Science of Computer Programming 130 (2016) 37–68

Contents lists available at ScienceDirect

Science of Computer Programming

www.elsevier.com/locate/scico

Lessons from Dutch IT-outsourcing success and failure

G.P.A.J. Delen, R.J. Peters, C. Verhoef, S.F.M. van Vlijmen

ARTICLE INFO

Article history:
Received 4 December 2014
Accepted 22 March 2016
Available online 27 April 2016

Keywords:
Empirical study
IT-outsourcing success
Controllable success determinants
Rigid success determinants
Outsourcing consultant

ABSTRACT

We present the findings from a Dutch field study of a representative sample of 30 outsourcing deals totalling to more than 100 million Euro, where both customers, corresponding IT-outsourcing providers and their intermediaries (if present) participated. Of the 30 deals, 18 were successful. As the sample is representative for the 700 Dutch IT-outsourcing deals between 2007–2010, the Dutch success rate was about 60% at that time. The main objective of the study was to examine what made some a success and some a failure and how to influence the outcome if possible. For a number of well-known factors we investigated whether they discriminate between IT-outsourcing success and failure in the early phase of service delivery and determined their impact on the chance on a successful deal. We investigated controllable factors to increase the odds during sourcing and rigid factors as a warning sign before closing a deal.

Based on 253 interviews with 516 different questions we collected 28 thousand data points (which took several FTEs). From the data and the perceived failure or success of the closed deals we investigated the discriminative power of the determinants (ex post). We found three statistically significant controllable factors that discriminated in an early phase between failure and success. They are: working according to the transition plan, demand management and, to our surprise, communication within the supplier organisation (so not between client and supplier). These factors also turned out to be the only significant factors for a (logistic) model predicting the chance of a successful IT-outsourcing. Improving demand management and internal communication at the supplier increases the odds the most, sticking to the transition plan only modestly. Other controllable factors were not significant in our study. They are: managing the business case, transfer of staff or assets, retention of expertise and communication within the client organisation. Of the rigid factors, the motive to outsource, cultural differences, and the type of work were insignificant. The motive of the supplier was significant: short-term motivations like increasing profit margins or business volume decreased the chance of success while long-term motivations like increasing market share or becoming a player increased the success rate. From the data we inferred that the degree of experience with sourcing did not show to be a convincing factor of success. Hiring sourcing consultants (intermediaries) worked contra-productive: it lowered chances of success. This illustrates that you better not outsource outsourcing.

1. Introduction

Outsourcing is becoming a common alternative for in-house development and maintenance for Information Technology (IT). The complete gamut from hosting services to entire business processes is available in the global outsourcing market.

E-mail address: s.f.m.vanvlijmen@uva.nl (S.F.M. van Vlijmen).

http://dx.doi.org/10.1016/j.scico.2016.04.001
0167-6423/

Next to the strongly evolving practice of outsourcing it has also become an established field of research. Nowadays there is an abundance of publications dedicated to IT-outsourcing in scientific journals which has been growing rapidly during the last decades. Typical research topics are determinants of IT outsourcing success, IT-outsourcing strategies, the risks of IT-outsourcing, and more [36].

The focus of our study is on the determinants of outsourcing success. We analysed the findings of 30 IT-outsourcing deals that were closed in the years 2007–2010 in the Netherlands. The 30 IT-outsourcing deals were followed during at least a two-year period stretching the decision phase to a point deep in the service delivery phase. Our study can therefore be characterised as a *longitudinal study*. To set up a longitudinal study is a costly and time-consuming undertaking and therefore there are just a few examples of such studies. The research period should be long enough to collect the research data required and a sufficiently large number of representative participants must be willing to participate. Most examples of longitudinal studies date to medical sciences. First a cohort of patients must be selected that is ready for monitoring during a fixed period (sometimes more than twenty years). Next, the medical data required must be collected during that period, which usually means that measurements must be made on the participants in the study by medical professionals, for instance, a longitudinal study on the treatment of breast cancer. In our case it took about 1.5 year to find a sufficiently large number of companies, outsourcing companies and service providers, that were ready to participate in our longitudinal study, and to find sponsors that were willing to finance the study. We claim that our longitudinal study into the determinants of outsourcing success in which so many companies participated is unique for this type of study.

Although it is generally recognized that two different parties are needed to play the IT-outsourcing game, most models primary focus on the role of the outsourcing company, leaving the role of the vendor underexposed. Our study was set up to gain information from all parties involved in a deal. So our approach is symmetrical in the way it deals in an equal way with the roles the outsourcer and service provider play in the IT-outsourcing process. We verified that our sample of 30 IT-outsourcing deals was a reasonably representative sample out of the population of Dutch IT-outsourcing deals in the years 2007–2010.

We conducted 253 interviews and we used questionnaires to semi-structure them. The participating companies delivered via interviews data of 30 IT-outsourcing cases. In each case at least two business partners were involved: the outsourcing company and its service provider. In half the cases (15 to be exact) also an IT-sourcing consultant was involved. Their task was to support the IT-outsourcing company in the vendor selection process and the conclusion of a service level agreement (contract negotiation). The foundation for the questionnaires was laid during meetings with representatives of the participating organisations and the sourcing consultants. Over at least a two-year period we collected 28 thousand data points. We completed the forms ourselves to be sure that the interpretation of the forms would be consistent. In Section 4 we describe the data collection process in detail.

In our study we make a distinction between *controllable success determinants* and *rigid success determinants*. For a controllable success determinant or factor the related circumstances can be changed in an early stage of the sourcing process during the service delivery phase by the outsourcing company and/or its service provider in order to reduce the chance on failure. A rigid success factor is fixed at the start of the sourcing process and cannot be affected during the service delivery phase anymore.

In our study we focused attention on well-known success factors. We investigated nine controllable factors and six rigid factors. Based on our measured success or failure, we investigated the discriminative power of these factors in an ex post manner. Namely, given the outcome, do the factors make a difference or not? We found three statistically significant controllable factors that discriminated in an early phase between failure and success. From the rigid factors also three factors appeared to be significant. After having discussed the related work to our study in Section 2, we are presenting the controllable and rigid success factors we tested in Section 3.

For the major part the success factors we tested are well-known. For instance, they are not only frequently mentioned in literature but also by the participating organisations of our study. Moreover we have encountered them in our advisory work. However, it is unknown whether they really possess power to discriminate between successful and failed deals. Of course, an often-mentioned success factor does not become more successful by mentioning it more often. In our research we determined to what extent such often seen factors indeed contribute to success or failure. Or, in other words, we tested whether they really matter or not.

Our study can be characterised as *practitioner-driven*. The research questions addressed in this study were mostly initiated by the 30 pairs of outsourcing organisations and their IT-service providers. The participating companies were primarily interested in warnings of a failure risk in an early stage of service delivery.

Next to the power of the controllable factors to discriminate between failure and success the participating companies in the study were interested in the odds for such outcomes. In other words, how much will the chance on a successful deal alter if a critical controllable factor is adapted? We used logistic modelling to answer this question.

All in all, we addressed the following three research questions in this paper:

1. Which controllable success determinants discriminate between successful and failed IT-outsourcing deals?
2. How much will the chance on a successful deal increase if a critical controllable factor is changed?
3. Which rigid factors discriminate between successful and failed IT-outsourcing deals?

An experimental design model was constructed to assess the impact of the controllable and rigid factors on the chance of success of an outsourcing deal.

To determine whether an IT-outsourcing project is a success or failure, sounds easy but is in fact complex. Clearly, success and failure are context-dependent concepts. If, for example, time-to-market is important and the project suffers from a 5 percent time-overrun, the project might be called a failure, regardless of its cost and functionality performance. To overcome these complexities we decided to measure the perception of success of both outsourcer and vendor and not to invent some quantitative measure of success based on plan accuracy or something else.

The rest of this paper is organised as follows. First, in Section 2 we present a review of the literature related to our work. Next, in Section 3 we go into our research process. We explain what made us to decide on the nine controllable success determinants and six rigid success factors of which the power to discriminate between successful and failed cases has been tested. In this section we also describe and motivate our approach chosen to measure success and failure as the perception of the involved parties. In Section 4 we explain how we collected the data of our study. We show that our sample is a representative sample of the Dutch IT-outsourcing practices during the research period (2007–2010). We also show that the size of our sample (30 outsourcing deals from a target population of 700 deals) is acceptable for the statistical tests we wanted to carry out. In Section 5 we show how we tested the discriminative power of the controllable success determinants and rigid success factors. In this section we also describe how we used logistic regression modelling to assess with what percentage the chance on success increases if more attention is paid to the critical controllable success factors. In Section 6 we discuss the findings presented in Section 5. We finish the paper with conclusions in Section 7.

2. Literature review of related work

There is an abundance of publications in scientific journals dedicated to determinants of outsourcing success. Before giving an overview of the work in this field we first pay attention to the views in literature regarding measuring outsourcing success. We fully agree to the statement of Schwarz that it makes no sense to look for determinants of success while the latter notion is still not properly defined [51].

Because of the many publications, also different terms are used for the same notion. In this paper we will adhere to the definitions and terminology introduced by Bergstra et al. [7,6]. In particular, we will use the term ‘sourcement’ to refer to the state arrived at after the act of transfer of a source. Further, note that we will use vendor, supplier, service provider and insourcer as synonyms. We will also use the terms ‘determinant’ and ‘factor’ interchangeable.

2.1. Views on measuring success or failure

To determine whether an IT-outsourcing project is a success or failure, sounds easy but is in fact complex. For example, Standish Group judges the success of an IT project by considering estimation accuracy of cost, time and functionality. However, their definition of success has serious drawbacks. We quote from [20]:

> Standish Group defines its success measure as a measure of estimation accuracy of cost, time, and functionality. In reality, the part of a project’s success that’s related to estimation deviation is highly context dependent. In some contexts, 25 percent estimation error does no harm and doesn’t impact what we would normally consider project success. In other contexts, only 5 percent overrun would cause much harm and make the project challenged. In that sense, there is no way around including more context (or totally different definitions) when assessing successful and challenged projects.

An early, and still often cited paper, defining success is Grover, Cheon and Teng’s work from 1996. These authors defined success of outsourcing in terms of the attainment of strategic, economic and technological benefits, and satisfaction, that need to be weighed against changes in transaction costs, flexibility, and conflicting objectives of the outsourcer [26]. They propose a general framework to judge success. They also acknowledge that softer factors as the relationship with the insourcer may play an important role.

Levina and Ross state that in 2003 the vendor’s perspective has hardly been explored, which is also our perspective many years later [46,43,41]. They investigated a single long-term successful applications management outsourcing engagement (but not over a longer timeframe). In our study we examined 30 IT-outsourcing engagements and followed them for a number of years. The 30 engagements are representative for the Dutch IT-outsourcing deals at the time of carrying out the research of in total 700 deals. Levina and Ross conducted 28 interviews; we carried out 253 interviews. The methodology of measuring success in Levina and Ross’ paper was to give a grade. In our case we asked multiple questions to both client and vendor measuring success or failure.

Dibbern et al. summarise in their survey of 2004 the state of affairs by identifying three ways to typify outcomes: satisfaction, expectations and their realization, and performance [17]. But, they conclude that work on the evolving definition and operationalization of outsourcing success is ongoing. Their summary gives soft factors a firm place under ‘satisfaction’. The same was done by Dahlberg and Nyrhinen in 2006 [13]. Dibbern et al. also warn that the three outcome types they distilled – satisfaction, realization of expectations, and performance – could well be dependent on one another. This is confirmed in a study by Rouse in 2006 [50] proposing to explain outsourcing purchasers’ dissatisfaction.

Table 1
The eight dimensions of IT outsourcing success as found by Schwarz [51].

1	The client acquires additional capabilities.
2	Objectives are achieved on time.
3	The client receives financial benefits.
4	Quality is improved.
5	The arrangement allows for flexibility to accommodate changing circumstances/needs.
6	The partners develop a mutually beneficial relationship.
7	Mutual satisfaction is achieved.
8	Service-level agreements are met or exceeded.

We observe that the success-failure discussion turns out to be far from settled. In 2008 Cullen, Seddon and Willcocks argue in favour for a more contextual and idiosyncratic view on success on the basis of an extensive study of both the literature and 49 cases [12]. Lacity, Kern and Willcocks, in their 2009 survey of the literature, give insight in the operationalization too [36]. They found that researchers started to differentiate between three units of analysis: the organisation, the IT-function and the specific project assessed, and also examined various metrics and variables related to these levels as stock price performance, increased service, and quality. Schwarz, in 2014, after arguing that it makes no sense to look for determinants of success while the latter notion is still not properly defined, synthesized eight dimensions of success on the basis of a thorough review of the literature, interviews of experts using questionnaires, and statistical analysis [51]. She also firmly brings in the vendor perspective. Her conclusion is that outsourcing success is a first-order reflective construct characterised by the eight dimensions shown in Table 1.

We feel that in Schwartz her work the framework of Grover et al. is still visible. Schwartz presents a well-argued refinement with more equal roles for client and vendor, recognizing the intricate interplay between the two, with more attention to the specific situation, IT-function and organisational level. It is again a step away from a general but narrow metric.

Given the material in the literature, we developed our own framework to measure success or failure. In Section 3.3 we present this framework. In Section 5.1 we show how we used the framework to divide the 30 cases of our study in success stories and failures.

2.2. Literature review on success determinants in IT-outsourcing

In this section we give a summary of the research into determinants of IT-outsourcing success during the last decades. In presenting determinants of IT-outsourcing success the authors whose work we discuss do not distinct between controllable success determinants and rigid success determinants, nor do we in our literature review.

