Agile Testing: Past, Present, and Future

Charting a Systematic Map of Testing in Agile Software Development

Theodore D. Hellmann, Abhishek Sharma, Jennifer Ferreira, Frank Maurer
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
{tdhellma, absharma, jen.ferreira, frank.maurer}@ucalgary.ca

***Abstract*—Testing has been a cornerstone of agile software development methodologies since early in the history of the field. However, the terminology used to describe the field – as well as the evidence in existing literature – is largely inconsistent. In order to better structure our understanding of the field and to guide future work, we conducted a systematic mapping of agile testing. We investigated five research questions: which authors are most active in agile testing; what is agile testing used for; what types of paper tend to be published in this field; how do practitioners and academics contribute to research in this field; and what tools are used to conduct agile testing? Of particular interest is our investigation into the source of these publications, which indicates that academics and practitioners focus on different types of publication and, disturbingly, that the number of practitioner papers in the sources we searched is strongly down since 2010.**

Keywords-agile software development, software testing, systematic mapping, empirical, test-driven development, testing tools.

I. INTRODUCTION

Systematic mapping studies have been gaining popularity recently in the fields of computer science and software engineering [1]. These studies set out to structure a field in order to gain deeper understanding of what has already been done and identify gaps in the existing literature where further research is needed. Because of this, these studies are useful for quickly orienting oneself to a field of study.

Testing is one of the cornerstones of agile software development and has received attention since early in the history of agile software engineering. Many agile practices rely on effective software testing [2]. This means that, beyond implications for the quality of the code and the effort that will be required to fix bugs post-release, the effectiveness of testing in an agile context can determine the outcome of an agile software development effort.

We chose to perform a systematic mapping of the field of agile testing given that meta-analyses of agile testing practices, like test-driven development, found that evidence for the effectiveness of these practices is conflicting and varies widely [3,4]. The goal of this paper, then, is to structure the body of literature on agile testing practices in order to better inform future work in this field. Specifically, we hope that this paper will enhance the focus and relevance of future studies in the area of agile testing such that future work will be better able to investigate the benefits and drawbacks of various agile testing practices.

The terminology used to describe this field is inconsistent. For instance, some terms have grown so closely associated with agile testing that they are taken for granted. Many papers do not specify that their agile testing is automated or that their test suites are used to detect regression errors because this already seems to be implied by the tests that are acceptable within an agile software development context. Because of this, we focus on terms that are not yet synonymous with agile testing.

In order to address our research questions, we extracted keywords describing research into agile testing such as the purposes that these tests serve, the type of evidence used to support assertions about agile testing, details about authors and year of publication, contexts in which agile testing is performed, and the tools used to conduct agile testing.

Of particular interest are the findings presented in subsection D of our Results section. These findings underscore low levels of publications by practitioners in academic venues – particularly beyond the context of experience reports. We believe that this is due in large part to the reality that practitioners tend to publish in other venues – in blog posts, in forums, in white papers, in books. This creates the possibility of a disconnect between the understanding of agile testing in academic and industrial contexts. In that section, we point out where gaps in existing literature already exist so that future work can focus on addressing the disconnect.

Section II orients our study to existing meta-analyses of agile testing and to systematic studies in software engineering. We then describe the strategy we used to conduct this study in Section III. Before presenting our results in Section V, we provide an overview in Section IV. The limitations of our study are discussed in Section VI. Promising directions for future work are presented in Section VII before our concluding remarks in Section VIII.

II. BACKGROUND

Meta-analyses, such as systematic literature reviews and systematic mappings, have been carried out to investigate the state of the art for agile software development and related fields, such as test-driven development (TDD). Dybå and Dingsøyr [5] present a systematic review of empirical studies of agile software development and present implications for

research as well as industry. Their review indicated a need for more empirical studies that focus on methods other than eXtreme Programming (XP). However, their review explicitly did not focus on agile testing.

In the meta-analysis by Shull et al. [4], the effectiveness of TDD is examined in terms of delivered quality, internal code quality, and productivity. However, the results reported in these meta-analyses are mixed. Reviewing the evidence for adoption of TDD, Causevic et al. [6] found the variation in reported results problematic for comparing between studies. This was also a found by Dybå and Dingsøyr [5], who identified large variation in how empirical studies were conducted and reported. Jeffries and Melnik [3] conducted a review of selected empirical studies on the effect of TDD on quality and effort. They found that TDD largely resulted in an increase in quality, but one study they identified showed instead that TDD resulted in a strong negative impact on quality. Additionally, while they showed that TDD could reduce the amount of effort required by up to 27%, most studies found an increase in effort of up to 100%.

A major consideration with meta-analyses, such as the ones referenced above, are issues of repeatability. Kitchenham et al. [7] carried out a case study to investigate the extent to which systematic literature reviews are repeatable. They found that the experience level of researchers impacted whether comparable sets of studies are selected for review, which subsequently impacted conclusions drawn by the researchers. For these reasons, it becomes imperative that methods used in systematic literature reviews and systematic mapping studies are clearly and transparently presented [7].

