Benefits from Modeling and MDD Adoption: Expectations and Achievements

Marco Torchiano
Politecnico di Torino
Torino, Italy
marco.torchiano@polito.it

Federico Tomassetti
Politecnico di Torino
Torino, Italy
federico.tomassetti@polito.it

Filippo Ricca
DIBRIS, Università di Genova
Genova, Italy
filippo.ricca@disi.unige.it

Alessandro Tiso
DIBRIS, Università di Genova
Genova, Italy
alessandro.tiso@disi.unige.it

Gianna Reggio
DIBRIS, Università di Genova
Genova, Italy
gianna.reggio@disi.unige.it

ABSTRACT

The adoption of Model Driven Development (MDD) promises, in the view of pundits, several benefits. This work, based on the data collected through an opinion survey with 155 Italian IT professionals, aims at performing a reality check and answering three questions: (i) Which benefits are really expected by users of modeling and MDD? (ii) How expectations and achievements differ? (iii) Which is the role of modeling experience on the ability of correctly forecasting the attainable benefits?

Results include the identification of clusters of benefits commonly expected to be achieved together, the calculation of the rate of actual achievement of each expected benefit (varying dramatically depending on the benefit) and the "proof" that experience plays a very marginal role on the ability of predicting the actual benefits of these approaches.

Categories and Subject Descriptors

D.2.10 [**Software Engineering**]: Design—*methodologies*

General Terms

Measurement, Languages

Keywords

Industrial survey, Model Driven Development (MDD), state of the practice, benefits

1. INTRODUCTION

Models are used in software development with the general goal of raising the level of abstraction. The approaches resorting on models are various and fall under different names: from simple modeling to model-driven development (MDD) [16], model-driven engineering (MDE) [17], and model-driven architecture (MDA) [13]. In practice, models can be transformed and code can be generated from them by means of (semi) automatic transformations. Alternatively, models can also be directly executed/interpreted (in that case they are called executable models). In the following, we will address all these related techniques with the abbreviation MD* [22].

There is a number of benefits commonly associated with the usage of models: they range from an improvement in the quality of documentation, to huge gains in productivity and reduction of defects [1]. Hype is frequently associated to software development processes/techniques until they are not yet mainstream and fully understood [4]; we think it is also the case for modeling and MD*. In our opinion it is important to distinguish which benefits associated with modeling and MD* are real and which contribute just to create hype.

The literature reports different success stories about MD* (e.g., [1, 8]). Those stories tell us which benefits we can get in the best-case scenario. What about the other cases? How frequently are the failures? How many practitioners were disappointed with MD* usage? How frequently the promises of MD* are not maintained in reality? We think it is important to answer these questions to provide guidance to practitioners and clarify what can be reasonably expected from modeling and MD* and what can possibly, but not so easily, be obtained.

The large number of methods under the MD* name is considered to be still evolving and not yet completely mature. The first success stories were heard a long time ago but the knowledge to make those successes consistently repeatable is still missing. Being the discipline not yet fully understood, and the underlying knowledge not yet codified, expertise is the only resource we can rely on when a MD* solution is designed. Thus, another interesting element to investigate is the role of expertise. Being expertise difficult and controversial to measure directly, we can use the number of years of experience as a proxy. The resulting question is: does the level of experience help when adopting modeling and MD*? In particular, does it help when forecasting the outcome of modeling?

In the next section, we present the design of the general survey, the research questions addressed in this work and the analysis we performed to answer them (Sect. 2). Then, we present the results found (Sect. 3). In Sect. 4, we discuss the results and later we compare them with previous work (Sect. 5). Finally, we draw our conclusions (Sect. 6).

2. SURVEY DESIGN

We conceived and designed the study with the goals of understanding:

G1 the actual diffusion of software modeling and MD* in the Italian industry,

G2 the way software modeling and MD* are applied (i.e., which processes, languages and tools are used), and

G3 the motivations either leading to the adoption (expected benefits) or preventing it (experienced or perceived problems).

The above goals cover a wide spectrum, which base been partly covered in previous works [19, 21]. The cited articles provide also more details about the design of the survey. In this work, we consider only a limited portion of those goals, in particular we focus on the benefits, that is the first part of goal G3.