An early, and still often cited paper, dealing with determinants of IT-outsourcing success is the literature survey of Dibbern, Goles, Hirschheim and Jayatilaka from 2004, which builds on, by and large, a super set of the material we based ourselves on in selecting the success determinants to be tested [17]. However, where our focus is on pinning down concrete factors and best practices for managing outsourcing projects, the survey follows an epistemological course unravelling the underlying theories to position and characterise the academic effort and papers surveyed while at the same time discussing and multi-level categorizing the apparently important notions and findings, and finally revealing the open questions and issues in need of attention from the community. The survey presents findings on the implementation phase and the ways to typify and assess the “outcome” of an outsourcement. Outcome is used as wide category of the results, encompassing both success and failure in narrower scopes. They listed determinants of outsourcing outcomes, too. However, they mention important determinants like contract management, service quality, risk factors, relationship management, end-user support.

We provide a number of overviews of determinants of success, published in the past six years. For instance, Cram performed a meta-analysis on success factors for information systems outsourcing in 2009 [11]. His focus is on the specific activities that empirical studies suggest organisations should perform in order to increase the likelihood of realizing the economic, strategic, and technological benefits traditionally associated with outsourcing. The aggregated factors postulated by Cram are listed below.

1. Communication
2. Strategy
3. Business relationship
4. Finance
5. Management
6. Duration
7. Human capital
8. Knowledge sharing
9. Technology
10. Quality of service

Table 2
Highest priority is 1. All items with the same priority number are not further ranked within that priority. The ranked list of the eleven critical success factors according to their importance as presented by Hodosi and Rasu in [29].

Priority	Critical success factors [29]
1	Outsource only when it makes good business sense. Have a detailed plan including all resources, competencies, and costs for the next several years before signing the contract.
2	Determine what to outsource and what to keep inside the organisation. Measure the contract fulfilment. Maintain good communication between the organisations. Select a supplier that fits to your business culture and size.
3	Use flexible contracts and update regularly. Prepare the personnel of the outsourcing company for the new role. Make sure that the provider uses the standard tools and processes defined as part of the operational model. Do not outsource broken processes; improve them first.
4	The outsourcer shall undertake a due diligence on itself to understand, quantify, and qualify its outsourcing needs before starting with request for information.

Cram suggests that "those organisations that focus on a robust strategy, a strong business relationship, mindful financial management, constructive knowledge sharing and an excellent quality of service are likely to experience a more successful outsourcing experience."

Four other studies are comparable to Cram's survey in that they centre around a set of determinants. Smuts et al. gathered critical success factors from the literature and synthesized factors tailored for the software development life-cycle [52]. They did so in order to set up and manage a large project in South Africa, which is not detailed further in the paper. First, they gathered the following factors from the literature: delivery performance, contract management, relationships, staff management, cost management, service level agreements, maintaining control, communication, understanding the customer, technical expertise, and 'being flexible'. Second, they perused through questionnaires a specific software development project and on the basis of the findings from this they tailored, refined and mapped their initial set to the software development life-cycle.

Nasir and Sahibuddin harvested determinants relevant irrespective of project size, domain or country. They do not assess their findings empirically in a case study or from theory [44]. Notably, they mention 'proper planning' explicitly. Planning, such a basic element in any project, does not get much attention in the literature we perused. In the Outsourcing Handbook, drawing up a plan and then executing it is a recurring issue (Power et al. [48]). As we will show in Section 5.2.2 it turns out to be a strong predictor in our field study. Ali and Khan listed success factors in a way similar to Nasir and Sahibuddin discussed above in that they order factors on their frequency of occurrence in the literature [2]. This paper focuses on success factors from the vendor perspective. The results presented show a further shift in the literature from hard factors as technical capabilities and contract towards shared values, commitment and understanding. Hodosi and Rasu also collected critical success factors from the literature. Their initial 91 factors were grouped into 19 factors which were valued on the basis on questionnaires sent to managers at the CIO level. This led to a subset of 11 factors perceived to be relevant by these managers and a subset of less relevant factors [29]. Their findings are presented in Table 2.

Many empirical studies point at the importance of information sharing, cooperation and communication for outsourcing success, for instance, the study by Lee and Kim in 1999 [38] and the study by Grover et al. also from 1999 [26]. More recent is work by Babar et al. on establishing and maintaining trust [3]. In most papers the importance of good communication between the outsourcer and its service provider is emphasized, whereas the communication inside the organisation of the outsourcer and its service provider is given less or no attention. We found only one publication by Urbach in 2012 in which the author explicitly indicates the importance of good communication inside the outsourcer's and vendor's organisation for a successful IT-outsourcing project [53].

In papers where later phases of the IT-outsourcing life-cycle are in focus, knowledge management is mentioned as a key success factor, for instance in [40,1] and Willcocks et al. [58].

There is ample literature on research to the motives of the outsourcer to resort to IT-outsourcing; see for example (Baldwin et al. [4]). In their review of the IT outsourcing literature Lacity et al. mention 143 articles in total that deal with motivations for outsourcing [36]. So the issue can be recognized as a widely researched topic. However, we found no publication on research into the relationship between the motivation of the outsourcer and IT-outsourcing success; a subject that is treated by us.

Completely different, we did not find a single paper in the IT-outsourcing literature that deals extensively with the motive of the vendor to engage in an IT-outsourcing deal. A number of articles discuss the trend in IT-outsourcing of blurring organisational boundaries between outsourcer and vendor due to the desire of both parties to enter into a strategic partnership (business relationship) [5,26,45,39].

According to the IT-outsourcing literature review of Lacity et al. evidence suggests that the types of IT functions outsourced affect IT-outsourcing success [36, p. 136]. See for example the article of Grover et al. [26] in which they report the finding of a strong relationship between IT-outsourcing success and outsourcing of systems operations and telecommunication.

One can find an overwhelming number of articles that is focused on the effect of cultural differences between vendor and outsourcer on the success of an IT-outsourcing deal. See for example [29,32,47,59,42,16] and the aforementioned literature survey articles by Dibbern et al. and Lacity et al. These papers address differences in cultures of the countries that the stakeholders are embedded in, for instance the difference between culture in Germany versus India.

We only found a few papers on IT-outsourcing that implicitly or explicitly deal with the capability of the service provider to imagine himself in the position of his client (vendor's empathy) [46,43,41]. Also Smuts et al. clearly refer with 'Understanding the customer' to vendor's empathy [52].

We tested the impact of hiring consultancy support on the chance of a successful deal. This is a rather under-exposed area of research according to the literature study we carried out. We just found a single article by Gable from 1996 that deals with the link between engaging external consultants and client success [22]. His paper is about the success in engaging an external consultant to assist with the selection of a computer based information system. In each researched case an external consultant was hired for assisting in the selection process of a computer based system.

From our literature study it had become clear that the focus of research into determinants of IT-outsourcing success usually is on factors that explain the result of an outsourcing deal in retrospect [11,27,2,52,29,34]. That knowledge is valuable as a warning sign before closing a deal in future cases, but is not helping in corrective control during the service delivery phase. We did not find much research that addresses the issue of early warnings, however. We just found one paper by Huber et al. that discusses the need to establish the positive link between controls and goal achievement for outsourced software developments projects [31]. Huber et al. discuss formal and informal control in IS-outsourcing from a control theory perspective. We quote: "The process of translating high-level project goals into specific formal controls becomes crucial for success or failure of IS outsourcing projects." Clearly, searching for specific formal controls in IT-outsourcing is equivalent to searching for controllable success determinants.

3. Research approach

3.1. Selection of controllable success determinants to be tested

In Section 1 we informed the reader about the nature of our study. We carried out a longitudinal study to gain insight in the factors that are determining whether a closed IT-outsourcing deal is going to be a success or a failure. In order to be able to timely take corrective measures in case of a failure risk the focus of the study was particularly on controllable success determinants. We define a controllable success factor as a factor that can be changed by the outsourcing company and/or its service provider in an early stage of the service delivery phase in order to reduce the chance on failure. After much debate with the participating companies in the study, it was decided to focus attention to nine controllable factors. In this section we substantiate why we selected these nine factors and not others. The selection process can be best described as a mix of factors discussed in other literature studies, input from the participating organisations, and our own real-world experience with outsourcing. The first author worked many years as an IT-sourcing consultant. He distilled from both continuous reviews of the literature and extensive practical experience a selection of success determinants spanning the entire life-cycle of outsourcing projects: from decision making, to service delivery, to follow-up or backsourcing [14]. For a detailed presentation of these factors we refer to the Handbook of Network and System Administration [15]. These factors are partitioned in four related phases: the decision and vendor selection phase, the control phase (during service delivery), a potential backsourcing phase, and the follow-up sourcing phase. The sets of factors are related in the sense that they are situation-specific for a smaller set of more abstract themes like: planning, business case, knowledge, assets, and quality. For the control phase, the focus of our study, we employ his factors. They are as follows.

1. Planned approach of the sourcing process
2. A positive business case for outsourcing
3. A positive business case for insourcing
4. Careful personnel transfer
5. Assets transfer
6. User change support
7. A solid contract
8. Clear financial agreements
9. Performance management
10. Retention of expertise

Also Cram's survey of success factors [11] for information systems outsourcing published in 2009 (see Section 2.2) and the eleven critical success factors collected by Hodosi and Rasu from interviews of managers at the CIO level (see Table 2 in Section 2.2) did serve as an important source of inspiration in the selection process. Given the results of our literature research, our own experience and knowledge and experience of the participating organisations in total nine controllable (potential) success factors to be tested were decided upon. They are working according to the transition plan, managing the business case of the outsourcer, managing the business case of the service provider, transfer of staff, transfer of assets,

demand management, retention of expertise in the organisation of the outsourcer, communication inside the organisation of the outsourcer, and communication inside the organisation of the service provider.

The first factor selected was 'Planned approach of the sourcing process'. Although planning, such a basic element in any project, does not get much attention in the literature as a success factor, it was generally felt that proper planning and sticking to it during the implementation phase of an IT-outsourcing deal is detrimental for success. In the Outsourcing Handbook, drawing up a plan and then executing it is a recurring issue (Power et al. [48]). Working according plan can be found back as the fifth factor 'Management' in Cram's list of success factors. Also Nasir and Sahibuddin mention 'proper planning' explicitly as a success factor.

Researchers have indicated 'managing the business case' during the IT-outsourcing implementation process as an important factor for outsourcing success. We refer to the work of Hodosi and Rasu [29] and Cram's list of success factors as presented in section 3.2. In Table 2 from Hodosi and Rasu the critical success factor 'Outsource only when it makes good business sense' has priority 1. Cram suggests that "those organisations that focus on a robust strategy and a strong business relationship are likely to experience a more successful outsourcing experience." This was in line with the general feeling of the representatives of the outsourcing companies and the service providers. It therefore was decided to include 'managing the business case of the outsourcer' and 'managing the business case of the service provider' in the list of controllable success factors to be tested.

Controlled transfer of staff and assets are seen by several researchers as important factors for outsourcing success. See for example 'Human capital' and 'Technology' on Cram's list of success factors. Also Hodosi and Rasu refer to controlled transfer of staff by including 'Prepare the personnel of the outsourcing company for the new role' in their list of Critical Success Factors. The representatives of the participating outsourcing companies and insourcing service providers agreed with these findings from the literature study. In the transfer of personnel ample attention should be given to cultural differences, communication, handling resistance, the workers council, and more. Transferring assets to a different environment may be a risky endeavour and should be managed carefully. For example, planning for back-up facilities just in case serious problems arise. Therefore 'transfer of staff' and 'transfer of assets' were listed as controllable success factors to be tested.

Almost all researchers have explicitly or implicitly indicated 'demand management' as a key factor for outsourcing success. Sourcement success in the long run is granted only when the outsourcer remains sufficiently able to manage the service provider that takes over the outsourced IT functions. Contract management, management of the financial agreements, and performance management form part of demand management. Cram suggests that "those organisations that focus on a strong business relationship, mindful financial management, and an excellent quality of service are likely to experience a more successful outsourcing experience." In Table 2 of Hodosi and Rasu the critical success factors 'Measure the contract fulfilment' and 'Make sure that the provider uses the standard tools and processes defined as part of the operational model' are listed with priority 2 and 3 respectively. Smuts et al. finally, gathered 'contract management', 'delivery performance', 'cost management', service level agreements', 'maintaining control', and 'being flexible' as critical success factors from the literature [52]. Accordingly, there was general agreement to include 'demand management' in the list of controllable success factors to be tested.

'Retention of expertise in the organisation of the outsourcer' was selected as controllable success factor to be tested for several reasons. First, it was generally felt that retention of expertise is important for the outsourcer in order to be able to manage the service provider during the service delivery phase. Next, without retention of expertise the ousourcer loses the free option of back-sourcing or changing from service provider at the end of the contract period and becomes permanently dependent on the one provider that he has (Willcocks et al. [58]).

Many researchers have pointed to the factor 'communication' as an important factor for outsourcing success. See for example the work of Dibbern et al. [17]; Lacity et al. [37]; Goo and Huang [25]; Grover [26]. Communication is an aspect of relationship management, i.e., paying close attention to the position of all different stakeholders who are involved in the IT-sourcing process. This finding from our literature study was generally endorsed by the representatives of the outsourcing companies and the service providers who participated in the meetings, in which the selection of the success factors to be tested was on the agenda, and it was therefore decided to pay specific attention in the set-up of the questionnaires to communication with the stakeholders inside both the outsourcer's and vendor's organisation. Of course, also the communication between outsourcer and service provider is of great importance for successful teamwork. This aspect of communication, however, was already covered by the success factor 'demand management'.