Following the advice by Kitchenham et al. [7], we closely followed the methods by Petersen et al. [8] and Dybå and Dingsøyr [5] in order to maintain transparency in our methods.

III. METHODOLOGY

This study is based on the guidelines provided by Petersen et al. [8]. As such, this section provides details on how we conducted each step of their process. Each step in our approach involved at least two, and usually three, of the authors. This was so that disagreements could be resolved through discussion.

A. Define Research Questions

The first step in this systematic mapping was to select a set of research questions to answer. These questions are used to define the focus and restrict the scope of the investigation. The research questions that drove this research were:

1. *Which authors have been most active in publishing research on agile testing?*
2. *What has agile testing been used for?*
3. *What types of papers on agile testing have been published?*
4. *Has research into agile testing been driven by industry, academia, or collaborations between the two groups?*
5. *What tools have been used for agile testing?*

B. Conduct Search

The second step in our research was to conduct a search for publications relevant to our research questions. To do this, we created a search string to use on popular research databases. We searched for the terms “agile” and “test” in the titles, keywords, and abstracts of papers and limited our search results to the domains of Computer Science and Engineering. While this search string is quite general, we felt that it was more important to make sure that as many relevant papers as possible were included than it was to try to restrict our result set to a reasonable number. We did not search for specific development methodologies (like XP) or practices (like TDD), so it is possible that we missed papers that were in fact agile, but did not use that word. Additionally, it is possible that publications to agile conferences did not use that word in their descriptions given that it was in the venue name. Because of this, future studies could extend this mapping into subfields of agile, include testing practices that are implicitly agile as search terms, and could manually search agile conference proceedings.

We executed this search string on the popular databases SciVerse Scopus[1] and IEEE Xplore[2]. Xplore was a natural choice in that the IEEE hosts many conferences related to computer science and software engineering. Scopus, on the other hand, is an important source because it indexes the results of other research databases (for example, the ACM Digital Library and Springerlink)[3]. As such, Scopus searches over 46 million entries[4] (compare with 3.1 million for Xplore[5], which is not indexed by Scopus). The results of this initial search are presented in Figure 1.

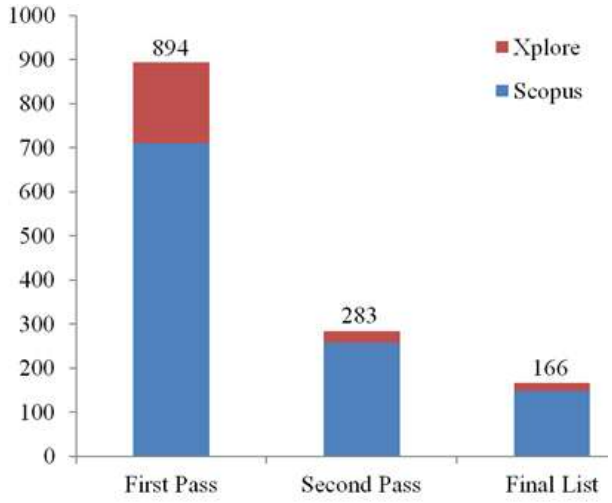

Figure 1: Papers remaining after each screening step.

[1] www.scopus.com
[2] ieeexplore.ieee.org
[3] http://www.info.sciverse.com/documents/files/scopus-training/resourcelibrary/xls/Publisherlist.xls
[4] http://www.info.sciverse.com/scopus/scopus-in-detail/facts
[5] http://ieeexplore.ieee.org

C. Screening Papers

After collecting a set of relevant papers, the next step was to screen out those publications not directly relevant to our research questions. This task was performed iteratively as the depth at which we looked at each paper increased. In the first phase, we excluded papers based only on their title while in the second pass we excluded papers based on a reading of their abstracts. The third pass occurred during the keywording process. The results of this process are shown in Figure 1, and the final paperset is available online at www.ase.cpsc.ucalgary.ca/~tdhellma.

It is worth noting that we ended up excluding a large number of papers in the domains of engineering test equipment, agile manufacturing techniques, or agile management styles. In the future, it may be worth specifically excluding papers including the following terms in our search strings: manufacturing, airplane, wavelet, waveform, spaceship, radar, sensor, missile, laser. It would also have helped to automatically exclude any paper published before the establishment of agile as a software development methodology (which we took to mean before 1997), although our final result set contained results from between 2003 and 2011 anyway.

D. Keywording Using Abstracts

After we finished reading each paper's title and then abstract to exclude unrelated papers, we performed keywording on each remaining paper's abstract in order to develop a framework for understanding the field of agile testing. Keywording is analogous to the process of open coding used in grounded theory [9] in that it is used to build a classification schema out of a large amount of freeform text data. In our context, keywording was performed identifying the key words in abstracts that describe the paper in terms of our research questions.

