

# The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature

Poorvi Rai  
North Carolina State University  
prai@ncsu.edu

## ABSTRACT

Systematic Literature Reviews (SLR) may not provide insight into the “state of the practice” in SE, as they do not typically include the “grey” (non-published) literature. A Multivocal Literature Review (MLR) is a form of a SLR which includes grey literature in addition to the published (formal) literature. This paper aims at raising the awareness for MLRs in SE.

## 1. INTRODUCTION

Systematic Literature Reviews (SLR) and Systematic Mapping (SM) studies have become common in software engineering (SE) to systematically collect evidence and to structure a given research area, respectively.

While SLR or a SM studies are valuable, other SE researchers have recently reported that “*the results of a SLR or a SM study could provide an established body of knowledge, focusing only on research contributions*” [1]. Since those studies do not include the “grey” literature (non-published, nor peer-reviewed sources of information), they do not provide insight into the “state of the practice” in SE. For a practical (practitioner-oriented) field such as SE, synthesizing and combining both the state-of-the art and practice is very important. Unfortunately, it is a reality that a large majority of software practitioners do not publish in academic forums [2], and this means that the voice of the practitioners would be limited in review studies if we do not consider grey literature in addition to academic. SLRs and SMs use as input only academic peer-reviewed articles, MLRs in addition also use sources from the grey literature, e.g., blogs, white papers and web-pages.

## 2. RESEARCH QUESTIONS & METHODOLOGIES

### 2.1.1 Research Questions

The goal of this study is to raise the need for (more) MLRs in SE. Based on the above goal, we raise the following two research questions (RQs):

- **RQ 1-** What types of knowledge (opportunities) are missed when a SLR does not include the multivocal literature in a SE field?
- **RQ 2-** What do we, as the SE community, gain when we include the multivocal literature in review studies and conduct MLRs?

### 2.1.2 Methodology for RQ1

To address RQ1, the methodology was to first select a small subset of SE SLRs which haven’t included grey literature, then search for grey literature in their focus areas, to find the types of missing information available via the grey literature sources that were not included in those SLRs.

Table 1- The three SLRs sampled (chosen) to answer RQ1

| Year | Topic  |
|------|--|
| 2013 | An SM on Graphical User Interface (GUI) testing [1]                  |
| 2015 | An SLR on using metrics in agile and lean software Development [2]   |
| 2015 | An SLR on definitions, precedents and outcomes of technical debt [3] |

### 2.1.3 Methodology for RQ2

From the small set of MLRs in SE, we sampled three MLRs to answer RQ2, as shown in Table 2. We also show the number of literature entries (formal versus grey) and the ratio of the grey sources for each MLR.

Table 2- The three MLRs sampled to answer RQ2

| Year | Topic  | Num. of lit. entries<br>F, G* (% of grey) |
|------|--|---|
| 2015 | A MLR on deciding when and what to automate in testing (ManAutoTest) | 26, 52 (66%)                              |
| 2015 | A MLR on test maturity and test process improvement (TM/TPI)         | 130, 51 (28%)                             |

\* F: Formal, G: Grey

To answer RQ 2 (assessing what the community would gain when we include multivocal literature), we reviewed, for each MLR, the types of contributions and evidence that were utilized from the grey literature to answer the RQs of each MLR.

## 3. RESULTS

### 3.1 RQ1: Knowledge missed when a SLR does not include multivocal literature

#### 3.1.1 Summary and meta-analysis of the three cases

An important area missed in our SLRs focusing on UI testing, Agile metrics, and Technical debt are available tools and their features. If industry is going to follow academic advice on these areas, it will need tools to do it efficiently and effectively. Thus, an area we suggest that academic SLRs should include to ease industrial adaptation is to look at the current and available tools that industry is using. Additionally, when our advice and tools match with the state-of-the-practice tool, it will ease the technology transfer from academia to industry.

Additionally, we would like to make a connection to five levels of ‘closeness’ between academia and industry as presented by Wohlin: Level 1: Not in Touch, Level 2: Hearsay, Level 3: Sales Pitch, Level 4: Offline, and Level 5: One Team. When the multivocal literature is not included in SLRs, the synthesis is conducted ‘closed’ and the results would not be very beneficial to practitioners, since the SLR contents will be mostly in the level 1 (not in touch) or at most in level 2 (hearsay), as per Figure 1. When grey literature is included, then the closeness can be characterized as Level 3 (Sales Pitch) or Level 4 (Offline).

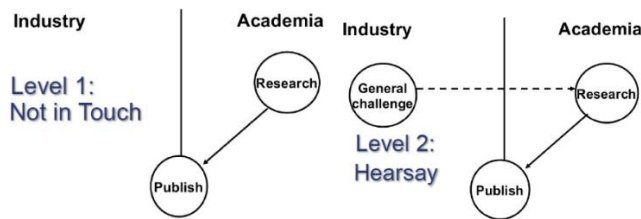


Figure 1-The first two levels of closeness between academia and industry, proposed by Wohlin

## 3.2 RQ 2: What the community would gain when we include multivocal literature

### 3.2.1 MLR on ManAutoTest

If we were to exclude the grey sources from the pool for this MLR, we would simply miss a major pile of experience and knowledge from practicing test engineers on the topic. To put this in quantitative terms, we partitioned the synthesis of a major output of that MLR by the type of source where they were mentioned in: either formal or grey literature, as shown in Figure 2.