Below, we provide a detailed description of the nine controllable success determinants of which the power to discriminate between successful cases and failed cases has been tested by us. Each determinant of success corresponds to a condition that must be satisfied sufficiently to safeguard the successful progress of the deal concluded between the outsourcer and his service provider(s). For the ease of reference we have numbered them D1–D9.

D1 Working according to the transition plan: After the outsourcer and service provider have decided to close an outsourcing deal, the transition of staff, software, hardware, or knowledge transfer can commence. It is of utmost importance that the transfer takes place in a controlled way according to a well-defined transition plan in time phases, which has been agreed upon by both outsourcer and service provider. Usually, the outsourcer takes the lead in creating the transition plan. The outsourcer and service supplier should adhere to the transition plan during the sourcing phase to safeguard the successful implementation of the IT-outsourcing arrangement.

D2 Managing the business case of the outsourcer: Success in the long run is granted only when the outsourcer continues to use his business case during the execution of the IT-outsourcing arrangement to evaluate the outsourcing process and the service delivery of the provider.

D3 Managing the business case of the service supplier: The business case of the service supplier is highly symmetrical to the business case of the outsourcer and should serve as a frame of reference during the service delivery phase.

D4 Transfer of staff: It is reasonable to assume that working according to the transition plan will be beneficial to a successful transfer of personnel. However, it is not a sufficient condition for that. Even without a transition plan or not sticking (fully) to it, the transfer of personnel may turn out well. It therefore makes sense to distinguish D4 apart from D1.

D5 Assets transfer: It is reasonable to assume that working according to the transition plan will be beneficial to a successful transfer of hardware and software and accompanying knowledge. However, it is not a sufficient condition for that. Even without a transition plan or not sticking (fully) this plan, the transfer of assets may turn out well. It therefore makes sense to distinguish D5 apart from D1.

D6 Demand management: Sourcement success in the long run is granted only when the outsourcer remains sufficiently able to manage the service provider that takes over the outsourced IT functions. This is called demand management. Contract management, management of the financial agreements, and performance management form part of demand management. An outsourcing contract is valid during the lifetime of the sourcing deal, and in the ideal case it remains in the filing cabinet during that period. A transparent cost price calculation is of paramount importance for a healthy outsourcing relationship [54,19]. Anything that is not explicitly included in the contract and the SLA is charged separately, so it is important to manage the concluded financial agreements carefully. Though we will find in Section 5 that D6 is an important discriminative factor, we do not discuss the question what is a good contract. A theoretical effort to do so has been put forward by Fitoussi and Gurbaxani in their work on properties of IT Outsourcing Contracts [21].

D7 Retention of expertise: After the transition, so during the service delivery phase, it becomes important for the outsourcer to retain his expertise. Without retention of expertise he loses the free option of back-sourcing or changing from service provider at the end of the contract period and becomes permanently dependent on the one provider that he has (Willcocks et al. [58]). Moreover, the expertise is needed to manage the service provider during the service delivery phase.

D8 Communication inside the outsourcer's organisation: for the outsourcer it applies that paying close attention to the position of all different stakeholders who are involved in the IT-sourcing process is of the utmost importance.

D9 Communication inside the vendor's organisation: also for the service provider it holds that paying close attention to the position of all different stakeholders who are involved in the IT-sourcing process is of the utmost importance.

We would like to stress that our list of controllable success factors should not be viewed as the ultimate list of factors that are signals for success during the service delivery phase of an IT-outsourcing project. Such a list simply does not exist. We think, however, that with our list we captured the most important factors that give a warning signal in case of an upcoming failure and can be influenced to decrease the risk of failure.

3.2. Selection of rigid success factors to be tested

Although the special interest of the participating companies was in testing the discriminative power of controllable success factors it yet was decided to add 6 rigid success factors to the list of controllable success determinants to be tested. The main reason for this was the wide range of publications in scientific journals dedicated to the potential effect of rigid factors on the outcome of an IT-outsourcing deal.

As discussed in Section 2.2 there is ample literature on research to the determinants of IT-outsourcing decisions: the motive of the outsourcer to engage in an IT-outsourcing deal. We therefore decided to test the power of the outsourcer's motive to discriminate between successful outsourcements and failed outsourcements. We decided to test the motive of the service provider likewise, although we hardly found any interest for this issue in the literature. We were rewarded for this decision: our study yielded a remarkable result. If the vendor's motive is long-term strategic, the chances for a successful deal increase, but if the motive is short-term the chance of success will decrease (see Section 5.3.2).

Many researchers have examined whether cultural differences between the outsourcing company and its service provider affects IT-outsourcing success. Most papers address differences in cultures of the countries where the outsourcing company and its service provider are embedded in, for instance the difference between culture in Germany versus India. Given the wide range of scientific publications dedicated to this issue we decided to examine whether the difference in culture play a role in predicting the chance on success of an IT-outsourcing deal. In doing this, we took a different view, however. In our study we tested the hypothesis whether the match or mismatch of management cultures affects the chance on success of the sourcing deal; and this turned out not to be significant.

In our literature study we were triggered by papers on IT-outsourcing that implicitly or explicitly deal with the capability of the service provider to imagine himself in the position of his client (vendor's empathy) [46,43,41]. Inspired by this it was decided to examine whether vendor's empathy affects the chance on success of the sourcing deal. We found vendor's empathy to be beneficial.

Table 3
Five different sources of information to measure success perception of both client and supplier.

#	Outsourcer:
1	The extent in which the outsourcer has reached his specified outsourcing objectives and his unspoken expectations
2	The extent of satisfaction of the outsourcer with the quality of the services
3	The trust of the outsourcer in further collaboration with his service supplier
	Supplier:
4	The extent in which the vendor has reached his insourcing objectives
5	The trust of the service supplier in further collaboration with his client

From the IT outsourcing review of Lacity et al. we learned that evidence suggests that the types of IT functions outsourced affect IT-outsourcing success [36]. It was therefore decided to examine in our study whether the type of outsourced work plays a role in predicting the chance on success of an IT-outsourcing deal. Surprisingly, we did not find any evidence for this.

Finally, it was decided to test the impact of hiring consultancy support on the chance of a successful deal. Apart from the outsourcing companies and service providers also IT-sourcing consultants were involved in the longitudinal study: half the cases with, and half without sourcing consultants. Therefore, we could measure their impact on the outcome of the deal. Our literature research had yielded just one single article by Gable from 1996 that deals with the link between engaging external consultants and client success [22]. However, his paper is about the success in engaging an external consultant to assist with the selection of a computer based information system, whereas in our case, the external people were specifically sourcing consultants, not to select an IT-system, as intermediaries between client and vendor in case of IT-outsourcing. We therefore conclude that research to the affect of hiring consultancy support on IT-outsourcing success is a rather under-exposed area of research. Surprisingly, hiring sourcing consultants turned out to be contra-productive: it lowered chances of success (see Section 5.3.6).

In summary, the following 6 rigid success factors to be tested were selected:

1. the motive of the outsourcer to engage in an IT-outsourcing deal;
2. the motive of the service provider to engage in an IT-outsourcing deal;
3. the match of the organisation cultures of the service provider and his client;
4. the type of outsourced work;
5. vendor's empathy capability;
6. hiring consultancy support.

3.3. Measuring IT-outsourcing success

In line with the views of many authors in recent literature (see Section 2.1) we adopted the approach to measure success and failure as the perception of the involved parties. We integrally measured IT-outsourcing outcomes including mutual, subjective and case specific aspects in concert with more abstract strategic, technological, economic and functional benefits. We considered five different sources of information to measure the perception of the outsourcer and his supplier about the success of the deal. In Table 3 we summarise the sources of information. As can be seen, we take both client and vendor into account.

Our approach is contextual because we look at the objectives met per case. In our case study both outsourcer and vendor were asked what their three key objectives were at the start of the project, and around two years later they were asked to what extent these were met. We verified whether these objectives given per case per stakeholder could easily be matched with the dimensions of Schwarz. This match is easily found for the outsourcer objectives. The vendor objectives are less simple. Schwartz states [51, p. 161] that while some of the dimensions are focused on only one party, i.e., the outsourcer, the dimensions 'The partners develop a mutually beneficial relationship' and 'Mutual satisfaction is achieved' demonstrate the inherent level of dependence within an outsourcing arrangement and the importance of developing a partnership to mitigate that risk. We feel the vendor perspective is than still limited. Vendor may have their own idiosyncratic goals as name recognition, or access to the market. Schwartz her definitions of success through her eight dimensions, are a step forward, but still miss out when it comes to encompassing such objectives. Compared to Schwartz her work our approach takes the extra step of finding a balance between client and vendor.

4. Data collection process

4.1. Participating companies

Sixty business partners participated in our research in an effort to gain insight in the discriminative power of determinants of IT-outsourcing success. The business partners delivered via interviews relevant data about the 30 IT-outsourcing cases. In each case at least two business partners are involved: the outsourcing party and the vendor. In half the cases (15 to be exact) also an IT-outsourcing consultant was involved.

Table 4
The involved clients, suppliers and subcontractors in the 30 deals.

Client	Supplier	Subcontractor(s)
Achmea (insurance)	Atos	Getronics, KPN
Aegon (insurance)	HCL (India)	
AKZO Car Refinishes	Atos	
BNG Bank	Ordina BPO (now Centric FSS)	
CWI (employee insurance)	Logica	IBM, Atos, Tele2
Da Vinci (school)	Goyello (Poland)	
Delta (utilities)	Atos	Accenture
Municipality of The Hague	Getronics	
Municipality of Almere	Atos	
Port of Rotterddam	Getronics	
Amsterdam University of Applied Sciences	SaNS	
ING	Tata CS (India)	
Innopay	Maxcode (Romania)	
ISS (facilities)	Interaccess	Logica, KPN
Ministry of Education, Culture and Science	Capgemini	
Ministry of Social Affairs and Employment	SSO-ICT, Ministry of the Interior and Kingdom Relations	
Ministry of Housing, Spatial Planning and the Environment	Getronics	Tele2, HP
Ministry of Health, Welfare and Sport	SSO-ICT, Ministry of the Interior and Kingdom Relations	
NRC (newspaper)	Conclusion	
NS (Dutch railways)	CSC	
Océ (now Canon)	Atos	
ProRail (rail infra)	GID	
Province of Northern-Holland	Siemens	
Rabobank	Ordina	Cognizant (India)
UVIT case 1 (health insurance)	Atos	
UVIT case 2 (health insurance)	Atos	
UWV-KCC (employee insurance)	KPN Newtell essence	
UWV (employee insurance)	Logica	
VKA (IT consulting)	van Veen	
Youropi	Goyello (Poland)	

In [8, p. 22, Table 1.3] all the clients and suppliers working with us in our case study are listed in Dutch. For the sake of ease we translated the table in English. In Table 4 all the client organisations, their IT-suppliers and if present also their subcontractors are shown. It is a mix of local, regional and federal government and private deals. Then there is insurance, banking, utilities, offshoring, local outsourcing, and more. There are large and small players, public–private, private–private, and public–public deals. In this section we show that the 30 engagements are a representative sample of acceptable size for the entire Dutch outsourcing market at the time of our investigations.

Data about the deals were collected via 253 interviews. In total it took several FTEs for making the appointments, travel time, lost time due to fairly frequent no-shows of participants, interview time of a few hours, transcription of the interviews, completing the questionnaires and making the data fit for statistical analyses. In this section we will provide global information about the types of inquiry forms used. The collected data enabled us to study a large number of outsourcing contracts from both the perspective of the outsourcing company and its vendor. The scope encompasses onshore as well as near- and offshore outsourcing. However, it remains limited to outsourcing deals closed in the Netherlands. The 30 IT-outsourcing deals were followed by us during different phases of the contract period. An introduction including first results can be found in Bergstra et al. [8].

The participating companies were involved in completing the questionnaires. According to the role they played three categories can be distinguished: outsourcers, service providers and IT-outsourcing consultants. In most cases the vendor–client relationship is of the type single vendor–single client. In seven cases the relationship was of the type multiple vendors–single client. Nevertheless, in these cases there was always one main contractor. In half the cases a consultant was involved in the decision phase of the outsourcing process to support the outsourcer in negotiating the best deal with its vendor(s).

All parties were familiar with the questions, as the questionnaires were discussed during preparatory meetings. Therefore, in principle, there was no need for trial rounds but to rule out any remaining misunderstandings regarding the meaning and purpose of the questions a trial inquiry was carried out. The questions of the questionnaires were in many cases answered by representatives from three different echelons, independently from each other: strategic management, tactical management, and operational management.

4.2. Questionnaires

Each outsourcer-vendor pair was asked to answer the questions of five different questionnaires. Questionnaire 1 was answered by outsourcer and vendor together, while questionnaires two and four and respectively three and five were answered separately by outsourcer and vendor. The structure of the questionnaires addresses many aspects of outsourcer and vendor

perceptions about the cooperation. The answers to our questions are a mix of Likert scale, nominal scale and open-ended answers. The five questionnaires cover a large part of the outsourcing life-cycle.

Questionnaire 1 collects information regarding the IT sourcing preparation phase. The questions are also meant to gather the characteristics of the cases, e.g.: type of work, sector of the economy, contract duration. Moreover, its purpose was to gain insight in the maturity of the outsourcer and vendor to start the business partner selection process.

