7-1-2013

Assessing The Assessors - An Overview And Evaluation Of IT Project Success Reports

Dirk Basten
University of Cologne, Cologne, Germany, basten@wiso.uni-koeln.de

Oleg Pankratz
University of Cologne, Cologne, Germany, pankratz@wiso.uni-koeln.de

Dominik Joosten
University of Cologne, Cologne, Germany, joosten@wiso.uni-koeln.de

ASSESSING THE ASSESSORS – AN OVERVIEW AND EVALUATION OF IT PROJECT SUCCESS REPORTS

Basten, Dirk, University of Cologne, Pohligstrasse 1, 50969 Cologne, Germany, basten@wiso.uni-koeln.de

Pankratz, Oleg, University of Cologne, Pohligstrasse 1, 50969 Cologne, Germany, pankratz@wiso.uni-koeln.de

Joosten, Dominik, University of Cologne, Pohligstrasse 1, 50969 Cologne, Germany, joosten@wiso.uni-koeln.de

Abstract

A great number of scholars in the information systems discipline motivate their research referring to poor performance of information technology (IT) projects, which is reported in prior success studies. However, the rigour of such success studies has either not been analysed to an adequate degree or in case of the most prominent study, the Chaos Report, widely criticised. In order to ensure a sound research foundation, there is a need for assessing the quality of existing success reports. We conduct a systematic mapping study of IT project success reports and provide an overview and a critical evaluation of existing reports as a result. We enable scholars to more carefully and thoughtfully motivate their research and help practitioners to assess published success rates at their true worth. By pointing out strengths and critical weaknesses of existing success reports, we hope to contribute to more considerate future inquiries.

Keywords: Information systems, project success, success reports, systematic mapping, quality assessment.

1 Introduction

A great number of scholars in the information systems (IS) discipline motivate their research referring to poor performance of information technology (IT) projects (e.g. Richard et al., 2012; Jukic and Jukic, 2010; Napier et al., 2009; Chiang and Mookerjee, 2004). As indicators of poor performance serve studies that report IT project success rates. The most prominent IT project success study (henceforth we use *success study* and *success report* synonymously) is the Standish Group's Chaos Report (The Standish Group International, 2009), conducted biennially since 1994. According to the current issue of this report, 32% of IT projects are successful (projects meet their goals without exceeding time and budget constraints). However, scholars have criticised this study for shortcomings like incomplete description of study design, lack of reporting project-selection criteria and insufficient success definitions (Eveleens and Verhoef, 2010; Glass, 2006; Jørgensen and Moløkken-Østvold, 2006). Despite its lack of scientific rigour, many scholars even in top IS outlets refer to the Chaos Report to motivate their research and to emphasize the importance of successful project management without mentioning any of the critical issues (e.g. Liu et al., 2010; Napier et al., 2009; Singh et al., 2009; Gemino et al., 2008; Huff and Prybutok, 2008; Chiang and Mookerjee, 2004).

Further studies exist that report IT project success rates (e.g. El Emam and Koru, 2008; Sauer and Cuthbertson, 2003) and are also commonly cited (e.g. by Richard et al., 2012; Jukic and Jukic, 2010; Stamelos, 2010). This bears risks as these reports have not been analysed to an adequate extent. Overall, scientific rigour of existing IT project success reports often seems questionable; still, a substantial number of scholars regularly use them for motivating research without further scrutiny. Research is thus in need to assess IT success reports' strength of evidence (i.e. whether further research is likely to confirm the results; cf. Atkins et al., 2004) to ensure that such sources provide a sound foundation for related research. An according overview and evaluation of such success reports – that would contribute to a more sound research foundation – are missing. Accordingly, we pose two research questions:

(RQ1) What studies exist reporting IT project success rates?

(RQ2) Which strength of evidence do the identified IT project success reports provide?

We aim to provide an overview and evaluation of existing studies reporting IT project success rates by conducting a systematic mapping study. As broad reviews of primary studies, systematic mappings "are designed to provide a wide overview of a research area, to establish if research evidence exists on a topic and provide an indication of the quantity of the evidence" (Kitchenham, 2007, p. 44). Especially in maturing research fields, it is important to provide summaries and overviews (Petersen et al., 2008). We note that a systematic mapping is a special type of systematic literature review that "allows the evidence in a domain to be plotted at a high level of granularity" (Kitchenham, 2007, p. 5). A comparison with systematic literature reviews in general reveals several differences (Kitchenham, 2007, p. vii): In systematic mappings, (1) it is more likely to find broader and multiple research questions; (2) search terms are less focussed and are likely to result in a greater number of identified studies; (3) along with the search terms, the data extraction is broader and comparable to classification stage; (4) analysing the identified study is more likely to be on a summarising level, not include in-depth analyses and focus on totals and summaries. One of the benefits that results from conducting such a mapping is "the identification of evidence clusters and evidence deserts to direct the focus of future systematic reviews and to identify areas for more primary studies to be conducted" (Kitchenham, 2007, p. 5; for further information see Petersen et al., 2008). Our mapping focuses on the strength of evidence of the different success reports and provides an overview of their results.