We used an open coding process to answer our second, fifth, and, to an extent, our third research questions. In open coding, key terms are drawn out of the source material in an emergent fashion. Open coding is good for gaining an understanding of data when it is unclear what to expect. On the other hand, we used closed coding to answer our fourth, fifth, and, to an extent, our third research questions. In closed coding, a list of key terms is determined beforehand and then assigned to papers. Closed coding is good for quickly coding a dataset when the possible codes are well-understood – as was the case with these fields.

The codes we used to describe papers for the fourth question were: industry, academia, and both. Papers were coded as "industry" when all authors were affiliated with companies at the time of publication; similarly, papers were coded as "academia" when all authors were affiliated with universities at the time of publication. When papers were written by a mix of industry and academic authors or when the authors were affiliated with a group in the grey area between academic and industry (for example, a government research institution), the paper was coded as "both."

We keyworded the third research question by performing both closed and open coding. Open coding was used to describe the type of *study* conducted in academic publications whereas closed coding was used to describe the type of *paper*. We used the paper types identified in [10]. Even though these categories were originally envisioned for product lines, they fit well for the types of paper identified in this study.

E. Data Extraction and Mapping

We used Endnote (www.endnote.com) for the actual keywording of our paperset because it makes grouping of keywords and analysis of the resulting categories easier. This allowed us to collect the hard numbers needed for creating the visualizations used in this paper. We then generated these visualizations using Excel (office.microsoft.com/excel).

IV. Overview of Results

From a very high level, our results allow understanding of the number of publications per year and the important terms associated with this field, as shown in Figures 2 and 3.

The number of papers published by each source

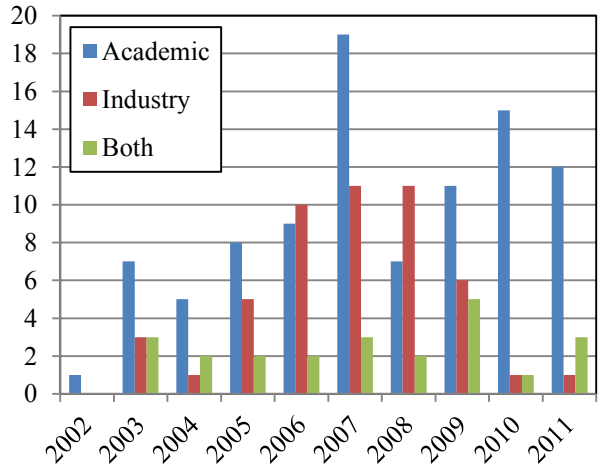

Figure 2 : Number of publications by source by publication year.

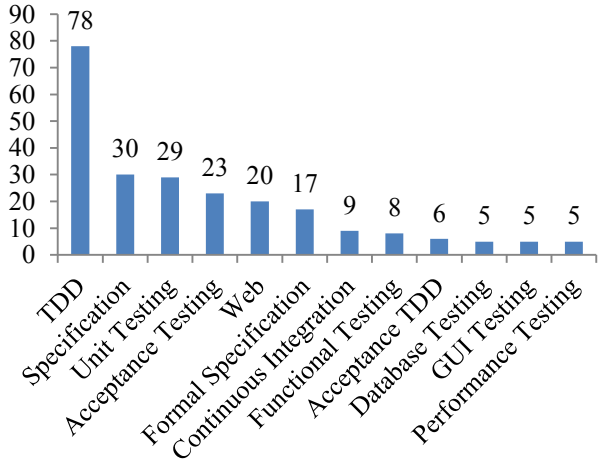

Figure 3: Frequency of type of testing.

(industry, academia, both) for each year in our paperset is shown in Figure 2. From this chart, we can see that, overall, academics have published roughly twice as many papers as have industry practitioners and that these practitioners have published twice as many papers as have been published through academia-industry collaborations. However, twice – once in 2006 and again in 2008 – practitioners published more work on agile testing than did academics. More recently, in 2010 and 2011, practitioners published very few papers. So, the question is: what caused practitioners to stop publishing papers on agile testing?

One possible reason is that the databases we searched primarily index peer-reviewed work. Peer-reviewed venues favour rigorous studies – usually the domain of academics. Practitioner publications tend to focus on sharing insights in a readable manner over rigorous study design. Given that there are few incentives for practitioners to invest such a high amount of effort in publications, it makes sense that our study returned fewer practitioner papers. From this perspective, it is possible that an increase in the rigor of the peer review process in the past two years is responsible for the near-absence of practitioner publications.

At a high level, it is also possible to get an overview of the important terms associated with the field of agile testing. Figure 3 shows the frequency of the occurrence of keywords relating to the type or purpose of testing identified in our final paperset. From this chart, we can clearly see that the majority of attention in agile testing has focused on test-driven development. Use of tests as specification or formal specification, for unit or acceptance testing, or to test web applications was also common. Much less common, however, was use of tests as part of continuous integration, acceptance test-driven development, database or GUI testing, or performance testing. Keywords with fewer than 5 occurrences are not shown.