Questionnaire 2 collects information from the outsourcer regarding the vendor selection process. The questionnaire also contains questions with respect to the carefulness of the transition of employees, software, hardware and knowledge to the vendor. Finally, the questionnaire contains a number of questions to measure the satisfaction of the outsourcing company with the performance of the vendor shortly after the work was started.

Questionnaire 3 collects information from the vendor regarding the client selection process. The same types of questions as asked to the outsourcer are asked to the vendor company. Also the vendor is asked to assess the opinion of its business partner (the outsourcer) about a number of matters, to measure the vendor's empathy.

Questionnaire 4 collects information from the outsourcer regarding his opinion about the execution phase for the second time at a later moment. There are also questions regarding the confidence of the outsourcer in further cooperation with the vendor.

Questionnaire 5 collects information from the vendor regarding his opinion about the state of affairs during the execution phase for the second time at a later moment. Finally, a number of rather open questions are posed. For example the vendor is asked what he expects to be happening in the future.

For a complete overview of all the questions of all the five questionnaires we refer the interested reader to [56]. All the data can be found on-line as well as a codebook [49,55]. Also facsimiles of the original questionnaires are available to the interested reader [57].

4.3. Size and representativeness of the sample

Size We strived for a sample size of at least 30 outsourcing deals from a target population of 700 deals. This is 4.3% of the total. Intuitively, in predicting the results of the election of parliament members a sample size of 1 percent of the voters population is generally accepted if the sample is sufficiently representative. Apart from this general guideline, there are also more refined reasons why our sample is of an acceptable size.

The Central-limit theorem, one of the most important theorems in statistics, tells us that if nothing is known about the form of the distribution of a population (apart from that it has a mean and a finite variance) the mean of a random sample from it will be approximately distributed as a normal variate with mean equal to the population mean and variance equal to that of the population variance divided by the size of the sample [30, p. 200]. The importance of the Central-limit theorem for practical applications is that it can be used to carry out statistical tests albeit the population distribution is unknown. A sample size of 25 to 30 is generally large enough to approach the normal distribution adequately. If we know more about the population distribution it can be lower. For instance, if it is symmetric, unimodal, and of the continuous type, a sample size of 4–5 may be adequate. If the population distribution is approximately normal, a sample size of 2–3 might even be used [30, p. 202].

In the sequel, we will carry out a number of statistical tests, such as χ^2-tests of independence. For such tests, one can calculate its so-called power, which is the probability of rejecting the null hypothesis when it is not true. A type II error is the reverse of the power: it is the probability of accepting a hypothesis when it is false. A power of 80% is a convention proposed for general use [10] just as a significance level of 5% is a general convention (also called a type I error).

We use the R-package called pwr which provides power analysis functions along the lines of Cohen [9]. In Fig. 1 we show the power curves of the χ^2-test for effect size of 0.5, 0.3 and 0.1. Those effect sizes are suggested by Cohen [10, p. 157] for a 2×2 contingency table as a first idea of effect size. On the horizontal axis we put the number of observations and on the vertical the power of the significance test for three effect sizes. The vertical line represents 30 observations, i.e., 30 outsourcing engagements. As can be seen, for a large effect size the power of a χ^2 test is around 80%. Of course, the larger the sample the better from a statistical point of view. But as can be seen from Fig. 1, its power will slow down so that for a bit more power an enormous amount of extra work is necessary for collecting data from many more cases. If the effect sizes are medium or small, 30 observations is not enough to obtain a power of 80%.

In our case we are looking for factors with a strong discriminative power, and thus we use the by Cohen suggested largest effect sizes at the 5% significance level for various tests. We are looking for factors that obviously contribute to the success or failure of an outsourcing deal, not whether an alternative treatment might have a bit less contra-indications (which is a small effect size). Since we are dealing with ordinal data, Cohen calculated that 26 observations is enough for a χ^2-test [10, p. 158, Table 2]. But also for other tests relatively low sample sizes around the 30 are suggested. We started with much more engagements of which we chose 35 for our longitudinal study. Five of the engagements had not enough interviews so these were left out. For our analyses 30 data-rich cases remained. Later on in this article we will show that our assumptions were often adequate: we will carry out a number of post hoc power analyses showing that 30 cases turned out to be acceptable.

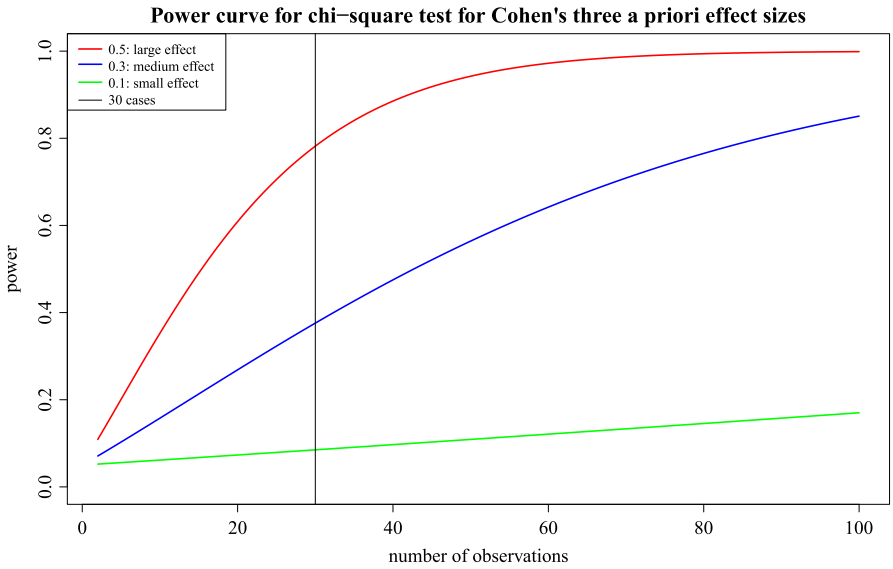

Fig. 1. The power curves for a χ^2-test of a 2×2 contingency table with Cohen's three a priori effect sizes.

Of course, if a factor is not significant, the effect size is likely to be small as well. This implies a low statistical power. In order to distinguish between a type II error and true insignificance the sample size should be increased. For small effect sizes, this implies an increase of the sample to hundreds of observations. For this study, insignificance is not the main focus. We are searching for significant factors that can help practitioners make the difference between failure and success. Later on we will show that an insignificant factor indeed needs hundreds of cases to know with sufficient power whether it is indeed insignificant and has a small effect or that there is no effect at all.

Finally, the population of IT-outsourcing deals is not easily accessible as a whole: information is scattered among many parties. It would have been highly impractical and prohibitively expensive to commit to much more pairs of companies involved in IT-outsourcing in those years. Even if one had tried to do so there would have been no certainty that all parties had been willing to cooperate. On top of that, an a priori power of 80% looked acceptable at 30 cases, and would only increase drastically with more than three times as much cases.

Representativeness The business partners delivered via interviews data of 30 IT-outsourcing cases that roughly took place in the years 2007, 2008, 2009 and 2010. In those years some 700 comparable deals were closed in the Netherlands. They were presented in the overviews of a well-known Dutch professional weekly on IT called Automatisering gids, say the Dutch Computer Weekly. One of the editors of this journal (Rolf Zaal) is an outsourcing watcher who gathered and compiled the material that we will use to show that our sample is representative [60–63].

It is important to know whether the sample is a representative cross-section of the Dutch IT outsourcing population. The representativeness of a sample is an important factor in determining how useful the sample is. In case of a biased sample we cannot make probabilistic statements about the entire population. In choosing the sample we proceeded as follows. First, the research team made the outsourcing community aware of the research project through many channels and invited them to join. Next, the team carefully picked a set of outsourcing organisations both small and large measured in the number of outsourcing deals. IT-service providers, domestic, near-shore, and offshore deals were chosen to participate. Also, IT-mediators were represented, since they are involved in sourcing as well. Next to that, we verified that the thus resulted sample is a reasonably representative sample out of the population of Dutch IT-outsourcing deals in the years 2007–2010. Thereto, we carried out three goodness-of-fit tests, which are described below.

First, we took the sectors of the Dutch economy that are represented in our sample. See Fig. 2 for a comparison of the distribution of the participants over the sectors of the Dutch economy with the benchmark compiled from data published by the Dutch Computer Weekly. We carried out a goodness-of-fit test to test the null hypothesis that the observed sample distribution does not differ from the distribution compiled from the benchmark data. We found the chi-square statistic of 17.18, which has the chi-square distribution with 14 degrees of freedom. Let $\chi_{.95}$ denote the 95th percentile of the χ^2 distribution. For 14 degrees of freedom $\chi_{.95} = 23.68$. Since the observed value of 17.18 is notably smaller than 23.68 we accept, at the 5 percent level of significance, that the observed sample distribution does not differ from the distribution compiled from the 700 deals published in the Dutch Computer Weekly. So the sample is representative for economic sectors of the Dutch economy where the 700 deals are carried out.

As an example we will calculate the power of the test we just carried out. Cohen suggested about 30 observations for it. The measured effect size is 0.76, which is even a larger effect size than Cohen's suggestion of 0.5 for this type of test.

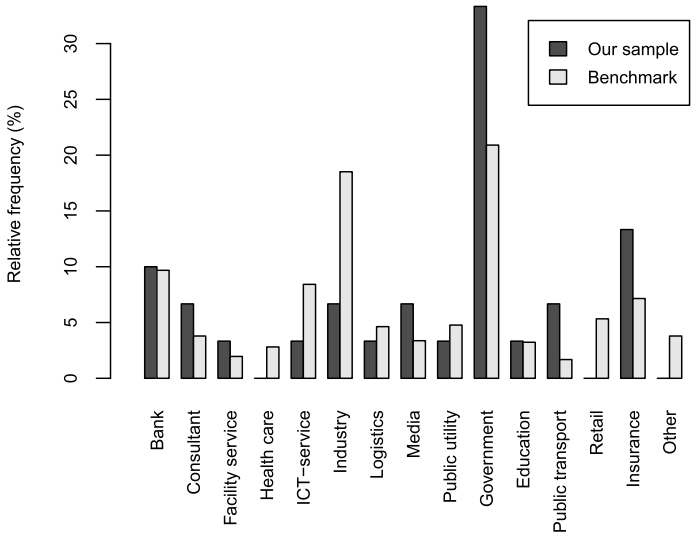

Fig. 2. Composite bar chart with a comparison of two relative frequency histograms re sectors of the Dutch economy: one compiled from our data and one compiled from the data published in the Dutch Computer Weekly.

Its power is 77%, so indeed in the order of about 80%. So the chance of making a type II error is acceptably low. If we had used 32 cases, the power would have been 80%, so the 30 cases are not too far off here.

Double check We used the chi-square goodness-of-fit test to determine whether the observed sample frequencies differ significantly from the expected frequencies specified in the null hypothesis. One of the assumptions of chi-square test is that the expected frequency counts in each category (the so-called theoretical frequency) are at least 5. Clearly, this assumption is not satisfied in our case, so there is a chance that the test is not accurate. Indeed, given the sample of size 30 and a total of 15 categories it is simply not possible to meet this test assumption for all categories. To be on the safe side, we carried out a double check with a test that does not need such assumptions. Appropriate non-parametric alternatives are the Wilcoxon Signed-Rank test or the Sign test. We used the Wilcoxon Signed-Rank test as this test is more powerful than the Sign test in our case. For the data we found the Wilcoxon Signed-Rank statistic of 60 ($p = 1$). Indeed when the critical value for the Wilcoxon test at 15 categories ($N = 15$) is 25 at the two-tailed significance level of 0.05, and our value is 60 which is much larger. So we can conclude that the difference between the probability distributions of our sample and that of the benchmark compiled from data published by the Dutch Computer Weekly is very unlikely to occur when the two distributions differ significantly from each other. In terms of the p-value, we found it to be close to 1. We would have rejected the null hypothesis when the p-value would have been less than the significance level of 0.05, but this clearly is not the case. So our χ^2 test is not leading to erroneous results.

Next, we compared the distribution of the contract durations of the sample with those of the 700 deals described in the Dutch Computer Weekly. Fig. 3 tells us that contract terms of 2–3 years seem underrepresented in our sample. On the other hand short-term contracts of 0–1 years seem overrepresented. The average contract duration of our cases is 4.53 while the benchmark average is 4.58. This difference is almost negligible. We carried out a goodness-of-fit test to test the null hypothesis that the observed sample distribution does not differ from the distribution compiled from the 700 durations published in the Dutch Computer Weekly. We found the chi-square statistic of 11.84, which has the chi-square distribution with 7 degrees of freedom. For 7 degrees of freedom $\chi_{.95} = 14.07$. Since the observed value of 11.84 is smaller than 14.07 we accept, at the 5 percent level of significance, that the observed sample distribution does not differ from the distribution compiled from the data published in the IT-weekly professional journal. So our sample is representative for the durations of the 700 deals.

We will also calculate the power of this test. The measured effect size is 0.66, which is a bit larger effect size than Cohen's suggestion of 0.5 for this type of test. Its power is 76%, so indeed in the order of about 80%. So the chance of making a type II error is acceptably low. If we had used 33 cases, the power would have been 80%, so the 30 cases are not too far off here.

