid Metrics
God Class Identification			
Time (minutes)	18	56	48
Hits	1 (33%)	2 (66%)	3 (100%)
False positives	2 (66%)	0 (0%)	0 (0%)