V. Results

As a result of our keywording process we were able to generate visualizations to answer our research questions. These results are presented in this section and are organized according to the research questions outlines earlier.

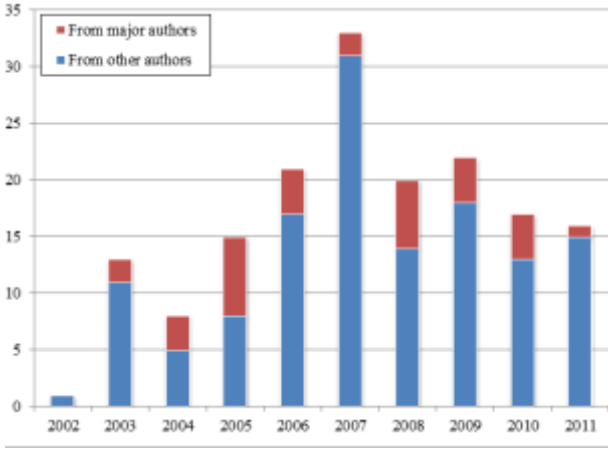

Figure 4: Publications per year by major/other authors.

A. *Which authors have been most active in publishing research on agile testing?*

From the final paperset we were able to find the six authors with the highest number of author credits in the field of agile testing. The top six were picked based on the number of papers that each author in the field had produced, as shown in Figure 5. Given their prominence in the field, we felt that it was important to gain a deeper understanding of these persons. Table 1 shows additional details about each major author: the years in which they have published papers; the types of testing in which they have been interested; and a link to their homepage. We hope that this will foster further contact and collaboration in this field.

Also, we created Figure 4 to compare the level of activity of the major authors we identified against the level of activity of the rest of the authors by showing the fraction of publications each year that can be attributed to them. This was done to understand if the field as a whole is the domain of a few or if the major authors are simply more active than others. Figure 4 helped us to identify two particularly interesting years: 2005 and 2007. In 2005, the field really was driven by the major authors – out of 15 total publications, 7 (nearly half) were from this small set of authors. 2007, on the other hand, saw the most publications total out of any year (33), yet very few of these came from major authors (2). Based on this, it would seem that the major authors can have a strong influence on the field as a whole, but that this influence varies widely and is not linked as strongly to the number of publications in a given year as might be imagined.

B. *What has agile testing been used for?*

Figure 6 was created in order to understand how agile testing has been used over time. Interest in TDD has consistently been high, which indicates its central role in agile testing. However, this interest is by no means overwhelming – there is no year in which TDD was the focus of a majority of publications. In contrast to this stability, interest in some other uses of agile testing shows distinct spikes. For example, there are gaps in the publication record for database, GUI, performance, and acceptance

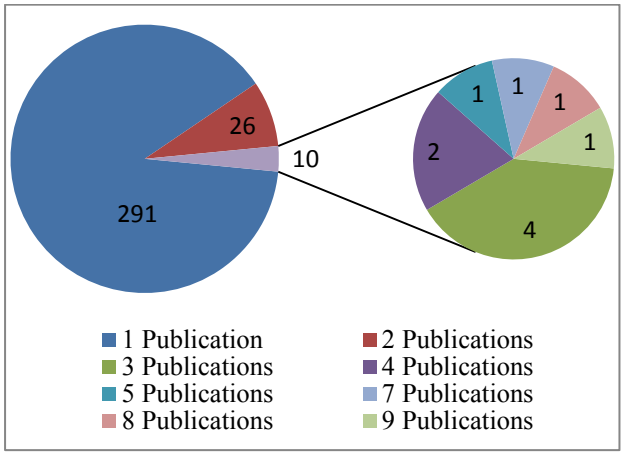

Figure 5: Number of authors with the given number of publications. Note: No authors had 6 publications.

TABLE 1. MAJOR AUTHORS IDENTIFIED BY THIS STUDY.

Author	Active Years	Interests	Link
Geras, A.	**2004, 2005, 2008**	**Manual Testing, Acceptance Testing, Non-Functional Testing, TDD**	http://ca.linkedin.com/pub/adam-geras/5/7a9/3b5
Janzen, D.	**2003, 2005, 2006, 2009**	**TDD**	http://works.bepress.com/djanzen/
Williams, L.	**2003, 2004, 2006, 2007, 2008**	**Performance Testing, TDD**	http://collaboration.csc.ncsu.edu/laurie/index.html
Smith, M.	**2004, 2005, 2008, 2009, 2010**	**Acceptance Testing, TDD, Non-Functional Testing, Unit Testing**	http://www.enel.ucalgary.ca/People/Smith/
Miller, J.	**2004, 2005, 2008, 2009, 2010**	**Acceptance Testing, TDD, Non-Functional Testing, Formal Specification, Unit Testing**	http://www.steam.ualberta.ca/
Maurer, F.	**2004, 2005, 2007, 2008, 2009, 2010, 2011**	**TDD, Acceptance Testing, Performance Testing, Acceptance Test-Driven Development, Specification, GUI, Formal Specification**	http://ase.cpsc.ucalgary.ca/ase/Frank.Maurer.php