Double check Also in this case the assumption of at least 5 is not satisfied. Given the sample of size 30 and 8 categories it is simply not possible to meet it for all categories. We used again the Wilcoxon Signed-Rank test and found a statistic of 21 ($p = 0.7422$). When we look in the table of critical values for the Wilcoxon test at $N = 8$ we see that the obtained value of 21 is much larger than the critical value of 4 at the two tailed significance level of 0.05. So we can conclude also from this test that the difference between the probability distributions of our sample and that of the benchmark compiled from data published by the Dutch Computer Weekly is very unlikely to occur when the two distributions differ significantly from

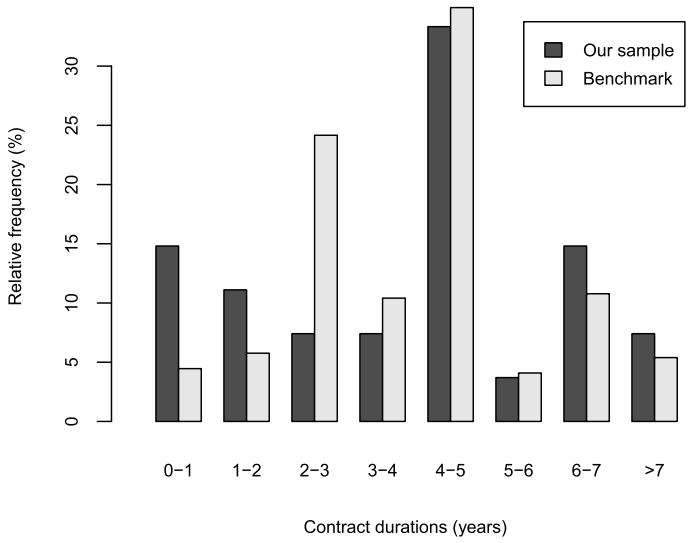

Fig. 3. Composite bar chart with a comparison of two relative frequency histograms re contract duration: one compiled from our data and one compiled from the data published in the IT-weekly professional journal.

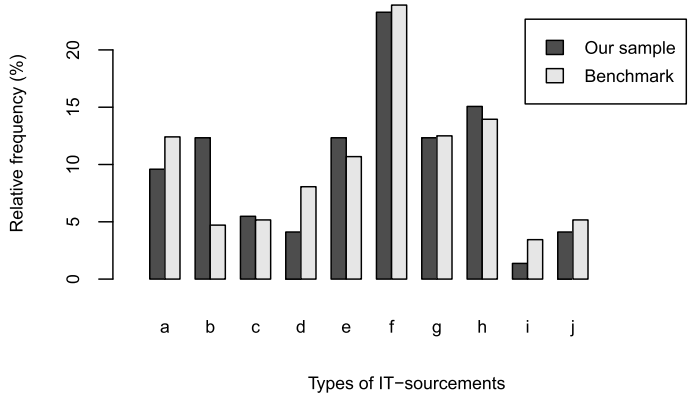

Fig. 4. Composite bar chart with a comparison of relative frequency histograms re sourcement focus: one compiled from our data and one compiled from the data published in the Dutch Computer Weekly. Legend: a. Application management, b. Application development, c. Application maintenance, d. BPO, e. Data communication and telephony, f. System and technical maintenance and management, g. Infrastructure, hosting and housing, h. End user services, i. General support and management services, j. Helpdesk.

each other. We would have rejected the null hypothesis when the p-value would have been less than the significance level of 0.05, but this clearly is not the case for $p = 0.7422$.

Finally, we focused on the type of the outsourced work. See Fig. 4, in which the distribution of the types of outsourced work derived from our data is compared with the benchmark distribution compiled from data published by the Dutch Computer Weekly. Fig. 4 shows that application development seems overrepresented in our sample compared with the benchmark. For the rest both distributions do not seem to differ much. We carried out a goodness-of-fit test to test the null hypothesis that the observed sample distribution does not differ from the distribution compiled from the data published in the Dutch Computer Weekly. We found the chi-square statistic of 12.23, which has the chi-square distribution with 9 degrees of freedom. For 9 degrees of freedom $\chi_{.95} = 16.92$. Since the observed value of 12.23 is smaller than 16.92 we accept, at the 5 percent level of significance, that the observed sample distribution does not differ from the benchmark. So the sample is representative for the type of outsourced work for the 700 deals.

The measured effect size is 0.40, which is somewhat smaller than Cohen's suggestion of 0.5 for large effect sizes. Its power is 56%. So the chance of making a type II error is present. In order to reach a power of 80% we need, given the degrees of freedom a sample size of 94 cases, which is not realistic in terms of cost/effort.

Double check Also in this case the assumption of at least 5 is not satisfied in our case for all categories. Given the sample of size 30 and 10 categories in total it is simply not possible to meet this assumption for all categories. We found the Wilcoxon Signed-Rank statistic of 23 ($p = 0.6953$). When we look in the table of critical values for the Wilcoxon test at $N = 10$ we see that the obtained value of 23 is much larger than the critical value of 8 at the two tailed significance level of 0.05. So we

Table 5
Cronbach's alpha for the success indicators.

Success indicator	Items in multi-item scale	Respondents	Cronbach's alpha
Outsourcer's success perception	10	27	0.9
Supplier's success perception	9	27	0.76

Table 6
Success indicator scores for all 30 deals.

Case #	Score outsourcer	Score supplier	Classification
2145	75.00	80.99	1
2146	63.89	40.44	0
2147	44.44	67.45	0
2149	55.56	66.29	1
2150	66.67	76.49	1
2151	69.44	78.35	1
2152	69.44	71.61	1
2153	72.22	49.26	0
2154	77.78	65.48	1
2155	46.43	79.09	0
2156	65.63	75.74	1
2157	66.67	56.25	1
2158	22.22	26.93	0
2159	55.56	29.84	0
2160	55.56	59.90	1
2161	38.89	32.14	0
2162	83.33	80.97	1
2163	61.11	68.68	1
2164	72.22	42.55	0
2165	69.44	70.60	1
2166	61.11	69.23	1
2167	69.44	63.10	1
2168	55.56	54.69	1
2169	50.00	40.36	0
2170	66.67	35.71	0
2172	52.78	38.10	0
2200	53.57	23.74	0
2201	52.78	66.22	1
2203	72.22	96.88	1
2204	61.11	66.15	1

can conclude that the difference between the probability distributions of our sample and that of the benchmark compiled from data published by the Dutch Computer Weekly is very unlikely to occur when the two distributions differ significantly from each other. In terms of the p-value, we found a p-value of 0.70. We would have rejected the null hypothesis when the p-value would have been less than the significance level of 0.05, but this clearly is not the case.

Conclusion In view of the tested null hypotheses, the power calculations, and the double checks, we conclude that our sample is a reasonably representative sample out of the Dutch IT-outsourcing population: the sample reflects the Dutch economic sectors, the duration of the deals and the types of work of the total population.

5. Results

5.1. Success perception measurement

Recall that we mentioned the five different sources of information we used to measure success perception of all parties already when we put them in context of the literature, see Table 3 in Section 3.3.

The respondents were asked to respond on a statement in terms of their own degree of agreement or disagreement. First, questions were asked to collect opinions formed in an early stage of the service delivery phase of the IT outsourcing life-cycle. These questions were asked once more later on during the service delivery phase to see how opinions might have changed in due course.

All answers to the questions are of the Likert scale type. A Likert scale question knows six possible answers: strong agreement, agreement, neutral, disagreement, strong disagreement, and non-applicable. We recoded these possible answers into 100, 75, 50, 25, 0 and NA. The questions that were answered with NA (non-applicable) were left out of consideration.

To verify the internal consistency of each set of questions, we calculated Cronbach's alpha reliability coefficient of internal consistency [24]. Table 5 shows the calculated alphas for the two success indicators.

Table 7
Partition of cases in successful cases and failed cases.

Result	Number of cases
Success	18
Failure	12
Total	30

Cronbach's alpha reliability coefficient normally ranges between 0 and 1, although there is actually no lower limit to the coefficient. According to the rule of George and Mallery a Cronbach alpha larger than 0.9 is excellent, between 0.8–0.9 is good, and between 0.7–0.8 is acceptable [23]. Below 0.5 the internal consistency of the items in the scale is unacceptable. From Table 5 we conclude the Cronbach's alpha reliability coefficients are high enough to have sufficiently trust in the measurements of both success indicators.

Table 6 shows the scores on the two success indicators for the 30 cases of our sample. We provide a case identifier in the first column. In the second and third column we list the scores of both outsourcer and supplier. Then we partitioned them in success (1) and failure (0) according to the following criterion:

1. If the scores on both success indicators are ≥ 50, then the case is called "successful";
2. All cases that do not satisfy condition 1 are called a "failure".

This leads to 18 deals that were perceived to be successful and 12 that were a failure; see Table 7. We found, therefore, a success rate of 60% and consequently a failure rate of 40% among the 30 deals. As the sample is representative for the 700 Dutch IT-outsourcing deals between 2007–2010, the success rate was about 60% at that time.

5.1.1. Validating perceptions

As indicated, we measured the perceptions of both outsourcer and supplier about the outcome of their deal via sets of questions giving scores for both vendor and client. Based on our scoring model we then inferred whether an engagement was a success or a failure. Apart from the already given considerations about the validity of our approach, we provide additional validation via the final outcomes of some of the deals that obtained media attention. We compared their final outcome as reported in the media with the outcome of our scoring model. Of course, not all cases were in the press and not all engagements were finalized at the time of writing. We checked four engagements in Table 4. In all cases our outcome coincided with the final outcome as reported on in the media.

Delta–Atos Delta is a utility company in the province of Zealand, delivering electricity, natural gas, water, cable (radio, TV, telephony, Internet). They engaged with ICT-supplier Atos Origin in a 7-year contract for ICT-outsourcing in 2006. This contract was prematurely terminated in 2012.[1] This does not sound as a success, although the word failure is not mentioned in the press. Based on our scoring model this was a failed engagement.

HvA–SaNS HvA is the Amsterdam University of Applied Sciences, and SaNS was a consortium developing the new student information system envisioned for five universities (see Table 4). The project started around 2004. Oracle and Atos Origin were implementation partners in this project. In 2009 one of the universities left the consortium because of cost overruns, time overruns, and technical complications.[2] The system was eventually delivered with high cost and time overruns, and the users were dissatisfied with it.[3] Based on our scoring model this was a failed engagement.

UWV–Logica Employee insurance provider UWV outsourced application development and maintenance for their subsidiary CWI to Logica (later CGI) in 2007. One of the developed systems was subject of a congressional investigation because of its failure. A reconstruction was published in 2014 in a Dutch newspaper.[4] Based on our scoring model this was a failed engagement.

VWS–SSO-ICT V&W The Ministry of Social Affairs and Employment (VWS in Dutch) outsourced their ICT (including desktops/laptops plus software) to a Shared-Services Organisation (SSO-ICT) of the Ministry of Transport, Public Works and Water Management (start around 2003). This SSO is now located at the Ministry of the Interior and Kingdom Relations and serves many more ministries. In 2014 about 22000 workplaces are supported with desktops, laptops and software by SSO-ICT.[5] Based on our scoring model this was a successful engagement.

[1] http://www.computable.nl/artikel/nieuws/outsourcing/4530118/1276946/delta-stopt-met-uitbesteding-ict-aan-atos.html.
[2] http://www.computable.nl/artikel/nieuws/erp/3198497/1276992/radboud-universiteit-stapt-uit-peoplesoftfiasco.html.
[3] http://www.cs.vu.nl/~x/knipselkrant/folia.pdf.
[4] http://www.nrcq.nl/2014/10/19/hoe-het-uwv-ingepakt-werd-door-ibm-en-logica.
[5] http://www.binnenlandsbestuur.nl/digitaal/nieuws/ssc-ict-beheert-nu-helft-ministerie-werkplekken.8504347.lynkx.

Table 8
Cronbach's alphas of the sets of multiple questions for the success determinants.

Success determinant	# Items in multi-item scale	# Respondents	Cronbach's alpha
D1	8	21	0.84
D2	7	23	0.75
D3	6	25	0.76
D4	7	16	0.84
D5	16	10	0.87
D6	13	24	0.70
D7	4	22	0.70
D8	15	23	0.66
D9	5	27	0.68

5.2. Discriminative power of the controllable success factors

In Section 3.1 we listed nine factors during the service delivery phase of the IT-outsourcing life-cycle which potentially determine whether an outsourcing deal is successful or not. In this section we show how we tested the power of these factors and whether or not they truly discriminate between successful and failed cases. In Section 5.2.1 we explain how we measured the success factors. In Section 5.2.2 we show how we statistically tested the discriminative power of the factors.

5.2.1. Measuring success determinants

To measure the factors we used multi-item Likert scales and summated ratings to quantify them. In their influential paper Gliem and Gliem state that it is better to use multiple questions for the same concept than to base the measurement on a single question [24]. In line with their research, each concept is measured with multiple questions. To give an idea of the type of questions, we provide below the complete set of the questions to measure the success determinant $D1$. As will be shown in Section 5.2.2 this controllable success determinant, together with $D6$ and $D9$, showed indeed strong power to discriminate between successful and failed cases. For the complete set of 516 different questions, the interested reader is referred elsewhere [49,55–57].

To measure $D1$ (Working according to the transition plan), the outsourcer was asked his opinion on the following statements using the Likert scale possibilities to answer (strong agreement, agreement, neutral, disagreement, strong disagreement, and non-applicable).

1. The transfer of the hardware was executed according to plan.
2. The transfer of the software was executed according to plan.
3. The transfer of staff was executed according to plan.

Likewise, the supplier was asked his opinion on the following statements using the same Likert scale possibilities to answer:

1. The plan for the transition of staff of the outsourcer has been followed.
2. The plan for the transfer of hardware of the outsourcer has been followed.
3. The plan for the transfer of the applications of the outsourcer has been followed.
4. The transfer of the applications of the outsourcer proceeds in clearly defined phases.
5. During the transition not only the software but also the related knowledge (e.g. the design) is transferred.