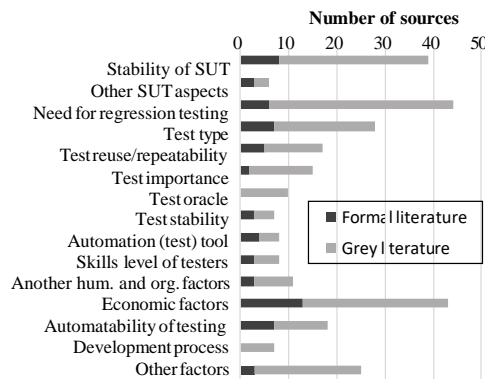


Figure 2-A major output of the MLR on ManAutoTest

### 3.2.2 MLR on TM/TPI

We would have missed information from practice on test maturity and test process improvement, if we were to exclude the grey sources. Figure 3 shows as a major output of the MLR the number of papers per TM/TPI model using or extending a source model. Without grey literature, the usage of TMap, several qualitative statements on drivers, impediments, objectives, and benefits of TM/TPI and some other models would not have been considered.

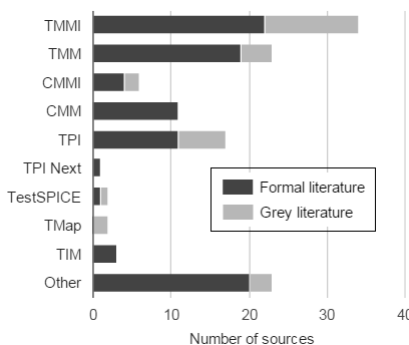


Figure 3-A major output of the MLR on TM/TPI

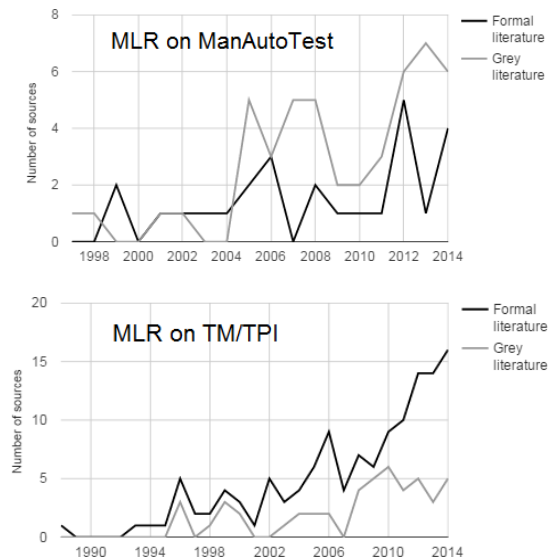
### 3.2.3 Summary and synthesis of the three cases

By synthesis of the above three cases, we discuss the answer to RQ2 in these aspects: (1) growth of the interest by academia versus industry to different SE topics, (2) usefulness of industry viewpoints, and (3) quality of evidence in grey literature

To assess the growth of the interest by academia versus industry for two areas of ManAutoTest and TM/TPI, Figure 4 shows the trends of the annual number of sources (formal versus grey literature) for their

corresponding two MLRs. As we can see, in the case of ManAutoTest, in terms of the level of interest, the grey literature has somewhat passed the formal literature.

Figure 4-Trends of the annual number of sources (formal versus grey literature) for two MLRs



Also, quite interestingly, the trends in Figure 4 resemble quite nicely to an abstract visualization of the relationship of “Theory versus practice” from the “Software Creativity 2.0” book by Glass and DeMarco [4], as shown in Figure 5. This seems to denote that sometimes the industry takes over academia in a certain field and then academic ‘catches up’ and vice versa. Thus, close linkage between the two ‘camps’ is indeed important and conducting MLRs is a good constructive effort in that direction.

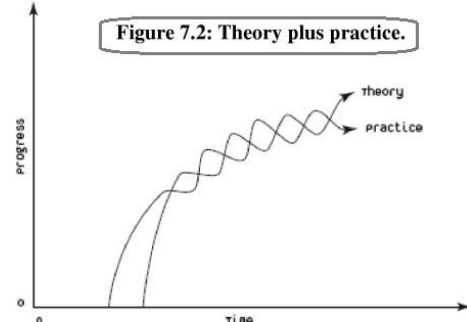


Figure 5-A figure adopted from the “Software Creativity 2.0” book by Glass and DeMarco [4]

## 4. CONCLUSIONS

While many researchers conduct empirical studies in the form of opinion surveys and semi-structured interviews to gather voice of practitioners, we believe that a vast knowledge base of public is already available online and can be, quite conveniently, analyzed and synthesized by SE researchers without investing major time and effort in conducting opinion surveys, as we conducted for the two MLRs (ManAutoTest and TM/TPI).

## 5. REFERENCES

- [1] I. Banerjee, B. Nguyen, V. Garousi, and A. Memon, “Graphical User Interface (GUI) Testing: Systematic Mapping and Repository,” *Information and Software Technology*, vol. 55, pp. 1679–1694, 2013.
- [2] E. Kupiainen, M. V. Mäntylä, and J. Itkonen, “Using metrics in Agile and Lean Software Development – A systematic literature review of industrial studies,” *Information and Software Technology*, vol. 62, pp. 143–163, 6// 2015.
- [3] E. Tom, A. Aurum, and R. Vidgen, “A Consolidated Understanding of Technical debt,” in *European Conference on Information Systems*, 2012, p. 16.