testing. This could indicate that subfields within the greater field of agile testing are at least somewhat disjoint. This is reinforced by the fact that an increase in the overall number of publications – for example, 2007 – does not result in an increase in all subfields. Interest in various subfields seems to rise and fall over time. For example, GUI testing was absent from 2006 to 2009 while interest in database testing was only shown during that same period. This is interesting in that it could demonstrate a switching of interest between subfields, though there are too few data points to draw strong conclusions in that regard.

C. *What types of papers on agile testing have been published?*

In [10], five basic types of paper are identified:

- Solution – a novel solution is proposed, but only a proof of concept is offered
- Validation – further investigates a solution, but is not evaluated in practice
- Philosophical – provides a new framework for understanding a field
- Opinion – presents the author's personal opinions
- Experience – describes the author's experience on a project
- Evaluation – investigation of a problem in practice

We used these same paper types to determine what type of publications regarding agile testing have been produced. Figure 7 is a visualization of what type of publications have been produced in the investigation of agile testing along with the year in which they were published. In this field, it's very important to note that opinion and experience papers have all but disappeared in the last two years and only one philosophical paper was published last year. This correlates strongly with the drop in industry participation in this field and is an unhealthy development. These types of publications should be encouraged: philosophical and opinion papers provide new directions for research, while experience papers are important to evaluating how techniques are working in practice. However, there was an increase in the number of solution papers in 2011. Since solution papers are important as a starting point to research into novel techniques, these papers could represent the first wave of research into new agile testing practices. It's worth noting that this spike in solution papers follows a spike in philosophical papers in 2010.

D. *Has research into agile testing been driven by industry, academia, or collaborations between the two groups?*

From Figure 2, we can see that, for most years, academics have contributed the largest portion of publications as indexed by the databases we used. As noted earlier, in 2006 and 2008 practitioners contributed more publications, but practitioner papers have been absent in the past two years. Additionally, publications by collaborations between industrial and academic authors have never made up the majority of publications in this field. Based on this, we feel that one major way of further developing this field in the future would be to find ways of encouraging more academic-practitioner collaborations in which the practitioner-authors are actively involved in both research and paper-writing.

In order to gain further understanding of the source of publications in the field of agile testing, we also investigated the type of publication with respect to its source. The results are shown in Figure . While this visualization is strongly influenced by the overall number of papers published in the field, we can still understand useful information from it. For example, we can see that practitioners have published virtually no evaluation or validation papers. This is somewhat counterintuitive since practitioners are in the best position to conduct this sort of research. However, rather than conducting formal studies, practitioners tend to publish experience papers as evidenced by the fact that a very strong majority of experience reports are from industry. Based on this, another way of enriching this field would be to encourage more practitioner involvement in evaluation research as well as more collaborations across the board.