2.1 Research questions

The goal we investigate in this paper, i.e., examine expectations and real achievements of benefits due to the adoption of modeling and MD*, can be detailed into three main research questions. First of all, we consider what benefits the adopters expect from modeling (RQ1), then we examine what is the actual frequency of achievements (RQ2). Finally, we consider if the experience can lead to more realistic expectations (RQ3).

- **RQ1: Which are the benefits expected from modeling and MD* adoption?**
  - **RQ1.1: Which are the most expected benefits?** We want to understand which are the anticipated benefits that also represent plausible motivations for adopting modeling and MD*.
  - **RQ1.2: Which are the relations between expectations?** We envision group of related benefits, i.e., benefits that are supposed to be achieved together.
- **RQ2: Which are the most frequently fulfilled expectations?** We aim at understanding how well confirmed benefits match expectations, in order to understand the capability of participants to predict the results and spot possibly hard-to-gain benefits.
- **RQ3: Does experience in modeling improves accuracy of benefits achievement forecasts?** Correctly forecasting achievable benefits is a key factor, e.g., in cost estimation, therefore we are aim to understand whether (or not) experience improves (or affects) the performance in this respect.

2.2 Instrument

We selected an opinion survey [6] with IT practitioners, administered through a web interface, as instrument to take a snapshot of the state of the practice concerning industrial MD* adoption. In the design phase of the survey we drew inspiration from previous surveys (i.e., [20] and [9]) and we followed as much as possible the guidelines provided in [12].

The survey has been conducted through the following six steps [12]: (1) setting the objectives or goals, (2) transforming the goals into research questions, (3) questionnaire design, (4) sampling and evaluation of the questionnaire by means of pilot executions, (5) survey execution and, (6) analysis of results and packaging.

For the specific purpose of this paper we analysed a few items contained in the questionnaire (a more detailed description is available in [19]).

An initial item (*Dev08*) concerned whether models are used in the organization for software development. For the respondents who provided a positive answer to such item we administered a further item measured using the question "*What are the benefits expected and verified from using modeling (and MD*)?*". This was designed as a closed option question; the list of benefit that we presented the respondents was compiled on the basis of the literature and includes:

- Design support
- Improved documentation
- Improved development flexibility
- Improved productivity
- Quality of the software
- Maintenance support
- Platform independence
- Standardization
- Shorter reaction time to changes

For each benefit the respondent could indicate whether the benefit was expected and/or verified.

To evaluate the experience in modeling we considered one item that measured the years since the initial adoption of modeling or MD* in the work-group.

2.3 Analysis

Whenever possible we addressed the research questions with the support of statistical tests. In all the tests we used we considered a $\alpha = 0.05$ as a threshold for statistical significance, that is we accept a 5% probability of committing a type I error.

RQ1.1: to answer this RQ we simply ranked the benefits by the number of respondents expecting that benefit in descending order. In addition, using the proportion test, we compute the estimate proportion of respondents who expect the benefit and the corresponding 95% confidence interval.

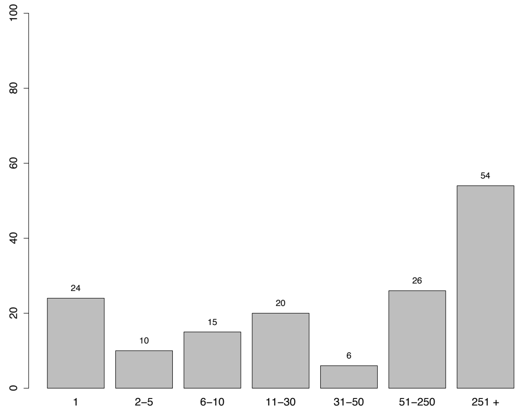

Figure 1: Size of respondents' companies

The interval is useful to understand the precision of the result.

RQ1.2: we looked at the relations between all possible pairs of benefits. We calculated the Kendall rank correlation coefficient between the expectations of each pair of benefits, obtaining a symmetrical measure of the strength of association between the expectations of the two benefits. Positive values represent a positive association while negative values represent a negative association. The absolute value of the correlation represent the strength of the association and it can vary from zero to one.

RQ2: to answer this question we examined for each benefit how frequently it was achieved when expected. We can look at the issue as a classification problem – expected benefits correspond to predictions and verified benefits to observations – then the above measure corresponds to the precision of the classifier.