The respondents answered the questions in terms of their degree of agreement or disagreement with a statement regarding the extent in which the condition for success concerned has been met. We recoded the Likert type answers into 100, 75, 50, 25, 0 and NA. We omitted the NA answers and calculated the average of the scores of the remaining questions to quantify D1. Likewise, we obtained the scores of the other control factors.

We verified the internal consistency of each of the sets of multiple questions with Cronbach's alpha reliability coefficient of internal consistency [24]. In Table 8 the Cronbach's alphas are presented. The alphas are all but two acceptable to good (0.7–0.9), and the latter two are just below, but round to acceptable [23]. We conclude that the alphas are high enough to have sufficient trust in the questions in that they measure the same concepts.

5.2.2. Testing on discriminative power

As shown in Section 5.1 we classified the cases of our sample into two groups of successful cases and failed cases. Given this dichotomy it is possible to test the discriminative power of the success determinants. An appropriate test for that is the Mann–Whitney test. We illustrate for D1, how we performed this test.

There are 18 successful deals, and 12 failed. We can look at the D1 scores of this group as a random sample of 18 out of the D1 scores of the population of successful IT-outsourcing cases. We do not know the size of the population of successful cases, but we know that we have drawn a sample of 18 out of that population. From the population of failed cases we have drawn a sample of 12. If the distribution of the D1 scores substantially differs for both groups, then the null hypothesis: "the

Table 9
D1 scores of the successful cases and the failed cases.

Result	D1 scores	Average
Successful	69.44, 97.5, 71.88, 91.67, 79.95, 100, 87.5, 75, 75, 70, 93.75, 50, 77.08, 75, 82.29, 79.17, 70, 80.99	79.23
Failure	66.67, 85.16, 79.17, 75, 71.88, 75, 46.88, 59.38, 52.05, 71.43, 70.83, 67.26	68.39

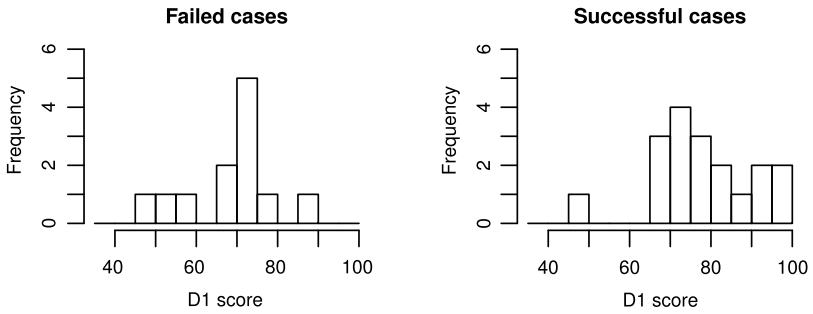

Fig. 5. Frequency histograms of D1 for successful cases and failures.

Table 10
Mann–Whitney p-values and meaning for all the factors.

Dn	Description	p-value two-sided	p-value one-sided	Relation
D1	Working according to the transition plan	0.0185	0.009248	strong
D2	Managing the business case for outsourcing	0.1493	0.07464	weak
D3	Managing the business case of the service supplier	0.5958	0.2979	absent
D4	Transfer of staff	0.1672	0.08359	weak
D5	Transfer of assets	0.1688	0.08441	weak
D6	Demand management	0.0028	0.001414	very strong
D7	Retention of expertise	0.5767	0.2883	absent
D8	Communication inside client organisation	0.7508	0.6406	absent
D9	Communication inside supplier organisation	0.0030	0.00151	very strong

two distributions do not differ substantially", will be rejected at a low percent level of significance. Usually one demands 5% level of significance.

We statistically tested our null hypothesis: "the two distributions of D1 do not differ substantially for the group of successful cases and the group of failures" by applying the Mann–Whitney test. This test is a non-parametric test for testing equality of population means between groups. The test makes minimal assumptions about the underlying probability distributions (e.g. a normal population is not assumed) and is a test for independent samples. The test requires at least an ordinal scale of measurement of the data and this requirement is met in our case.

In Table 9 we provide the scores for $D1$ for the cases presented in Table 6. On top of that we provide both average scores for the partition of successful and failed deals. It shows that the mean score of D1 is different for the two groups. The average score of the group of successful cases is clearly higher than the average score of the group of failed cases. See also Fig. 5 for the frequency histograms of D1 of the group of successful cases and the group of failed cases.

Fig. 5 strongly suggests that the D1 distributions of the 2 groups substantially differ, but visual inspection is no formal proof. The Mann–Whitney test provides a more formal answer. We tested the null hypothesis that the location parameters of the distributions of D1 are the same in each group and found the low p-value of 0.0185. This p-value is certainly low enough to reject the null hypothesis if one demands the usual 5% level of significance. In fact, our grading system is designed in such a manner that a higher score ideally correlates with higher odds on a successful outsourcing deal. So we also test whether the successful scores are higher than the failed ones. The Mann–Whitney test gives a p-value of 0.009248 for this. From this we conclude that there is a strong statistical indication that D1 has power to discriminative between successful and failed outsourcing deals and that higher scores lead to more success than lower scores.

As an example we calculate the power of the one-sided Mann–Whitney test for $D1$ by means of the power of the t-test. The power of the Mann–Whitney test is related to that of the t-test in the sense that at worst it is 84.6% of the power of the t-test [28]. The effect size is 0.86, which is in line with the 0.8 suggested by Cohen for a large effect size when performing a t-test [10]. The achieved power of the test is 73%, which is acceptable but somewhat lower than the 80% guideline. We carried out Shapiro–Wilk normality tests to check whether the normality assumptions of the t-test are violated by our data. These tests did not give low p-values so presumably the normality conditions of the t-test are not too much violated. Therefore, we could also have carried out a t-test (which gives similar p-values) hence applying the worst-case bound of 84.6% is not needed.

We obtained the Mann–Whitney p-values for the other control factors analogously. In Table 10 we present our findings. From these tests it follows that there is strong statistical evidence that, besides $D1$, the decision factors $D6$ and $D9$ also have significant power to discriminate between successful and failed cases. Or, in other words, besides $D1$ also $D6$ and $D9$ are useful predictors of the chance of success and failure in an early stage of the service delivery phase of the IT-outsourcing life-cycle. Also in the cases of $D6$ and $D9$, the one-sided Mann–Whitney test gave very low p-values, indicating that higher scores must be associated with success and lower with failure (p-values are 0.001414 and 0.00151 respectively). The effect sizes are 1.17 and 1.09 respectively, and the powers are 92.2% for $D6$ and 88.4% for D9 (normality tests indicated that a correction of 84.6% of the t-test power was not needed). A number of other factors have a weak correlation with success and failure: the two-sided test was above the 0.1 significance level, but the one-sided version below it, but still above 0.05. A few factors did not show any correlation, also not in the one-sided test. So the relation is absent, or in case of a type II error, its effect is presumably weak.

The Mann–Whitney test is useful in finding what control factors have discriminative power to distinguish between successful and failed cases and gives us the direction: higher scores lead to higher success chances. This test, however, does not tell us how much the chance on success will increase if one succeeds in improving scores. In the next section we use logistic regression to answer this question.

5.2.3. Logistic regression

In this section we show how we applied binary logistic regression as modelling technique to predict the chance of success given the scores of the critical success factors (risk drivers). A binary variable can only take two values: 0 and 1. Binary logistic regression is a specialized form of regression that is designed to predict and explain a binary (categorical) variable (coded as 0 or 1) rather than a continuous variable.

We have used the logistic regression procedure of the statistical R package[6] to carry out our computations. The logistic regression procedure of R does not only yield the estimated regression coefficients of the regression equation, but also the standard errors for the estimated parameters. If the sample size is sufficiently large, each regression coefficient is normally distributed by approximation.

The simplest estimator of the probability of success is obtained by counting how many projects in the sample displayed success and divide it by the total number of projects. For our 30 projects we obtain $p = 18/30 = 60\%$ chance of success. This is a very simple constant model for failure risk: given a new project, the chance on failure is 40%. We call this model the null model.

We have much more information at our disposal and it is likely that we can do a better job in predicting the chance on success than the null model. Therefore, we take into account all variables in our dataset $D1, D2, \ldots, D9$ that potentially influence the chance on success of the outsourcing deal. If we allow all nine potential risk drivers to enter the logistic regression model and search for the optimal solution then we find the following logistic regression equation:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \underset{(13.953)}{-36.7664} + \underset{(0.050)}{0.06806 D1_i} + \underset{(0.101)}{0.22886 D6_i} + \underset{(0.102)}{0.22926 D9_i}$$

Here p_i is the probability that case i will be successful. The standard errors of the estimated regression coefficients are given in brackets. If one tests the difference of the estimated model from the null model (the start model in which no explanatory variables are included but just the constant), one finds that the estimated model is a significant improvement at the $\alpha = 0.01$ level when comparing to the null model.

In Fig. 6 we provide a still of an animation of the logistic formula where for varying D1, D6, and D9 the chance of success is presented. On the horizontal axis the demand management score (D6) varies from 0–100, on the depth axis the score on internal communication within the supplier ranges from 0–100 (factor D9). By playing the animation the score for the transition plan (D1) increases from 0–100, the still is at an 80% score for D1. On the vertical axis the chance of success for the deal is presented, which is the outcome of the logistic regression formula.

Now that we provided some intuition about possible outcomes of the formula, we discuss it further. The variables $D2$, $D3$, $D4$, $D5$, $D7$ and $D8$ are not statistically significant based on the likelihood-ratio test. This is what would be expected given the results of the Mann–Whitney tests that show very low p-values found for $D1$, $D6$, and $D9$ and large p-values for the other control factors (see Table 9 in Section 5.2.2). The Mann–Whitney test already told us that $D1$, $D6$ and $D9$ have strong discriminative power. Surprisingly, the standard error of the regression coefficient of $D1$ is relatively large. From the Mann–Whitney test one would expect $D1$ to have a strong discriminative power (p-value of 0.0185) but the estimated regression coefficient of $D1$ does not appear significantly different from 0, using a significance level of 5%.[7]

The coefficients of the determinants can be interpreted meaningful. They tell us that better management increases the chance on a successful progress of the cooperation between outsourcer and service provider. A logistic coefficient can be interpreted as the change in the log odds associated with a one-unit change in the explanatory variable. So, one can estimate the increase of the chance of success due to an increase of the score of $D1$, $D6$, or $D9$. The most important risk driver is

[6] www.r-project.org.

[7] To be sure we calculated the exact p-value of the Mann–Whitney test, just in case the approximate p-value was erroneous. But also in the exact case the p-value was low: 0.01623.

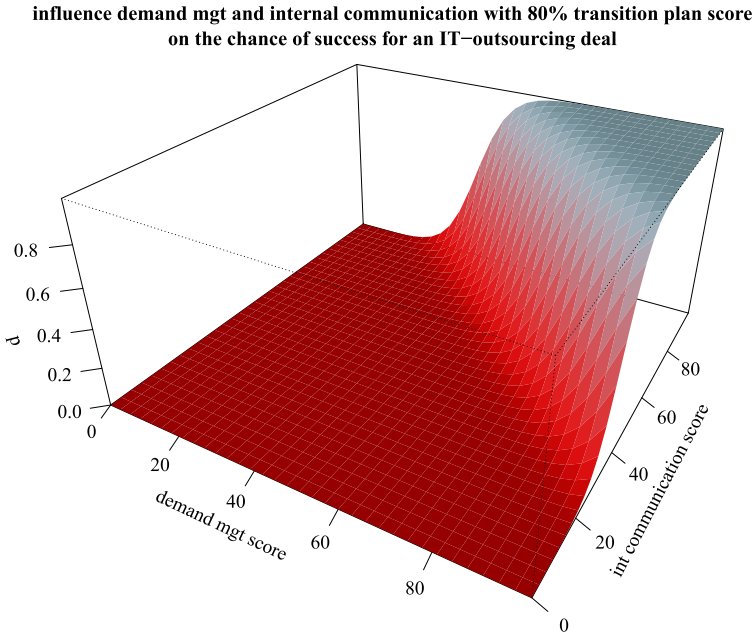

Fig. 6. Still of the animation of the logistic formula expressing the chance of success by varying D1, D6, and D9. The animation is available as supplementary electronic appendix on http://dx.doi.org/10.1016/j.scico.2016.04.001.

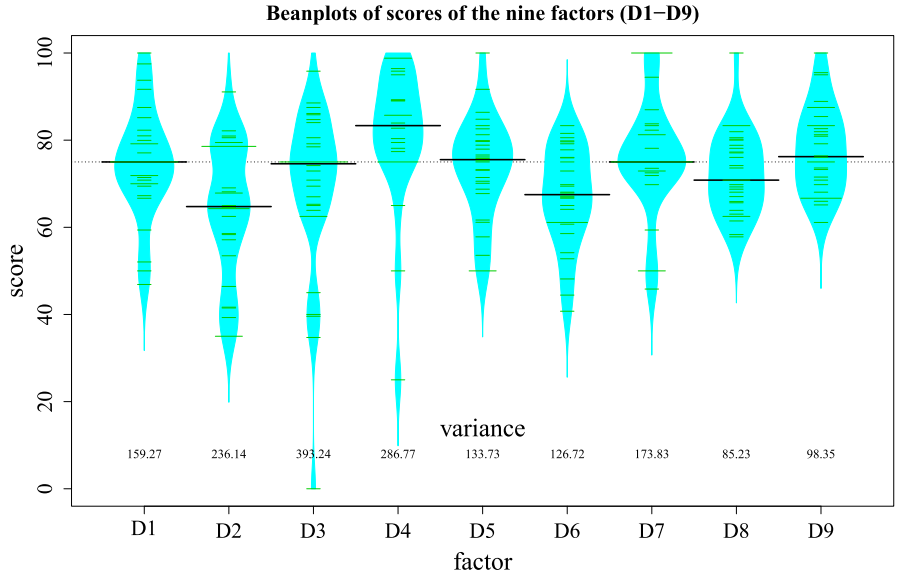

Fig. 7. Beanplots showing the range of values of the nine factors, plus their variance.

communication inside the supplier organisation, which, if organised well, significantly increases the chance on success. This finding is rather surprising as we were not aware of any suggestion in this direction in the literature. Next, there is demand management, which, if managed well, increases the chance on success of the outsourcing deal to a large extent. Last in ranking is strictly working according to the transition plan, which increases the chance on a productive cooperation.