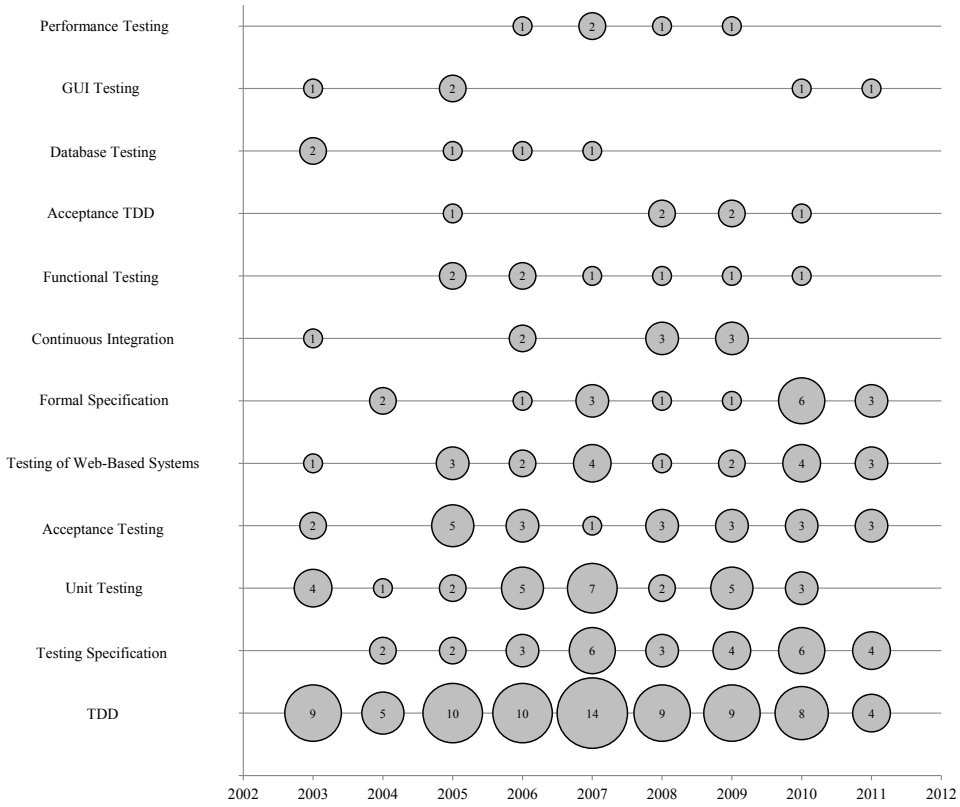

Figure 6: Use of testing over time.

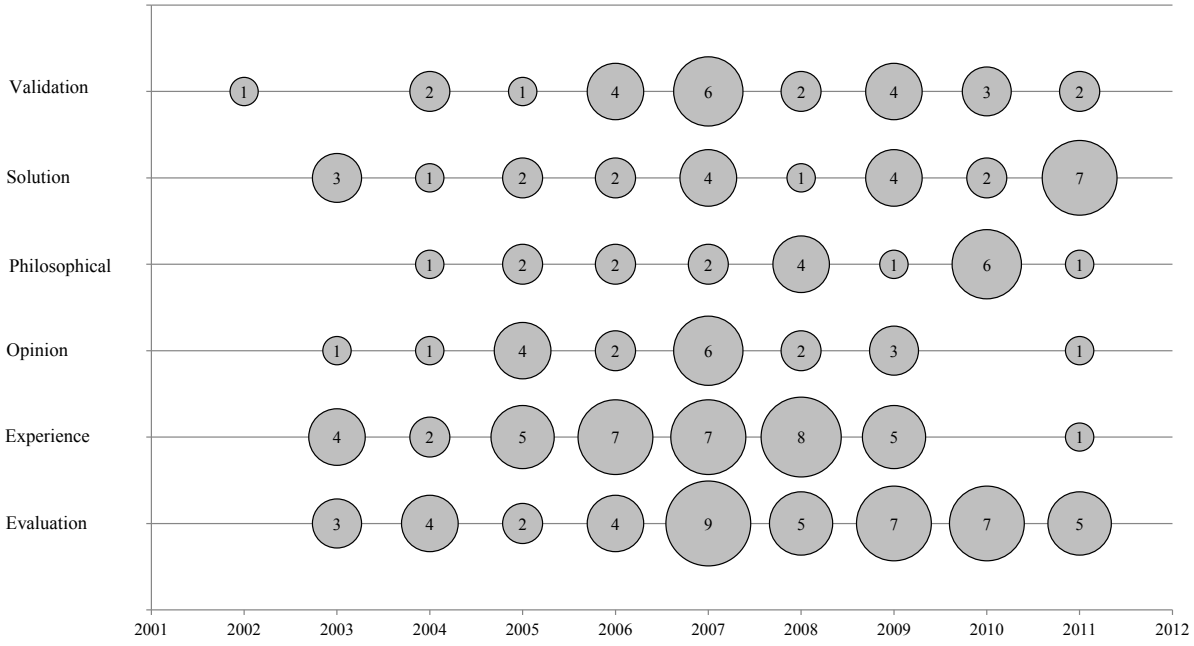

Figure 7 : Type of paper by year of publication.

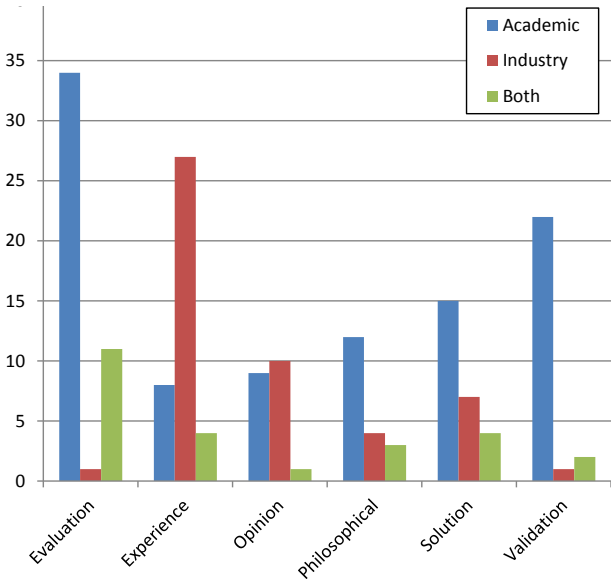

Figure 8: Types of paper by source of paper.