RQ3: to answer this question we considered the factor experience in modeling, so we divided the respondents in two groups: low experienced practitioners (i.e., < 5 years of experience in modeling) and high experienced practitioners (i.e., ≥ 5 years of experience in modeling). Finally, we built the contingency table and performed the Fisher test considering the two groups (low and high experience) and the number of correct and wrong forecasts done by each group.

3. RESULTS

In summary, over a period of 2 months and half, we collected 155 complete responses to our questionnaire by means of an online survey tool[1].

The most of the companies where the respondents work are in the IT domain (104), then come services (15) and telecommunications (11). The distribution of the companies size where the respondents work is presented in Figure 1.

Among the respondents, on the basis of item *Dev08* we were able to identify 105 respondents using modeling and/or MD* techniques. We apply the analysis described above only to the information collected from respondents who adopted modeling.

[1]LimeSurvey: http://www.limesurvey.org

3.1 RQ1: Which are the benefits expected from modeling adoption?

RQ1.1: Which are the most expected benefits? In table 1 we report for each benefit the frequency of expectation (column *Freq.*) and the corresponding percentage of respondents (column *Estimate*).

Improved documentation is the most expected benefit, with almost 4 out of 5 respondents anticipating it. Also *Design support*, *Quality of the software*, *Maintenance support*, and *Standardization* are frequently expected. For all of the top 5 benefits we are 95% sure that more than 50% of modeling adopters expect them: in fact the confidence interval (C.I.) lower bounds are larger than 50%. The remaining benefits, *Improved development flexibility*, *Improved productivity*, *Shortened reaction time to changes*, and *Platform independence* are less popular, with the latter typically expected by less than 40% of respondents.

RQ1.2: Which are the relations between expectations? We report the statistically significant relations among benefits in the graph shown in Figure 2: the nodes represent the individual benefits, the edges represent a statistically significant relation, with the length of the edge inversely proportional to the Kendall rank correlation coefficient (KC), which is reported as edge label. The benefit expected together ($KC > 0$) are linked by continuous black lines, while the benefits whose expectations tend to exclude each other ($KC < 0$) are linked by dashed red lines, with circles at the ends.

All the significant relations were positive except one, that between *Improved documentation* and *Improved development flexibility*: who expects one of these two benefits tend to not expect the other one.

By observing Figure 2, we can note two distinct clusters: the first includes *Improved documentation*, *Design support* and *Maintenance support*. The second one includes *Improved development flexibility*, *Shorter reaction time to changes*, *Platform independence*, *Standardization* and *Improved productivity*. *Quality of the software* appears to be a transversal benefit, connecting the two clusters.

The two cluster contain three maximal cliques[2]: the smallest (left side) cluster correspond to a three-vertexes maximal clique, while the largest one (right side) correspond to a four-vertexes and a three-vertexes cliques that share a node (*Reactivity to changes*).

3.2 RQ2: Which are the most frequently fulfilled expectations?

This RQ concerns how often the verification of a benefit met the expectation. It is measured as the frequency of verified

[2]From Encyclopedia of educational psychology (Sage Publications): in the social sciences, the word "clique" is used to describe a group of 2 to 12 (averaging 5 or 6) persons who interact with each other more regularly and intensely than others in the same setting.