Providing an overview and a critical assessment of IT project success reports, we contribute to research and practice in two major ways. First, researchers and practitioners gain a picture of the studies' quality as well as of the representativeness of each report compared to others. Accordingly, we enable scholars to more carefully and thoughtfully motivate their research and help practitioners to

assess published success rates at their true worth. Second, we point out strengths and weaknesses of existing success reports and thus hope to contribute to more considerate future studies by sensitizing researchers for critical aspects in such inquiries.

The remainder of this paper is structured as follows. In section 2, we outline related work on IS project success. In section 3, we present our research design by describing our systematic mapping. Section 4 provides the results of our mapping, that is the overview and critical assessment of the identified success reports. We discuss identified strengths and weaknesses of the reports in section 5. Section 6 concludes with limitations of our study and implications for future research and project management.

2 Related Work

Concerning success of projects in IT-related context, researchers have used a variety of terms to denote this concept. Common examples include *project performance* (Aladwani, 2002), *software project success* (Agarwal and Rathod, 2006), *IT project success* (Thomas and Fernández, 2008; Wateridge, 1995), *development project success* (Saarinen and Sääksjärvi, 1992) and *IS development success* (Banker and Kemerer, 1992). Wateridge (1998) demonstrates the disunity of denoting this term by referring to *IS/IT project success* in his paper title. Nowadays, the term (IT) project success is widely applied (e.g. Ika, 2009; Judgev and Müller, 2005) and seems to prevail. However, despite being topic of discussion for a long time, an agreed-on understanding of what actually constitutes this concept is missing. The variety of measurement approaches (e.g. Cuellar, 2010; Agarwal and Rathod, 2006; Aladwani, 2002; Wateridge, 1998) illustrates the lack of a commonly applied definition.

Beginning with the traditional view of using adherence to planning (time, budget and requirements) to assess project success (Wateridge, 1995; Pinto and Slevin, 1988), understanding and measurement of project success have apparently changed (Judgev and Müller, 2005). Researchers argue that measuring success in terms of adherence to planning is inappropriate (Agarwal and Rathod, 2006) or insufficient (Judgev and Müller, 2005; Pinto, 2004; Shenhar et al., 2001; Wateridge, 1995), leading to inadequate success evaluations (Shenhar et al., 2001; Dvir et al., 1998). Arguments for the inappropriateness of adherence to planning include this criterion's short-term character, which does not account for customer benefits (Agarwal and Rathod, 2006), inaccuracy of underlying estimates (Basten and Mellis, 2011) and not accounting for the subjective nature of success (Agarwal and Rathod, 2006). Considering this subjectivity, anecdotal and empirical evidence exist for what Nelson (2005) calls successful failures and failed successes. Some projects that fail according to traditional criteria are perceived successful and others that fulfil these criteria are perceived as failures (Ika, 2009; Pinto and Slevin, 1988) by various stakeholders. Consequently, researchers argue for assessing project success as a multidimensional construct (Ika, 2009; Thomas and Fernández, 2008; Aladwani, 2002) including additional criteria that surpass the limited traditional view of adherence to planning.

Several authors differentiate between the concepts process success and product success (e.g. Thomas and Fernández, 2008; Baccarini, 1999; Wateridge, 1998; Saarinen and Sääksjärvi, 1992). To cover these concepts, researchers propose additional criteria to complement adherence to planning instead of replacing it (Agarwal and Rathod, 2006; Wateridge, 1995), for instance customer satisfaction (Pinto, 2004; Pinto and Slevin, 1988) or stakeholder satisfaction in general (Judgev and Müller, 2005; Baccarini, 1999). Although many researchers equate adherence to planning and process efficiency (e.g. Dvir and Lechler, 2004; Freeman and Beale, 1992), they should be evaluated separately as exceeding budget or schedule does not necessarily result from wastage of resources. According to IS project managers, there is a clear difference between these concepts (Basten et al., 2012), which should be reflected in success assessments. As Gray (2001, p. 104) puts it: "If a project fails to meet an impossibly tight budget, but is efficiently delivered without wastage, to what extent should it be said to have 'failed'?". An extensive overview of success criteria potentially relevant for IT projects is given in Joosten et al. (in press).

Nevertheless, the traditional criteria are still frequently applied. This is typically justified by arguing that the objectivity of the traditional criteria leads to increased comparability of according study results (Joosten et al., in press; Pinto and Slevin, 1988). However, results of studies using these criteria are comparable to a limited degree only due to different criteria interpretations. Beside the criteria applied, an important aspect concerns the flexibility of defining success. In this context, flexibility can refer to tolerance levels for exceeding criteria or to the question whether all criteria need to be met. As Eveleens and Verhoef (2010) illustrate, applying inflexible criteria might lower success rates inadequately. For instance, different studies allow for different overrun tolerance levels (e.g. no budget overrun The Standish Group International, 2009, 10% budget overrun Miller et al., 2007 or 30% budget overrun Whittaker, 1999 allowed for considering a project a success).