E. What tools have been used for agile testing?

A wide variety of tools were used for a variety of purposes in our paperset, and these tools are shown in Figure . Many of these tools focus on enabling unit and acceptance testing in various circumstances. For example, many of the tools reported are versions of xUnit for different programming languages. Relatively few tools provide support for advanced testing techniques in an agile environment and instead focus on providing support for basic testing techniques. For example, only Jumble provides support for mutation testing and it only works with Java. This indicates a substantial blind spot in the agile testing literature. It is important that future work look into why this is the case, but we speculate that this slow uptake could be due to the heavyweight nature of many advanced testing tools. Perhaps formal testing tools are not suitable for an agile software development environment. If this is the case, it is very important that future work find ways to make these tools more compatible with agile development practices.

VI. LIMITATIONS

As with any meta-analysis, the first limitation of this study is the risk that we failed to identify all relevant peer-reviewed publications on agile testing. We took steps to minimize this risk – for example, we used a high-level search string that returned 728 results that would eventually be eliminated. This approach meant that we ended up manually excluding a large number of papers in subsequent steps. However, it also increased the likelihood of identifying relevant papers in the first place. One way of increasing the number of publications included in this study could be to follow citation chains of papers in order to create a larger web of related literature. However, this represents a very significant amount of work and has been left for future work. This would not increase the amount of practitioner literature included in this study, though. Because of the sources we chose to search for relevant literature, a large amount of practitioner literature was overlooked. Practitioner experiences are seldom captured in peer-reviewed publications, but are often expressed through books, technical papers, blog posts, and the like. In future work, this second body of knowledge should be systematically explored to further increase our knowledge of how agile testing is used in practice.

The second limitation of this study is that we explicitly made the assumption that the abstract of a paper will contain precise information concerning its contents. This assumption allows us to extract the information we present in this paper in a timely manner. However, the abstracts we encountered were not always sufficiently specific or detailed enough to allow us to find all of the information in which we were interested. In cases where sufficient information relevant to our research questions was not available through an abstract, we additionally scanned the paper's introduction and conclusion. This step introduces an imbalance in the amount

Figure 9: Tools mentioned in abstracts in our paperset.

of information considered for each paper and could in fact have led to papers with lower-quality abstracts receiving higher focus in our evaluation. In order to restore this balance, we would need to perform a full systematic literature review. However, this is left for future work.

Third, we noted during our search process that many abstracts use the term "agile" to describe a software development environment or practice. In order to accurately understand the context of agile testing, it is important that authors take care to precisely describe what practices an agile team is using, or what about the way in which a practice is used makes it specifically agile. This is important given the rising popularity of meta-analyses like the present study in that it is difficult for us to talk about agile testing when we cannot be more specific with respect to what we mean when we discuss an agile environment or practice.

VII. Future Work

Through the process of conducting this systematic mapping, we identified a variety of strong opportunities for future work. We identified these opportunities through both our keywording process and the maps we generated.

The first opportunity for future work is the investigation of problems that agile testing practices are expected to address. Related to this, a mapping could also be created showing the expected benefits of different types of agile testing. However, it was not possible for us to investigate this relationship in our systematic mapping due to the large number of abstracts that do not provide this information. To investigate the problems addressed by and benefits associated with agile testing, it would be necessary in future work to conduct a systematic literature review.

Certain questions cannot be answered through a systematic mapping. For example, we cannot make suggestions about the best way to perform agile testing. In order to better understand these sorts of questions, it would be necessary to use systematic literature reviews.

Research questions related to where the centers of agile testing research are located could also be addressed in future work. This study does not seek to identify which countries, universities, or companies are more actively involved in agile testing research and this could be addressed in future systematic mappings.

In this study, we investigated the tools used in agile testing. However, this represents only a subset of the tools that could be relevant for this purpose. A survey of testing tools based on Internet searches, searches of open-source software repositories like CodePlex and SourceForge could be combined with an analysis of their applicability to agile testing situations would help increase our understanding or available tool support and areas where more tool support is needed.

VIII. Conclusion

In this paper we present the results of a systematic mapping study of the field of agile testing. The purpose of this study was to provide guidance for future work in this field. We used five research questions as our axis of investigation into this topic.

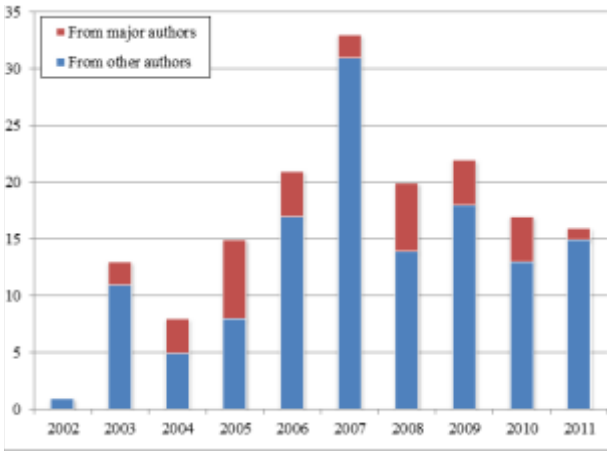

Figure 8: Publications by year by major / other author.