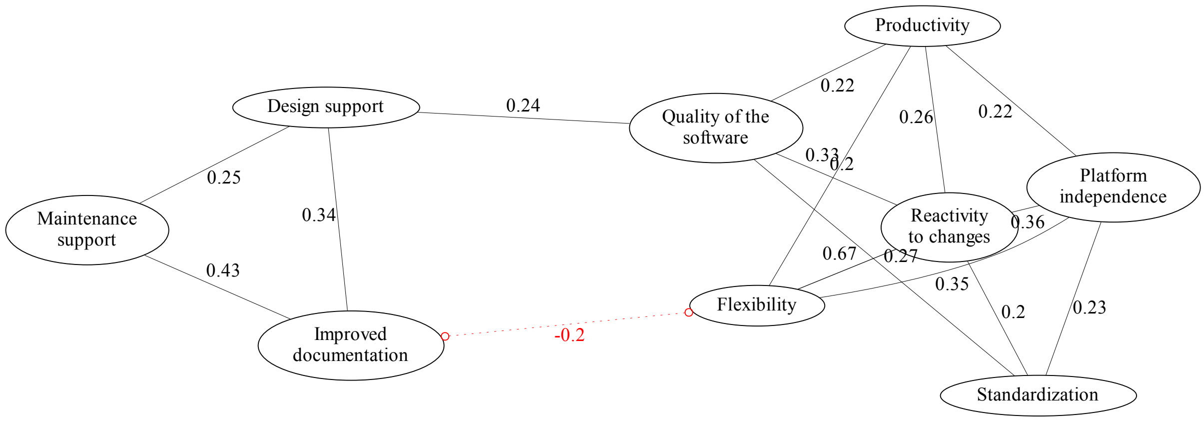

Figure 2: Relations among benefits expectations.

benefit given the benefit was expected. Results are reported in the rightmost column of Table 1 (Fulfillment rate).

Design support has the highest fulfilment rate: 60 respondents out of the 81 who reported to expect it (i.e., 78%) actually achieved the benefit. Also *Documentation improvement* is consistently verified when expected, the same is not true for all the other benefits. *Standardization* and *Maintenance support* are just above the parity (it means are slightly mainly achieved than not achieved, when expected) and all the others are more often not achieved than achieved. *Platform independence* and *Reactivity to changes* have a really low fulfilment rate, representing very often a delusion for practitioners.

3.3 RQ3: Does experience in modeling improves accuracy of benefits achievement forecasts?

The low experienced practitioners group (<5 years of experience in modeling) is constituted by 50 persons, the high experienced practitioners group (i.e., ≥ 5 years of experience in modeling) by 55. Thus, the two groups are balanced.

Applying the Fisher test to the built contingency table, even adopting a loser threshold of 0.1, it is not possible to find any statistically significant difference. Therefore, we conclude that experience does not improve the precision in forecasting the attainable benefits.

4. DISCUSSION

The rate of expectation among benefits varies considerably. The most commonly expected are the benefits deriving from a descriptive use of models (e.g., *Improved documentation* and *Design support*) as opposed to those deriving from a prescriptive use of models (e.g., *Improved productivity* and *Shorter reaction time to changes*). This tells us indirectly how practitioners use models and for what.

It is interesting to note how this distinction between the usage of models in a descriptive or a prescriptive way emerges also from the relation between benefit expectations, where two distinct clusters are clearly depicted (Figure 2). These strong relations between benefit expectations suggest us that practitioners are trying to achieve a set of different benefits at the same time. It remains to understand how often those benefits are contrasting and how difficult is to devise MD* approaches able to permit the achievements of all those benefits at the same time.

The strongest relation is between *Improved development flexibility* and *Shorter reaction time to changes* (KC = 0.67), the intensity of this relation is so strong that we can deduce the two benefits are either essentially considered synonyms

Table 1: Frequency of expectations

Benefit	Freq.	Proportion Estimate	Proportion 95% C.I.	Fulfillment Rate
Improved documentation	81	77%	(68% , 85%)	68%
Design support	77	73%	(64% , 81%)	78%
Quality of the software	75	71%	(62% , 80%)	49%
Maintenance support	66	63%	(53% , 72%)	52%
Standardization	64	61%	(51% , 70%)	52%
Improved development flexibility	51	49%	(39% , 58%)	45%
Improved productivity	42	40%	(31% , 50%)	45%
Shorter reaction time to changes	41	39%	(30% , 49%)	37%
Platform independence	32	30%	(22% , 40%)	34%

or they are intimately related. The next strongest relation (KC = 0.43) is between *Improved documentation* and *Maintenance support*, this link seems to implicitly confirm the common wisdom about documentation being an enabler of maintenance activities.

The rate of achievement is constantly higher than 50% for benefits of descriptive models while it is much lower for benefits of prescriptive models. In the latter case, the rate of achievement can be as low as one out of three for *Platform independence* and slightly higher for *Reactivity to changes* and *Improved flexibility*. A few pragmatic questions arise from the perspective of a project manager, that deserve further investigation:

- is it reasonable to expect those less fulfilled benefits from the adoption of modeling and MD*?
- what are the possible causes of low fulfilment rate for those benefits?
  - limited experience in modeling,
  - lack or inadequacy of tools,
  - simply not obtainable through MD* approaches.