As regards the most prominent success study, the Chaos Report, previous works have criticised its lack of scientific rigour – among others, incomplete description of study design, lack of reporting project-selecting criteria and insufficient success definitions (Eveleens and Verhoef, 2010; Glass, 2006; Jørgensen and Moløkken-Østvold, 2006). The extensive analysis of this report can be attributed to its prominence. While other success reports (e.g. El Emam and Koru, 2008) have included the criticism in their own analysis, they have not been analysed themselves.

3 Systematic Mapping

3.1 Data Sources and Search Process

To identify success reports, we conducted a systematic keyword-based search of the following databases (in alphabetical order): *ACM Digital Library*, *AIS Electronic Library*, *EBSCOHost*, *EmeraldInsight*, *Google Scholar*, *IEEEXplore*, *ProQuest*, *ScienceDirect*, *SpringerLink*, *Web Of Knowledge*, *Wiley InterScience*. The search pattern, which relevant articles needed to meet in title, abstract or keywords, can be found in our Online Appendix (http://tinyurl.com/ecis2013success).

However, the results were sparse as most success reports seem not to be published in scientific outlets. We thus also applied our search to Google. Although this approach is different as complex search terms need to be separately handled, it yielded the most potentially relevant success reports. Identified success reports were reviewed in a two-step procedure. (1) Two authors independently identified relevant reports by screening the search results according to the inclusion/exclusion criteria as described below. (2) We merged both lists and discussed diverging views. Additionally, we applied a search backward and forward (Webster and Watson, 2002) to identify further success reports. A detailed overview of our search process can be found in our Online Appendix (http://tinyurl.com/ecis2013success).

3.2 Inclusion and Exclusion Criteria

We selected relevant studies by screening their full text. Included studies meet the following criteria:

- *Reporting success rates of IS or IT projects is one of the primary objectives of the study*. Only studies setting the objective to provide such success rates can be scrutinised with regard to the quality of achieving this objective. Studies that used success rates in their work – and sometimes even investigated those rates themselves – but did not pose the objective to report on success rates in the first place were therefore excluded. For example, although also investigating success rates, the KPMG study 1997 was excluded from our analysis as in the only available report (Whittaker, 1999) the single primary objective was to outline *reasons* for project failure.
- *English-language empirical research*. Studies in other languages as well as non-empirical studies like technical reports, discussions and opinion papers were excluded to ensure a homogeneous and well-founded set of reports under scrutiny.

- *Reporting on cross-organisational success rates of multiple IS or IT projects*. We applied this criterion to prevent potential biases that can result from considering only single projects or projects from only one organization.

In case of publications with multiple editions and varying amount of information on the research approach (e.g. Chaos Report has been biennially published since 1994 and the provided level of detail concerning study design in the respective issues varies considerably), we collected relevant data from all publications assuming that study design remains the same. Concerning success rates, we provide the most recent available data.

3.3 Quality Assessment

We assess the quality of each included success report applying an instrument adapted from Dybå and Dingsøyr (2008). This approach has been used for quality assessment in other systematic reviews and mappings (e.g. Chen and Babar, 2011; Alves et al., 2010). It contains questions covering aspects of study design concerning rigour, reporting, credibility and relevance of success studies. For the purpose of better traceability of our results, these questions are listed in section 4.2 (cf. Table 3). We omitted one of the questions (Is the paper based on research (or is it merely a "lessons learned" report)?) as this was assured by our inclusion/exclusion criteria (cf. section 3.2). For the remaining ten questions, we use three possible responses (yes, to some extent, no) and consequently apply a three-point scale (1, 0.5, 0). The sum of the scores of one study builds the overall quality score for that study.

3.4 Data Collection

Two authors independently extracted all data required to answer the research questions (cf. section 1). With this way of data triangulation, we aimed to ensure the completeness and validity of our analysis. For identifying relevant studies for our overview (RQ1), this involved reviewing results of the search in the specified databases and applying our inclusion and exclusion criteria. For assessing the strength of evidence of identified studies (RQ2), it involved applying the quality assessment instrument (cf. section 3.3) to all studies identified in RQ1.

4 Results

We structure the results according to our research questions. While section 4.1 provides an overview of studies reporting IT project success rates (RQ1), section 4.2 is dedicated to the strength of evidence of these studies (RQ2).

4.1 IT Project Success Reports

We identified six success studies of interest for our mapping study (listed in Table 1). Considering the latest publication of each study, the success reports have been published between 2003 and 2009.

The description of the unit of analysis is rudimentary in most identified studies. Five studies refer to IT projects without defining this term (The Standish Group International, 2009; Miller, 2008; Miller et al., 2007; Sauer and Cuthbertson, 2003; Sonnekus and Labuschagne, 2003). El Emam and Koru (2008, p. 84) describe the following project type: "The unit of measurement for this study was the software project (not process improvement, organizational change, or business-process-reengineering projects)".