1. *Which authors have been most active in publishing research on agile testing?*

In order to answer this question, we provide created Figure 4 and Figure 5. Table 1 provides the top 6 authors in this field as well as the years in which they published papers and keywords describing their research interests. Further, this table also provides a link to the home page of these authors. It is our hope that this will make it easier for readers interested in performing future research to orient themselves to the relevant literature.

2. *What has agile testing been used for?*

In order to answer this question, we provide a Figure 3 and Figure 6, which provide the top 10 keywords and tracks the top 10 keywords over time, respectively. This was done to provide an orientation to the field's major concepts and also an indication of which topics have been important both historically and recently.

3. *What types of papers on agile testing have been published?*

This research question was designed to provide insight into which types of testing evaluation methods have been used to investigate the field of agile testing. Figure 7 and Figure address this question by mapping the types of paper published both in each individual year and by each source. Figure in particular highlights the difference in the type of papers published by academics and practitioners. In the future, ways of encouraging further participation by industrial authors in the publication of types of paper other than experience reports must be found.

4. *Has research into agile testing been driven by industry, academia, or collaborations between the two groups?*

The purpose of this question was to investigate how academics and practitioners have contributed to the development of the research field of agile testing. While there have been years when industrial authors have contributed more publications in this subject area than their academic counterparts, academics have tended to contribute

more publications in most years and overall. However, what is most striking is the low number of collaborative publications involving authors from both academic and industrial settings. It is our opinion that the number of these collaborations must be increased in the future to advance our knowledge of agile testing.

5. What tools have been used for agile testing?

Unsurprisingly, our analysis of the tools used for agile testing showed a focus towards unit and acceptance testing tools. In the future, research should be done to see if the techniques proposed within the testing community in general can be successfully applied to agile software development efforts. In general, understanding of what makes a tool suitable for use in agile environments would greatly enhance our ability to take advantage of developments in other fields of software engineering.

Finally, particularly in subsection D of the Results section, we have tried to highlight areas in which differences between practitioner and academic literature were apparent. As a result, we can see from Figure that academics tend to focus on evaluation papers where practitioners tend to produce experience reports. However, in the past two years, practitioner participation in this field has been very low. This could be an artifact of our search process, but it could also be due to changes in the types of papers that are being accepted for publication. It is crucial to the future of this field that a strong dialog be established between academics and practitioners researching agile testing.

IX. WORKS CITED

[1] S. MacDonnel, M. Shepperd, B. Kitchenham, and E. Mendes, "How Reliable Are Systematic Reviews in Empirial Software Engineering?," *IEEE Transactions on Software Engineering*, vol. 36, no. 5, pp. 676-687, Sep. - Oct. 2010.

[2] Glenn Vanderburg, "A Simple Model of Agile Software Practices - or - Extreme Programming Annealed," in *Object-Oriented Programming, Systems, Languages, and Applications*, New York, 2005, pp. 539-545.

[3] Ron Jeffries and Grigori Melnik, "Guest Editors' Introduction: TDD - The Art of Fearless Programming," *IEEE Software*, vol. 24, no. 3, pp. 24-30, May-June 2007.

[4] F. Shull et al., "What Do We Know about Test-Driven Development?," *IEEE Software*, vol. 27, no. 6, pp. 16-19, Nov. - Dec. 2010.

[5] Tore Dybå and Torgeir Dingsøyr, "Empirical Studies of Agile Software Development: A Systematic Review," *Information and Software Technology*, vol. 50, no. 9-10, pp. 833-859 , January 2008.

[6] Adnan Causevic, Daniel Sundmark, and Sasikumar Punnekkat, "Factors Limiting Industrial Adoption of Test Driven Development: A Systematic Review," in *International Conference on Software Testing, Verification, and Validation*, Berlin, 2011, pp. 337-346.

[7] Barbara Kitchenham, Pearl Brereton, Li Zhi, David Budgen, and Andrew Burn, "Repeatability of Systematic Literature Reviews," in *Evaluation & Assessment in Software Engineering*, Durham, UK, 2011, pp. 46-55.

[8] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson, "Systematic Mapping Studies in Software Engineering," in *12th International Conference on Evaluation and Assessment in Software Engineering*, Bari, Italy, 2008, pp. 71-80.

[9] Barney G. Glaser and Anselm L. Strauss, *The Discovery of Grounded Theory*. Chicago, United States of America: Aldine, 1967.

[10] Roel Wieringa, Neil Maiden, Nancy Mead, and Colette Rolland, "Requirements Engineering Paper Classification and Evaluation Criteria: A Proposal and a Discussion," *Journal of Requirements Engineering*, vol. 1, no. 11, pp. 102-107, 2005.