In Table 2 we show, side by side, the position of each benefit among the rank of the most expected benefits (Table 1, 2nd column) and the rank of the most reliably predictable benefits (Table 1, last column). As can be seen, the two rankings are very similar, with most expected benefits being also the most reliably predictable, and the least expected being also the least reliably predictable.

The only relevant difference involves *Quality of the software* and *Standardization*. The former is the 3^{rd} more frequently expected benefit but it proved to be not so easily attainable, while the relation is inverted for the latter. Therefore we can say that concerning the improvements of the software quality through the usage of modeling there are greater expectations than it is realistic, while the benefits in terms of standardization are generally underestimated.

Finally, the lack of effect of experience on the ability of predicting the outcome could be due to the immaturity of model-driven techniques, which are still evolving. Is it possible that developers who have more experience rely on assumptions which were valid for old-fashioned model-driven approaches and are not more valid with the most recent ones.

5. RELATED WORK

In the literature is possible to find anecdotal reports of inflated expectations on software by stakeholders [2]. High expectations and consequent disillusion were reported also for other highly-hyped approaches, as for example for agile methods [5]. We believe this is true also in the SOA context [14].

The effects of expertise on forecast of the outcome were proved to be at the best uncertain in different domains. Camerer and Johnson state that in many domains expert judgments is worse than the simplest statistical models [3].

Table 2: Comparison between expectations and rate of achievement.

Benefit	Exp.	Rate ach.
Improved documentation	1°	2°
Design support	2°	1°
Quality of the software	3°	5°
Maintenance support	4°	4°
Standardization	5°	3°
Improved development flexibility	6°	7°
Improved productivity	7°	6°
Shortened reaction time to changes	8°	8°
Platform independence	9°	9°

Hammond [7] stated that "in nearly every study of experts carried out within the judgment and decision-making approach, experience has been shown to be unrelated to the empirical accuracy of expert judgments"; such a statement fits very well the findings of our study, and in particular with RQ3.

While in general, expert judgment seems not to work particularly well, in the context of software development, effort estimation conducted by experts outperforms sophisticated formal methods [11]. The reasons provided by Jørgensen in [10] are: (i) the importance of highly context-specific knowledge in software development, (ii) the instability of relationship in software development (e.g., between effort and size) which lead to a very unpredictable field. The effect of expertise on judgment of other aspects of the software development process are rarely studied, as reported by Loconsole and Borstler in [15]. In their work they examine how expectations on requirements volatility matched the actual number of changes, resulting in a lack of statistical correlation between the expectation and the real outcome.

We have no data for explaining why it is so difficult forecasting the benefits of modeling and MD*. We can only report the work from Shanteau and Stewart [18]; they suggest that experts rely on heuristics in making judgments that could lead to systematic biases.

6. CONCLUSIONS

In conclusion, the results of this survey reveal that:

RQ1: *Improved documentation* and *Design support* are the most expected benefits from practitioners using modeling and/or MD*. Also *Quality of the software*, *Maintenance support*, and *Standardization* are frequently expected. On the contrary, other important benefits, such as *improved productivity* and *platform independence*, are not so much expected. That result tell us, indirectly, for which reason IT practitioners use models.

RQ2: The benefits having the highest fulfilment rate are still *Improved documentation* and *Design support* (Fulfilment rate $> 65\%$). However, considering all the benefits the average fulfilment rate is not high.

RQ3: Experience in modeling does not help in forecasting the benefits.

Probably the expectations are currently inflated by the amount of hype around MD*. It is possible that in the future practitioners will learn to focus on a smaller set of benefits and they will be able to actually achieve them more reliably.

All in all, this uncertainty about the outcomes of modeling and the fact that it affects also practitioners with many years of experience in the field is probably hampering the adoption of these approaches, which are always predicted to become mainstream in a never reached next future.

As a future work, it could be interesting to understand how much of the difficulty in forecasting the benefits of modeling and MD* depends on the immaturity of those approaches. Is that difficulty inherent in experts' judgement or is it worse in this particular field?

7. REFERENCES

[1] P. Baker, L. Shiou, and F. Weil. Model-driven engineering in a large industrial context - motorola case study. In L. Briand and C. Williams, editors, *Model Driven Engineering Languages and Systems*, volume 3713 of *Lecture Notes in Computer Science*, pages 476–491. Springer Berlin / Heidelberg, 2005.