At this point, information about the projects' context (industries, geographical regions, organization size, project size, respondents' position etc.) is also of interest. In the recent Chaos Report, organization sizes and geographical regions are provided only roughly and in percentage values without reporting the sample size. Information is given neither on the included industries nor on the

Success Study	Summary
Sauer and Cuthbertson, 2003: The State of IT Project Management in the UK	Web survey of 421 IT projects in 2002-2003 in UK, investigating the state of project performance, characteristics of project managers and performance drivers.
Sonnekus and Labuschagne, 2003: The Prosperus Report	Study among IT project managers in South Africa investigating success rates of IT projects and the maturity of IT project management.
The Standish Group International, 2009: CHAOS Summary 2009. The 10 Laws of CHAOS.	Cross-national (concentrated on US and Europe) study of IT projects in 2008, investigating project success rates and also providing recommendations (Laws) for achieving success.
Miller, 2008: IT Project Success & Failure	Study of 131 IT projects in 2005-2007 primarily in the UK, rest of Europe and US, investigating success rates & factors and failure cost & causes.
Miller et al., 2007: IT Project Success – The Computing Holy Grail	Study of 163 IT projects in UK in 2005-2007, investigating success rates as well as how the applied development model, using PRINCE2 or the industry affect project success.
El Emam and Koru, 2008: A Replicated Survey of IT Software Project Failures	Replicated international Web survey of IT projects in 2005 and 2007, investigating project cancellation and success rates, failure factors and impact of project size on success.

Table 1: Analysed IT Project Success Reports

distribution of successful/challenged/failed projects over single industries. In contrast, Miller et al. (2007) provide percentages of successful and failed projects for single industries. Three studies investigate projects in one geographical region (twice UK (Miller et al., 2007; Sauer and Cuthbertson, 2003) and once South Africa (Sonnekus and Labuschagne, 2003)), the other three report cross-national data. Overall, the provided amount of context information varies considerably. Some studies provide detailed overviews of according data while others do not mention certain aspects at all.

	Success Criteria					Success Rates			
Success Study	Budget	Schedule	Requirements	Stakeholder Satisfaction	Staff Productivity	Successful	Challenged	Failed	
Sauer and Cuthbertson, 2003	X	X	X			16% [1]	75% [2]	9% [3]	
Sonnekus and Labuschagne, 2003	X	X	X			42.87% [1]	34.97% [2]	22.17% [3]	
The Standish Group International, 2009	X	X	X			32% [1]	44% [2]	24% [3]	
Miller, 2008	X	X	X	X*		31.30% [1]	55.73% [2]	12.97% [3]	
						38.93% [4]		61.07% [5]	
Miller et al., 2007	X	X	X			44% [4]		56% [5]	
El Emam and Koru, 2008	X	X	X	X	X	34.23% [6]	12.12% [7]	12.23% [8]	11.54% [9]

[1] Project completed on-time and on-budget, with all features and functions as initially specified
[2] Project completed and operational but over budget, over time or with less functionality than initially specified
[3] Project is cancelled at some point during the development cycle
[4] If not a failure according to [5]
[5] Project exceeds the initially estimated delivery date or the initially estimated budget by 10% or more, does not deliver full functionality of original specification within a 10% allowance or is cancelled
[6] Rated poor or fair on zero or one criterion (poor and fair being two lowest points on a four-point Likert scale)
[7] Poor or fair on two or three criteria
[8] Unsuccessful: Poor or fair on four or five criteria
[9] Cancelled: No usable functionality at first release
*Data collected on stakeholder satisfaction does not apply to concrete projects but is concerned with participants' general experiences. As such, it does not impact the success rates provided above.

Table 2: Overview of Success Criteria Applied and According Success and Failure Rates

For a meaningful comparison of different studies, it is important to consider which criteria have been applied and in which ways they have been combined to assess project success. Table 2 shows the concept matrix (Webster and Watson, 2002) concerning the success reports and the applied success criteria. Moreover, Table 2 provides success and failure rates according to the success definitions of the different success reports. We emphasise that success rates in same columns do not necessarily imply the same success definitions (cf. the according numbers). Comparable success rates (due to equal definitions) are highlighted by the same colour.

Considering success criteria, the most common ones (applied in all six studies) are adherence to budget, adherence to schedule and conformance with specified requirements. Additionally, selected studies apply criteria concerning satisfaction of different stakeholders (El Emam and Koru, 2008; Miller, 2008) and staff productivity (El Emam and Koru, 2008). Each study distinguishes between successful and failed projects. Except for the report by Miller et al. (2007), all studies additionally classify projects as challenged if they are neither completely failed nor completely successful. Beside these similarities, there are also some differences. Whereas some studies classify projects as failed if they are cancelled (The Standish Group International, 2009; Sauer and Cuthbertson, 2003; Sonnekus and Labuschagne, 2003), one study distinguishes between successful, challenged and unsuccessful projects only if they are not cancelled (El Emam and Koru, 2008). Whereas Miller et al. (2007) only differentiate between successful and failed projects, Miller (2008) analyses collected data in both ways (cf. Table 2) to enable comparisons with both types of classifications.