Sensitivity analysis To provide more intuition for the found logistic regression equation, let us calculate the probability of success for different values of D1, D6 and D9 in the range of 0 to 100. In doing so we make the simplification that varying the other 6 controllable factors do not have any effect on the chance on success of the case. In general it might be that the variance in some scores is too small so that the variance in the outcomes of the cases, success or failure, could not be explained from the variances of the scores of these factors.

In Fig. 7, we depict the result of our analysis of the variance of the nine factors. For each factor we gave their values plus the chance of occurrence as a contour via a so-called beanplot [33]. The dotted horizontal line is the overall median,

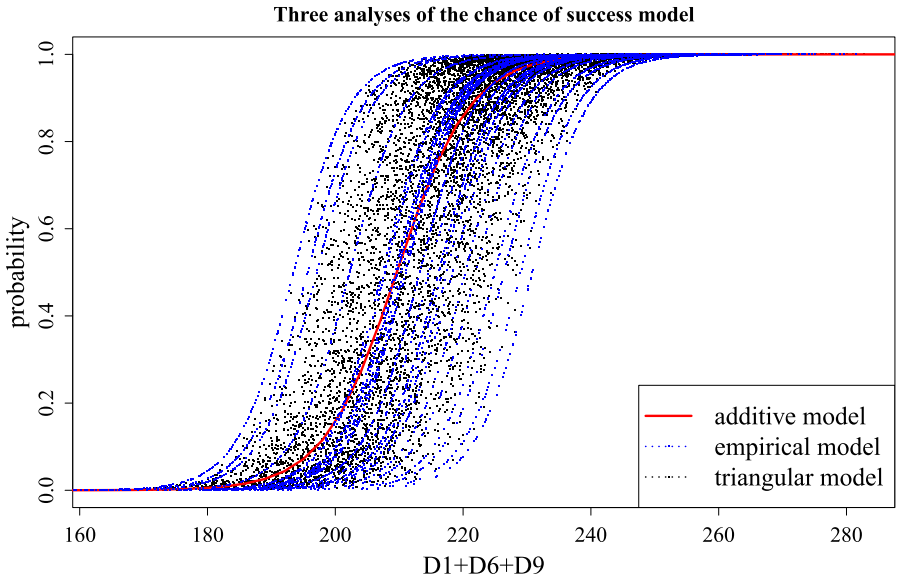

Fig. 8. Three analyses of the outcomes of the logistic regression equation.

and the thick solid ones are the individual medians. The stacked short lines are the observed values, and the contours give their chance of occurrence. Visually it is clear that the insignificant factors have similar spread as the significant ones. We also calculated the variances of all the factors and the ones that did not enter the logistic regression equation do not differ notably in order of magnitude from the variances of the scores of D1, D6 and D9 that entered the regression equation indeed. From this we conclude that there are strong indications that the insignificant control variables did not enter the regression equation because the fluctuations in their scores did simply not correlate with the fluctuations in the outcomes of the cases. Of course, the scores of insignificant factors may have some effect on the real chance of success but that effect must be small in comparison to the effect of changing the scores of D1, D6, and D9. We therefore think it makes sense to focus on our logistic regression model using only D1, D6, and D9 in exploring the effect on the chance on success of increasing or decreasing the scores of the controllable factors.

The result of our sensitivity analysis is shown in Fig. 8. On the horizontal axis we put the sum of the three factors D1, D6 and D9; so 0, 3, 6, 9, …, 300. Vertically we put the chance of success calculated with our logistic regression equation. We carried out three analyses to obtain an idea of the success chances that our logistic formula displays. The first analysis is just an *additive model*, where D1, D6, and D9 simultaneously obtained the same value between 0–100. For each of these values the probability of success was calculated leading to the solid S-curve in the figure (red in the electronic version of this article). But of course, not all factors need to be the same. Therefore we also took the actual observations for D1, D6 and D9 each 10000 times and for all these combinations we calculated the chance of success and dotted them in the same figure (blue dots in the electronic version of this article). This, we call the *empirical model* because we took the observed values. Of course, sampling 10000 times 3×30 observed values leads to systematic patterns. The empirical model shows multiple similar S-curves. Finally, we took the minimum, maximum and median values of the three factors from which we constructed three triangular distributions, this we call the *triangular model*. A random sample using 10000 values per factor gives the black dotted pattern in the same region as the empirical model but without the systematic S-curve pattern.

The analyses show that scores for the three factors need to be high in order to obtain a chance of success at all. For instance, the additive model shows a chance of success of 51% if all factors score at 70 points (so the sum is 210). At 225 points (so all factors have 5 more points) the model predicts a chance of success of 94%. From the other models we learn that a high chance of success might also be obtained by a lower score and also that a sum of 225 of the three factors does not guarantee a chance of success of 94%. The chance of success does not depend on the sum of the scores of the three factors but on how the sum has been made up. Of course the probability of a high chance of success is much higher if the sum of the 3 factors is 225 or higher than in case of a lower total, say 210.

The formula shows that communication inside the supplier matters the most, shortly followed by demand management. Sticking to the transition plan matters the least. Of course, good scores on internal communication and demand management are not nearly as useful if the transition plan is absent or not at all followed. It is ill-advised to use our logistic formula to obtain the highest chance with the lowest possible scores. We do advise, however to make an effort of maximizing on each of the factors D1, D6 and D9 to improve the chance of success for outsourcing deals.

Goodness-of-fit of the model In the case of a linear regression model the correlation coefficient, R^2 is a number between zero and one, which provides a quality measure for how good a linear regression model fits. Also for logistic regression such so-called goodness-of-fit measures have been developed. We used McFadden's Pseudo R^2 as goodness-of-fit metric.

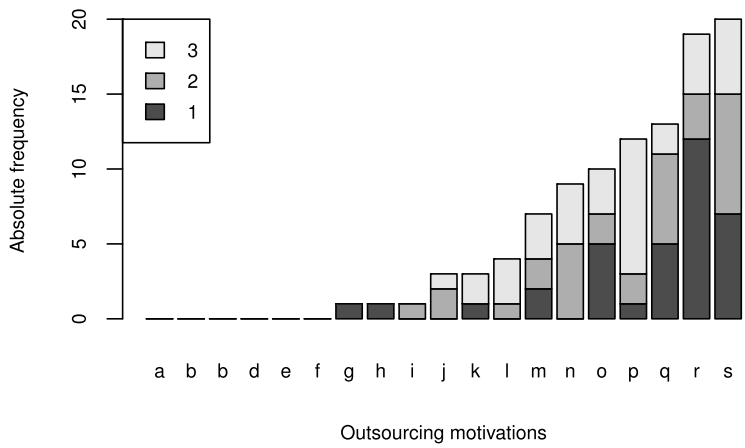

Fig. 9. Frequency histogram of the motives at rank 1, 2 and 3 of the outsourcer. Legend: a. pressure by shareholders, b. injection of capital, c. shared services centre, d. follow-up sourcing, e. pressure by regulatory supervisors, f. extended working hours, g. career of IT staff, h. pressure from higher management, i. mandatory public tender, j. short-term reduction of costs, k. guarantee business continuity, l. continuity in case of illness and turnover of staff, m. cover IT risks, n. predictability of costs, o. access to specialized staff, p. more flexibility, q. access to knowledge and experience of supplier, r. focus on core activities, s. long-term reduction of costs.

McFadden's R^2 statistic is one of the simplest and most popular R^2 measures for logistic regression. It depicts the amount of reduced deviance by the estimated model as a percentage of the deviance of the null model. In our case McFadden's Pseudo R^2 amounted to 0.594. Considering that a pseudo R^2 is to be expected much less than what would be expected in an ordinary least square model, a pseudo R^2 of approximately 0.6 indicates a reasonably good model.

5.3. Rigid success determinants of IT-outsourcing success

We have also tested the discriminative power of the rigid success determinants.

1. the motive of the outsourcer to engage in an IT-outsourcing deal,
2. the motive of the service provider to engage in an IT-outsourcing deal,
3. the match of the organisation cultures of the service provider and his client,
4. the type of outsourced work,
5. the capability of the service provider to put himself in the position of his client (vendor's empathy capability),
6. hiring consultancy support.

These case characteristics differ from the success determinants discussed in Section 3.1 as they cannot be adjusted or improved by the provider or the vendor during the service delivery phase. We therefore did not include them among the potential explanatory variables in carrying out the logistic regression analysis.

5.3.1. Motivation of the outsourcer

In Fig. 9 we show a multiple bar chart of the different motives mentioned by the outsourcers to outsource (some of) their IT functions. The partition of the diagrams shows how many times a motive was ranked on the first place, the second place and the third place. We did not depict fourth, fifth, ..., ranking as this turned out to have no added value.

Most reasons relate to cost savings. The next popular reason is strategic: focus on core capabilities, increasing flexibility, safeguarding continuity, etc. Access to high quality employees comes at the third place. We clustered the 19 motives mentioned by the respondents at the first rank into three main groups:

1. strategic motives;
2. financial motives;
3. HRM (Human Resource Management) motives.

An interesting question is whether there is a relationship between the outsourcing motive and the success or failure of the outsourcing deal. We applied the test of independence for contingency tables (chi-square test) to obtain the answer. In Table 11 the observations are classified by two characteristics: the motive of the outsourcer to engage in an outsourcing deal and the success of the deal.

The figures in brackets refer to the frequencies one would expect if both classification factors are completely independent from each other. The other figures refer to the observed frequencies. If the observed frequencies do not differ too much from the figures in brackets, then there is a strong indication that the two classification factors are mutually independent. We

Table 11
Relating outsourcer's main motives and success classes.

RESULT \ MOTIVE	Strategic	Financial	HRM	Total
Successful	9 (9.6)	5 (4.2)	4 (4.2)	18
Failed	7 (6.4)	2 (2.8)	3 (2.8)	12
Total	16	7	7	30

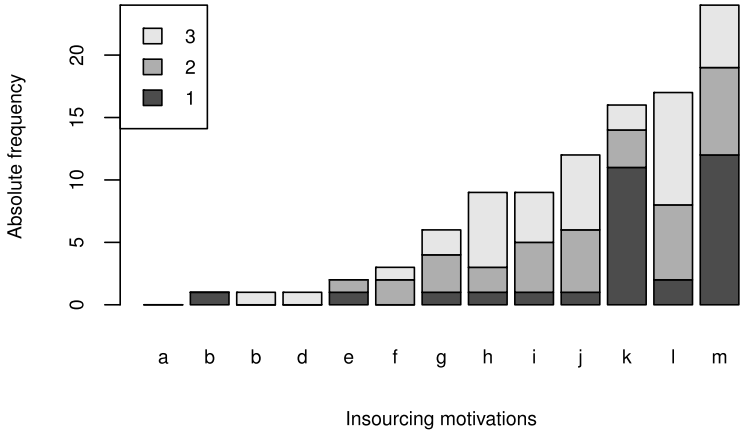

Fig. 10. Frequency histogram of the motives at rank 1, 2 and 3 of the supplier. Legend: a. professionalizing of services, b. reinforcement of market position, c. growth, d. reciprocal business opportunity, e. external pressure, centralization of public services, f. acquisition of specialized staff, g. attractive profit margin, h. knowledge acquisition, i. short-term turnover, j. synergy with existing activities, k. access to the market, l. visibility as supplier, m. long-term turnover.

Table 12
Relating suppliers' main motives and success classes.

RESULT \ MOTIVE	Long-term ambitions	Short-term ambitions	Total
Successful	12 (8.4)	6 (9.6)	18
Failed	2 (5.6)	10 (6.4)	12
Total	14	16	30

tested the null hypothesis that the two classification factors are independent (chi-square test) and found the high p-value of 0.887. From this we conclude at the 5% level of significance that the sample does not contradict the null hypothesis. In other words there is strong statistical evidence that there is no relationship between the motive of the outsourcer to engage in an outsourcing deal and the success of the deal.

5.3.2. Motivation of the service provider

In Fig. 10 we show a multiple bar chart of the first, second and third ranked motives of the service suppliers to insource IT functions.

Among the first ranked motives "become well known", "access tot the market", and "long-term turnover" come at the first place. At the second place comes increasing volume of business in the short term. The next popular reason is building up (specific) knowledge. Also increasing the profit margin and synergy benefits are frequently mentioned motives. An interesting question is whether there is a relationship between the motive of the service supplier and the success or failure of the outsourcing deal. We clustered the motives into two groups:

1. reaching long-term (strategic) ambitions
2. reaching short-term ambitions

Examples of long-term (strategic) motives are: "become well known", "access to the market", and "long-term turnover". Examples of short-term addressed motives are: "increasing volume of business", "attractive profit margin", and "short-term turnover".