[2] B. Boehm. The art of expectations management. *Computer*, 33(1):122 –124, jan 2000.

[3] C. F. Camerer and E. F. Johnson. The process-performance paradox in expert judgment: How can the experts know so much and predict so badly? In K. A. Ericsson and J. Smith, editors, *Towards a general theory of expertise: Prospects and limits*. Cambridge University Press, 1991.

[4] T. Dowling. Are software development technologies delivering their promise? In *IEE Colloquium on "Are Software Development Technologies Delivering Their Promise?"*, pages 1/1 –1/3, mar 1995.

[5] H. Esfahani, E. Yu, and M. Annosi. Capitalizing on empirical evidence during agile adoption. In *Agile Conference (AGILE), 2010*, pages 21 –24, aug. 2010.

[6] R. M. Groves, F. J. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. *Survey Methodology*. John Wiley and Sons, 2009.

[7] K. R. Hammond. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford University Press, USA, Oct. 2000.

[8] J. Hossler, M. Born, and S. Saito. Significant productivity enhancement through model driven techniques: A success story. In *Enterprise Distributed Object Computing Conference, 2006. EDOC '06. 10th IEEE International*, pages 367–373, oct. 2006.

[9] A. Jelitshka, M. Ciolkowski, C. Denger, B. Freimut, and A. Schlichting. Relevant information sources for successful technology transfer: a survey using inspections as an example. In *First International Symposium on Empirical Software Engineering and Measurement, 2007. (ESEM 2007)*, pages 31–40. IEEE, September 2007.

[10] M. Jørgensen. Estimation of software development work effort: Evidence on expert judgement and formal models. *Int. Journal of Forecasting*, 2007. In press.

[11] M. Jørgensen and S. Grimstad. Software development effort estimation: Demystifying and improving expert estimation. In O. L. Aslak Tveito, Are Magnus Bruaset, editor, *Simula Research Laboratory - by thinking constantly about it*, chapter 26, pages 381–404. Springer, Heidelberg, 2009.

[12] B. Kitchenham and S. Pfleeger. Personal opinion surveys. In F. Shull and Singer, editors, *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer London, 2008.

[13] A. G. Kleppe, J. Warmer, and et al. *MDA Explained: The Model Driven Architecture: Practice and Promise*. Addison-Wesley Longman Publishing Co., Inc, 2003.

[14] M. Leotta, F. Ricca, M. Ribaudo, G. Reggio, E. Astesiano, and T. Vernazza. SOA Adoption in the Italian Industry. In *Proceedings of 34th International Conference on Software Engineering (ICSE 2012)*, pages 1441–1442. IEEE, 2012.

[15] A. Loconsole and J. Borstler. Are size measures better than expert judgment? an industrial case study on requirements volatility. In *Software Engineering Conference, 2007. APSEC 2007. 14th Asia-Pacific*, pages 238 –245, dec. 2007.

[16] S. Mellor, A. Clark, and T. Futagami. Model-driven development - guest editor's introduction. *Software, IEEE*, 20(5):14 –18, sept.-oct. 2003.

[17] D. C. Schmidt. Guest editor's introduction: Model-driven engineering. *Computer*, 39:25–31, 2006.

[18] J. Shanteau and T. R. Stewart. Why study expert decision making? some historical perspectives and comments. *Organizational Behavior and Human Decision Processes*, 53(2):95–106, 1992.

[19] F. Tomassetti, A. Tiso, F. Ricca, M. Torchiano, and G. Reggio. Maturity of software modelling and model driven engineering: a survey in the italian industry. In *Int. Conf. Empirical Assessment and Evaluation in Software Eng. (EASE12)*, 2012.

[20] M. Torchiano, M. Di Penta, F. Ricca, A. De Lucia, and F. Lanubile. Migration of information systems in the italian industry: A state of the practice survey. *Information and Software Technology*, 53:71–86, January 2011.

[21] M. Torchiano, F. Tomassetti, A. Tiso, F. Ricca, and G. Reggio. Preliminary findings from a survey on the MD* state of the practice. In *International Symposium on Empirical Software Engineering and Measurement (ESEM 2011)*, pages 372–375, 2011.

[22] M. Völter. MD* best practices. *Journal of Object Technology*, 8(6):79–102, 2009.