4.2 Strength of Evidence

We assess the success studies' quality using our instrument (cf. section 3.3), which is illustrated in Table 3. The according results are shown in Table 4. As stated above, two researchers independently answered the according questions to reduce the subjectivity of assessments. Our inter-coder reliability (85.5%; number of agreements divided by the total number of agreements and disagreements Miles and Huberman, 1994, p. 46) is above the threshold of 80% (Nunnally, 1978, pp. 245-246) which can be seen as recommended standard for the majority of purposes (Lance et al., 2006). Deviating assessments were discussed until agreement was reached.

No	Question
Q1	Is there a clear statement of the aims of the study?
Q2	Is there an adequate description of the context (including the unit of analysis) in which the study was carried out?
Q3	Was the research design (including specified success criteria and their flexibility) appropriate to address the aims of the study?
Q4	Were the recruitment strategy and sample appropriate to the aims of the study?
Q5	Are the results compared to results of other success studies?
Q6	Was the data collected in a way that addressed the research issue?
Q7	Was the data analysis sufficiently rigorous?
Q8	Was the relationship between researchers and participants sufficiently considered?
Q9	Is there a clear statement of findings?
Q10	Is the study of value for research and practice?

Table 3: Questions used for quality assessment checks (adapted from Dybå and Dingsøyr, 2008)

In addition to the quality assessments, Table 4 provides the strength of evidence that we assessed for each study. There are four possible values for the strength of evidence (Atkins et al., 2004)):

- Very low: the results are very uncertain
- Low: Further research is very likely to have an important impact on the confidence in the results
- Moderate: Further research is likely to have an important impact on the confidence in the results
- High: Further research is very unlikely to change the confidence in the results

Whereas our strength-of-evidence assessments are mainly based on the quality assessments, they are our overall evaluations of the studies' strength of evidence (despite high quality assessments, a success report's strength of evidence could be rated *very low*, if for instance success rates of only five projects are reported). Two researchers also conducted these assessments independently, with agreement in each case.

Study	Quality Assessments											Strength of Evidence
	Cluster 1			Cluster 2				Cluster 3				
	Q1	Q6	Q9	Q10	Q7	Q2	Q4	Q5	Q3	Q8	∑	
Sauer and Cuthbertson, 2003	1	1	1	1	1	0.5	1	0.5	0.5	0.5	8	Moderate
Sonnekus and Labuschagne, 2003	1	1	1	0.5	0.5	0.5	0.5	0.5	0	0	5.5	Low
The Standish Group International, 2009	1	0.5	0.5	0.5	0.5	0.5	0	0	0	0.5	4	Low
Miller, 2008	1	1	1	1	0.5	0.5	1	1	1	0.5	8.5	Moderate
Miller et al., 2007	1	1	1	0.5	0.5	0.5	0.5	1	0.5	0	6.5	Low
El Emam and Koru, 2008	1	1	1	1	1	1	0.5	0	0.5	0.5	7,5	Moderate
∑	6	5.5	5.5	4.5	4	3.5	3.5	3	2.5	2		

Table 4: Quality Assessment Results and Strength of Evidence

For our analysis, we arranged the overall scores of the questions (undermost row in Table 4) in descending order and clustered those according to the sum of all studies.

- Cluster 1: Questions with a sum greater than or equal to 5 ($X \geq 5$): Q1, Q6, Q9
- Cluster 2: Questions with a sum less than 5 and greater than 3 ($5 > X > 3$): Q10, Q7, Q2, Q4
- Cluster 3: Questions with a sum less than or equal to 3 ($X \leq 3$): Q5, Q3, Q8

In general, the analysed success studies provide clear statements of their aims (Q1), collect data in an adequate way according to these aims (Q6) and present a clear statement of their findings (Q9) (cf. cluster 1 in Table 4). Considering cluster 2, the success reports are in general of interest for research and practice (Q10) but partly neglect to report their benefits. They partly lack an adequate description of the study context (Q2). In four out of six studies, little information is given concerning the analysis of data (Q7). As regards the question whether the recruitment strategy and sample were appropriate to the aims of the study (Q4), there is lack of information at least in terms of justification. Cluster 3 contains aspects with the highest need for improvement. Most studies do not compare their results (Q5) or focus on a single reference study, mainly the Chaos Report. The research design (Q3) is insufficient in most reports mainly due to a limited set of success criteria, an inadequate aggregation of those and the inflexibility of exceeding the criteria (cf. section 2). While we considered the relationship between researchers and participants (Q8) to be regarded to some extent (value 0.5) if researchers were not actively engaged in data collection (e.g. using a web survey), none of the analysed studies actually provided any information regarding this aspect.

5 Discussion

Our systematic mapping led to interesting insights gained from six different success reports published within the past decade. We discuss our results and point out strengths and weaknesses of the identified studies below.

As regards cluster 1 (cf. Table 4), almost all identified success reports provide clear information concerning the according assessment questions (clear statement of study aims, appropriate way of data collection, clear statement of findings). Especially the clear statements of study aims and results are not surprising as these are evident in the context of IT project success rates. Questionnaires were used for data collection, partially complemented by interviews. In total, the aspects covered in cluster 1 can be considered to be the strengths of identified studies.

The four analysed aspects in cluster 2 (value for research and practice, adequate context description, rigorous data analysis, appropriate recruitment strategy and sample) reveal shortcomings in at least

three out of the six identified studies (cf. Table 4). Some studies' value for research and practice is limited due to, for instance, potential biases inherent to the chosen sample. We found only one success study (El Emam and Koru, 2008, cf. section 4.1) providing a delineation of the unit of analysis (by referring to a software project and excluding specific project types). Beyond this delineation, a clear definition of the unit of analysis is not given. Other studies simply refer to IT projects without any specifications of this term. Regarding the term projects, they might develop, extend or adapt IT, and these different project goals are likely to have an impact on project characteristics and according success rates. Also, it is not unambiguously clear what is meant by IT. One reason for the lack of a definition might be that the term information technology is defined in dictionaries. However, whereas the Oxford Dictionary defines IT as "the study or use of systems (especially computers and telecommunications) for storing, retrieving, and sending information", a definition in the IS discipline is as follows: "information technology consists of all the hardware and software technologies a firm needs to achieve its business objectives" (Laudon and Laudon, 2009, p. 45). As the success studies refer to IT projects in general without precise definitions, it is unclear whether for instance mere hardware projects are included. Partially neglecting to provide detailed context information (geographical regions, industries, organization size and age, project size (e.g. in number of team members) and date) further adds to the findings' bleariness. The same applies to the transparency of the data analysis and recruitment strategy – the amount of provided information varies between detailed extensive description and rudimentary mentions. All these ambiguities impede deriving meaningful statements concerning the state of IT projects.

The greatest potential for improvement, however, lies in the aspects in cluster 3 (results comparisons, research design, researcher-participant relationship). We believe that it is crucial to compare results to those of other success reports as the reasons for potential deviations are likely to impact the implications derived from the results. Miller (2008) provides a good example for this by analysing collected data in different ways to enable comparisons with several prior studies. Contrarily, most studies do not compare their results or focus on a single success study, mainly the Chaos Report. As this report has been widely criticized (cf. section 1 as well as our assessment in Table 4), the benefits of such comparisons are questionable. Ideally, researchers should classify the analysed projects according to the categorization of previous success reports and discuss reasons for deviations. Accordingly, it is important to consider comparability aspects when defining the applied success criteria in the research design the first place. Our analysis of the applied success criteria revealed a predominant focus on the classic criteria of adherence to planning: budget, schedule and requirements (cf. Table 2). As these are also the criteria that are widely applied in organizations (Joosten et al., in press; Thomas and Fernández, 2008), this finding is not surprising. However, following the predominant opinion in the on-going discussion concerning these classic success criteria (cf. section 2), adding further success criteria is desirable for a more holistic assessment of IT project success rates. A positive finding in this regard was the inclusion of stakeholder satisfaction (El Emam and Koru, 2008; Miller, 2008) and staff productivity (El Emam and Koru, 2008) as success criteria. Whereas the former accounts for the subjective nature of success assessments (Myers, 1995), the latter implies a differentiation between adherence to planning and the efficiency of the development process (which is reasonable as argued in section 2). Overall, assessing success of IT projects should include a set of success criteria that cover the overall success concept instead of using the limited set of adherence to planning. Beyond the limited set of success criteria, we found some of the applied aggregation approaches of these criteria to be rather inflexible. For instance, the Chaos Report defines IT projects as failed even in case of minimal budget or schedule overruns. Miller et al. (2007) and Miller (2008) counteract this problem by allowing for a 10% exceedance. Compared to all other success reports, the study by El Emam and Koru (2008) provides a different picture concerning the overall success definition as it provides another kind of flexibility concerning the success rates. By differentiating the number of criteria that are rated poor or fair, IT projects are categorised as successful (zero or one criterion), challenged (two or three criteria), unsuccessful (four or five criteria) or cancelled (no usable functionality at first release). Future success studies should thus consider including an adequate degree of flexibility in their definitions of success and failure. Finally, none of

the studies mentions considerations of the relationship between researchers and participants. Whereas we rated it to be considered to some extent if the researchers were not actively engaged in data collection (e.g. if a web survey was used), this aspect especially requires attention in situations with high bias potential (e.g. interviews). Therefore, researchers should explicitly consider potential effects of the researcher-participant relationship when investigating IT project success rates.

Overall, the strength of evidence of three success reports was rated low (The Standish Group International, 2009; Miller et al., 2007; Sonnekus and Labuschagne, 2003) or moderate (El Emam and Koru, 2008; Miller, 2008; Sauer and Cuthbertson, 2003), respectively. Our results thus confirm previous assessments questioning the approach used in the Chaos Report (cf. section 1). In this regard, we are aware of practitioners who for instance refer to the Chaos Report to motivate the acquisition of consulting services. Practitioners can improve the foundation of such motivations by using one of the higher rated studies or by collecting data in a way that avoids the various shortcomings outlined in this paper.

6 Conclusion

Considering our study's limitations, our analyses are limited to what is reported. We have no insight into what the investigators have done additionally. Moreover, as we identified most success reports using Google, our findings might be biased to this search engine. Besides, we had to omit unavailable success reports (that assumingly meet our inclusion criteria). For instance, one study that among others investigated success rates had to be excluded as the only available paper focuses only on reasons for project failure (Whittaker, 1999), while further potentially relevant reports were not available at all (e.g. OASIG Survey 1995). Here, our results are limited to what is available. Finally, while we identify specific shortcomings and derive according recommendations, our study does not provide a holistic framework, that is an extensive set of guidelines for conducting and reporting success studies.

Our findings entail important implications. Due to partly different definitions of success and failure (see Table 2 for a detailed comparison), it is not meaningful to derive an overall success rate of IT projects from the investigated studies. Depending on success definitions, the IT project success rates range between 16% and 44% (cf. Table 2). With 32%, the often cited and criticized Chaos Report stands in the middle of this interval. However, our assessment confirms previous criticism concerning the rigour of this study. Moreover, concerning all investigated studies, adequate interpretation of success rates is impeded as it is not clearly evident to which types of projects these rates actually apply. Lack of providing context information and project type definitions leads to the perception that specific success rates apply to IT projects in general. Also, scholars should more carefully assess whether cited studies used to motivate research by referring to the poor state of IT projects are credible and representative. A high citation count does not ensure a high quality of the source. Although it might for instance be appealing to use the name Chaos and the according success rates to motivate IT consultancy services, IT professionals should first and foremost avoid using success rates that might be based on false assessments. Assessing success only in terms of adherence to planning is probably not what customers want. Thus, we call for more substantiated success reports with various success criteria and rigorous study design.

References

Agarwal, N. and Rathod, U. (2006). Defining 'success' for software projects: An exploratory revelation. International Journal of Project Management, 24 (4), 358-370.

Aladwani, A.M. (2002). An Integrated Performance Model of Information Systems Projects. Journal of Management Information Systems, 19 (1), 185-210.

Alves, V., Niu, N., Alves, C. and Valença, G. (2010). Requirements engineering for software product lines: A systematic literature review. Information and Software Technology, 52 (8), 806-820.

Atkins, D., Best, D., Briss, P.A., Eccles, M., Falck-Ytter, Y., Flottorp, S., Guyatt, G.H., Harbour, R.T., Haugh, M.C., Henry, D., Hill, S., Jaeschke, R., Leng, G., Liberati, A., Magrini, N., Mason, J., Middleton, P., Mrukowicz, J., O'Connell, D., Oxman, A.D., Phillips, B., Schünemann, H.J., Edejer, T.T.-T., Varonen, H., Vist, G.E., Williams, J.W. and Zaza S (2004). Grading quality of evidence and strength of recommendations. British Medical Journal, 328 (7454), 1490-1496.

Baccarini, D. (1999). The Logical Framework Method for Defining Project Success. Project Management Journal, 30 (4), 25-32.

Banker, R.D. and Kemerer, C.F. (1992). Performance Evaluation Metrics for Information Systems Development: A Principal-Agent Model. Information Systems Research, 3 (4), 379-400.

Basten, D., Joosten, D. and Mellis, W. (2012). Managers' Perceptions of Information System Project Success. Journal of Computer Information Systems, 52 (2), 12-21.

Basten, D. and Mellis, W. (2011). A Current Assessment of Software Development Effort Estimation. In Proceedings of the 5th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, September 22-23, Banff, Canada.

Chen, L. and Babar, M.A. (2011). A systematic review of evaluation of variability management approaches in software product lines. Information and Software Technology, 53 (4), 344-362.

Chiang, I.R. and Mookerjee, V.S. (2004). A Fault Threshold Policy to Manage Software Development Projects. Information Systems Research, 15 (1), 3-21.

Cuellar, M. (2010). Assessing Project Success: Moving Beyond the Triple Constraint. In International Research Workshop on IT Project Management, St. Louis.

Dvir, D. and Lechler, T. (2004). Plans are nothing, changing plans is everything: the impact of changes on project success. Research Policy, 33 (1), 1-15.

Dvir, D., Lipovetsky, S., Shenhar, A.J. and Tishler, A. (1998). In Search of Project Classification: A Non-Universal Approach to Project Success Factors. Research Policy, 27 (9), 915-935.

Dybå, T. and Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. Information and Software Technology, 50 (9-10), 833-859.

El Emam, K. and Koru, A.G. (2008). A Replicated Survey of IT Software Project Failures. IEEE Software, 25 (5), 84–90.

Eveleens, J.L. and Verhoef, C. (2010). The Rise and Fall of the Chaos Report Figures. IEEE Software, 27 (1), 30-36.

Freeman, M. and Beale, P. (1992). Measuring project success. Project Management Journal, 23 (1), 8-17.

Gemino, A., Reich, B.H. and Sauer, C. (2008). A Temporal Model of Information Technology Project Performance. Journal of Management Information Systems, 24 (3), 9-44.

Glass, R.L. (2006). The Standish report: does it really describe a software crisis? Communications of the ACM, 49 (8), 15-16.

Gray, R.J. (2001). Organisational climate and project success. International Journal of Project Management, 19 (2), 103-109.

Huff, R.A. and Prybutok, V.R. (2008). Information systems project management decision making: The influence of experience and risk propensity. Project Management Journal, 39 (2), 34-47.

Ika, L.A. (2009). Project Success as a Topic in Project Management Journals. Project Management Journal, 40 (4), 6-19.

Joosten, D., Basten, D. and Mellis, W. (in press). Measurement of Information System Project Success in German Organizations. International Journal of Information Technology Project Management.

Jørgensen, M. and Moløkken-Østvold, K. (2006). How large are software cost overruns? A review of the 1994 CHAOS report. Information and Software Technology, 48 (4), 297-301.

Judgev, K. and Müller, R. (2005). A Retrospective Look at Our Evolving Understanding of Project Success. Project Management Journal, 36 (4), 19-31.

Jukic, B. and Jukic, N. (2010). Information System Planning and Decision Making Framework: A Case Study. Information Systems Management, 27 (1), 61-71.

Kitchenham, B. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report.

Lance, C.E., Butts, M.M. and Michels, L.C. (2006). The Source of Four Commonly Reported Cutoff Criteria. What Did They Really Say? Organizational Research Methods, 9 (2), 202-220.
Laudon, K.C. and Laudon, J.P. (2009). Management Information Systems. Managing the Digital Firm. 11th Edition. Pearson, Upper Saddle River.
Liu, S., Zhang, J., Keil, M. and Chen, T. (2010). Comparing senior executive and project manager perceptions of IT project risk: a Chinese Delphi study. Information Systems Journal, 20 (4), 319-355.
Miles, M.B. and Huberman, A.M. (1994). Qualitative data analysis. Sage, Thousand Oaks.
Miller, K., Dawson, R. and Bradley, M. (2007). IT Project Success – The Computing Holy Grail, Loughborough, Leicestershire.
Miller, M. (2008). IT Project Success & Failure, Loughborough, Leicestershire.
Myers, M.D. (1995). Dialectical hermeneutics: a theoretical framework for the implementation of information systems. Information Systems Journal, 5 (1), 51-70.
Napier, N.P., Keil, M. and Tan, F.B. (2009). IT project managers' construction of successful project management practice: a repertory grid investigation. Information Systems Journal, 19 (3), 255-282.
Nelson, R. (2005). Project retrospectives: evaluating project success, failure, and everything in between. MIS Quarterly Executive, 4 (3), 361–372.
Nunnally, J.C. (1978). Psychometric theory. 2nd Edition. McGraw-Hill, New York.
Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. In 12th International Conference on Evaluation and Assessment in Software Engineering, p. 1, Bari.
Pinto, J.K. (2004). The Elements of Project Success. In Field Guide To Project Management (Cleland, D.I. Ed.), p. 14, Wiley, Hoboken.
Pinto, J.K. and Slevin, D. (1988). Project Success: Definitions and Measurement Techniques. Project Management Journal, 19 (1), 67-72.
Richard, P., Coltman, T. and Keating, B. (2012). Designing IS service strategy: an information acceleration approach. European Journal of Information Systems, 21 (1), 87-98.
Saarinen, T. and Sääksjärvi, M. (1992). Process and Product Success in Information Systems Development. Journal of Strategic Information Systems, 1 (5), 266-275.
Sauer, C. and Cuthbertson, C. (2003). The State of IT Project Management in the UK 2002-2003, Oxford.
Shenhar, A.J., Dvir, D., Levy, O. and Maltz, A.C. (2001). Project Success: A Multidimensional Strategic Concept. Long Range Planning, 34 (6), 699-725.
Singh, R., Keil, M. and Kasi, V. (2009). Identifying and overcoming the challenges of implementing a project management office. European Journal of Information Systems, 18 (5), 409-427.
Sonnekus, R. and Labuschagne, L. (2003). The Prosperus Report 2003: IT Project Management Maturity versus Project Success in South Africa, Johannesburg.
Stamelos, I. (2010). Software project management anti-patterns. Journal of Systems and Software, 83 (1), 52-59.
The Standish Group International (2009). CHAOS Summary 2009. The 10 laws of CHAOS.
Thomas, G. and Fernández, W. (2008). Success in IT projects: a matter of definition. International Journal of Project Management, 26 (7), 733-742.
Wateridge, J. (1995). IT Projects: A Basis for Success. International Journal of Project Management, 13 (3), 169-172.
Wateridge, J. (1998). How can IS/IT projects be measured for success? International Journal of Project Management, 16 (1), 59-63.
Webster, J. and Watson, R.T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26 (2), xiii-xxiii.
Whittaker, B. (1999). What went wrong? Unsuccessful information technology projects, Toronto, Canada.