Subsequently, we applied the test of independence for contingency tables (chi-square test) to test the null hypothesis that there is no relationship between the motive of the service provider and the success or failure of the outsourcing deal. The contingency table is shown in Table 12.

Table 13
Relating matching of organisation cultures and success classes.

RESULT \ MATCH	No difference	Strong difference	Weak difference	Total
Successful	9 (7.8)	3 (4.2)	6 (6)	18
Failed	4 (5.2)	4 (2.8)	4 (4)	12
Total	13	7	10	30

Table 14
Relating type of outsourced work and result case.

RESULT \ OUTSOURCED WORK	Software	Techn. Infra.	End User Sup./Off.Aut.	Total
Successful	7 (7.8)	6 (6)	5 (4.2)	18
Failed	6 (5.2)	4 (4)	2 (2.8)	12
Total	13	10	7	30

We found the very low p-value of 0.001, which strongly indicates that the two classification criteria are not independent of each other. In other words, the nature of the motive of the service supplier increases or decreases the chance on success of the deal.

As an example we will calculate the power of the test in Table 12. Cohen suggested about 30 observations for it. The measured effect size is 0.49, which is completely in line with Cohen's suggestion of 0.5 for large effect sizes. Its power is 77%, so indeed in the order of about 80%. So the chance of making a type II error is acceptably low.

5.3.3. Matching of organisation cultures

An interesting question is whether there is a relationship between the match of organisation culture and the success or failure of the outsourcing deal. We first clustered the cases in three groups:

1. No difference in organisation culture: informal ↔ informal, between ↔ between, or formal ↔ formal.
2. Strong difference in organisation culture: informal ↔ formal
3. Weak difference in organisation culture: informal ↔ between or formal ↔ between

Table 13 shows the result of the clustering. We tested the null hypothesis that the two classification factors are independent (chi-square test) and found the p-value of 0.503. From this we conclude at the 5% level of significance that the sample does not contradict the null hypothesis. In other words there is statistical evidence for our sample of 30 cases that there is no relationship between the matching of the organisation cultures of outsourcer and vendor and the success of the deal.

As an example we calculate the effect size of the test in Table 13, which is 0.21. This is a small to medium effect size according to Cohen [10]. From Fig. 1 we know that the power will be low, given 30 observations. Indeed it is 16%, so the chance of a type II error is present. If you would want to know whether this insignificance is representative for the entire population of 700 cases at a power of 80%, we need 220 cases (7 times more cases), which is prohibitively expensive. But even if this factor does contribute to success or failure, its influence is presumably relatively modest, given the measured effect size and the representativeness of our 30 engagements. We are interested in major factors that make a strong difference between success and failure.

5.3.4. Does the type of outsourced work matter?

In questionnaire 1 (see Section 4), the outsourcer is asked to give details about the type of outsourced work. See Fig. 4 in Section 4.3 for the 12 different types of outsourcing work mentioned by the business partners of the 30 IT-outsourcing cases.

We clustered the specified types of outsourced work in three groups:

1. Software (13)
2. Technical Infrastructure (10)
3. End User Support/Office Automation (7)

In Table 14 the observations are classified by two criteria: the type of outsourced work and the result of the deal.

We tested the null hypothesis that the two classification factors are independent and found the high p-value of 0.887. From this we conclude at the 5% level of significance that the sample does not contradict the null hypothesis. So there is strong statistical evidence that there is no relationship between the type of outsourced work and the result (success or failure) of the deal. It seems logical that the type of outsourced work influences the chance on success of an outsourcing deal, but our finding shows that there is no such influence.

Table 15
Assessment capability of the two different success groups.

RESULT	AC scores	Mean
Successful	12.5, 12.5, 12.5, 12.5, 25.0, 16.7, 8.3, 16.7, 12.5, 25.0, 16.7, 16.7, 16.7, 41.7, 0.0, 0.0, 0.0	14.0
Failed	8.3, 8.3, 33.3, 16.7, 16.7, 25.0, 12.5, 16.7, 75.0, 25.0, 16.7	23.1

5.3.5. Is vendor's empathy a success factor?

In questionnaire 3 (see Section 4), the vendor is asked to assess the opinion of his client about a number of matters. For example, first the outsourcer is asked whether his top management appreciates the quality of the delivered services. Subsequently, the vendor is asked what he thinks his client has answered on this question. By comparing the answers of the outsourcer and its vendor an idea is formed about the capability of the vendor to put himself in the position of his client.

Recall that all answers to the questions are of the Likert scale type and the possible responses are recoded into the numbers 100, 75, 50, 25 and 0. If the vendor judges his client perfectly the difference between his assessment of the answer of his client and the actual answer of his client is 0. In case of underestimation the maximal difference is 100. The same holds for overestimation.

To measure the capability of the vendor to put himself in the position of its client we compute the average of the underestimations, overestimations and correct assessments. If each of the 3 questions has been answered by both the vendor and outsourcer we compute the average of three numbers. In case of missing data (vendor or outsourcer or both have not answered some question(s)), the concerned question(s) is (are) left out of consideration in calculating the norm. In that case the computed assessment capability is the outcome of the division of two numbers or just consists of one number. In a formula:

$$AC = \frac{\Sigma_{j=1}^{n} |x_j - y_j|}{n} \tag{1}$$

Here AC stands for the vendor's assessment capability, x_j for the vendor's score on question j, y_j for the score of the outsourcer on question j, and $||$ for the absolute difference between the answers, and n stands for the number of questions that have been answered by both the outsourcer and service supplier.

The assessment capability cannot exceed the number 100 and its minimal value is 0. We have made the following classification. If the assessment capability is in the range of 0–25 the vendor shows that he is very well capable to put himself in the position of the client. An assessment capability in the range of 25–50 indicates a moderate empathy of the vendor. An assessment capability in the range of 50–100 indicates that the vendor is unable to put himself in the position of his client.

We statistically tested the null hypothesis that the assessment capability scores of the cases that come into the failure class are statistically significant higher than the assessment capability scores of the cases that come into the class of successful cases. To that we applied the Mann–Whitney test on the data recorded as in Table 15. We found a p-value of 0.08. This p-value is not low enough to reject the null hypothesis that there is no relationship at the 5% level of significance. However, at the 10% level the null hypothesis would not have been rejected. The effect size of this test is 0.59, which is no longer a large effect in terms of Cohen's suggestion for a t-test: 0.5 is medium, and 0.8 is large. The underlying data is not normally distributed as Shapiro–Wilk tests show. Therefore, we use the worst-case bound of 84.6% for the approximation of the power of the Mann–Whitney test by the t-test, which leads to a power between 50–59% (at the 10% level of significance). To reach a power of 80% at the 5% level for an effect size of 0.59, each group should at least contain 36 cases, which is prohibitively expensive, apart from the difficulties in obtaining the commitment of so many organisations in the first place.

5.3.6. The effect of hiring sourcing consultants

In 15 cases the outsourcer hired a consultant. In 11 cases a consultant was hired to compensate for lack of experience of the outsourcer. In 4 cases a consultant was hired although the outsourcer was experienced. The information regarding the experience status of the outsourcing companies that participated in the longitudinal study was collected through the various questionnaires.

One would expect that the chance on success of an outsourcing deal increases when the outsourcer has built up experience with IT-outsourcing in the past or hires experience from consultancy offices to compensate the lack of in-house experience. In this section we analyse the joint effect of experience with outsourcing and hiring consultancy support. The question to be answered is whether experience built up in the past is helpful to gain success in the future and/or whether consultancy support is helpful when in-house experience is lacking. To that end we started with classifying our sample of 30 cases according to 3 classification criteria:

1. Criterion A: whether or not the case was successful;
2. Criterion B: whether or not the outsourcer had built up experience with outsourcing;
3. Criterion C: whether or not a sourcing consultant was hired.

Table 16
Three-way contingency table to test whether the three criteria A, B, and C are mutually independent.

		Result case:				
		Success		Failure		
		Maturity:				
		Experience	No experience	Experience	No experience	
Consultant	yes	2 (3.9)	4 (5.1)	2 (2.6)	7 (3.4)	15
	no	8 (3.9)	4 (5.1)	1 (2.6)	2 (3.4)	15
		10	8	3	9	30

Table 17
Two-way contingency table to test whether the A classification is independent of B and C.

	Experienced/ no consultant	Experienced/ consultant	Inexperienced/ no consultant	Inexperienced/ consultant	
Success	8 (5.4)	2 (2.4)	4 (3.4)	4 (6.6)	18
Failure	1 (3.6)	2 (1.6)	2 (2.4)	7 (4.4)	12
	9	4	6	11	30

To test whether these 3 classification criteria are mutually independent we constructed a $2 \times 2 \times 2$ contingency table; see Table 16.

The numbers in brackets show the theoretical frequencies if there is no relationship between the 3 classification criteria. Based on our sample of 30 cases the best estimate of the probability of a case to be successful is 18/30; the best estimate of the probability that a consultant is hired amounts to 15/30, and the best estimate of the probability that the outsourcer has gained experience with outsourcing in the past is 13/30. If the three classification criteria are mutually independent one would expect $(18/30 \cdot 15/30 \cdot 13/30) \cdot 30 = 3.9$ experienced cases with a consultant to be successful. The other numbers in brackets are computed in the same fashion. We now test the null hypothesis that the A, B and C classifications are independent. Or, in a formula, we test the null hypothesis: $p_{i,j,k} = p_i \cdot p_j \cdot p_k, \Sigma p_i = 1, \Sigma p_j = 1, \Sigma p_k = 1$. Here $p_{i,j,k}$ is the joint chance of which we assume it is the product of the individual chances p_i, p_j, p_k; that of course sum up to 1.

The chi-square statistic is computed from the eight categories as follows:

$$\chi^2 = \frac{(3.9-2)^2}{3.9} + \frac{(5.1-4)^2}{5.1} + \frac{(2.6-2)^2}{2.6} + \frac{(3.4-7)^2}{3.4} + \frac{(3.9-8)^2}{3.9} + \frac{(5.1-4)^2}{5.1} + \frac{(2.6-1)^2}{2.6} + \frac{(3.4-2)^2}{3.4} = 9.847$$

Note that as a rule of thumb the χ^2 approximation is adequate if the minimum theoretical frequency is 2 [18]. The statistic has a chi-square distribution with 4 degrees of freedom. Let $\chi_{.95}$ denote the 95th percentile of the χ^2 distribution. For 4 degrees of freedom $\chi_{.95} = 9.49$. Since the observed value of 9.847 is larger than 9.49 we reject the hypothesis of mutual independency of the three classification criteria at the 5 percent level of significance.

As an example we calculate effect size and power for the test in Table 16. As it is a three-way table, we use Cramér's effect size formula: $\sqrt{(\chi^2/N(k-1))}$, where k is the minimum of the table dimensions (which is 2 since it is a $2 \times 2 \times 2$-table). This leads to an effect size of 0.57, so a bit higher than Cohen's suggestion of 0.5. The power of this test turns out to be 71%, which is in the order of 80% but a bit lower.

Clearly, there is statistical evidence that the three classification criteria are not mutually independent. However, without further examination it is not clear to which classification criteria(s) the dependency has to be attributed. When we look at Table 16 it strikes that the observed frequency of 8 successful cases without consultant is much larger than the frequency one would expect if there is no relationship between success, experience of the outsourcer and hiring a consultant, which would be 3.9. The same holds for the observed frequency of the inexperienced outsourcer that hired a consultant. One would expect 3.4 failures, but the observed number of failures is twice as high, namely 7. So there is a good reason to look into more detail at the impact of hiring a sourcing consultant on the result of a case. For this purpose we carried out 4 additional contingency table tests.

In connection with Table 16 we can also test whether the A classification is independent of B and C. For that we constructed a two-way contingency table; see Table 17.

The figures in brackets again refer to the frequencies one would expect if both classification factors are completely independent from each other. For example, one would expect $(18/30) \cdot 9 = 5.4$ experienced outsourcers without consultant to be successful if the two classification criteria are independent. If we compute the χ^2-statistic from the eight categories

Table 18
Two-way contingency table to test whether hiring a consultant and the outcome of sourcing are independent.

	Consultant	No consultant	
Success	6 (9)	12 (9)	18
Failure	9 (6)	3 (6)	12
	15	15	30

Table 19
Two-way contingency table to test whether experience of the outsourcer with outsourcing and the outcome of the case are independent.

	Experience with outsourcing	No experience with outsourcing	
Success	10 (7.8)	8 (10.2)	18
Failure	3 (5.2)	9 (6.8)	12
	13	17	30

Table 20
Two-way contingency table to test whether hiring a consultant and the experience of the outsourcer with outsourcing are independent.

	Experience with outsourcing	No experience with outsourcing	
Consultant hired	4 (6.5)	11 (8.5)	15
No consultant hired	9 (6.5)	6 (8.5)	15
	13	17	30

we obtain the number 5.968. For 3 degrees of freedom $\chi_{.95} = 7.81$. Since the observed value of 5.986 is less than 7.81 we accept the null hypothesis of independency of the two classification criteria at the 5 percent level of significance. So the outcome of the deal is independent of B and C.

Subsequently we tested the dependency between the A classification and the B classification; the dependency between the A classification and the C classification, and finally, for the sake of completeness, the dependency between the B classification and the C classification. To that end we constructed the three two-way contingency Tables 18, 19 and 20.

From Table 18 it follows that the χ^2 statistic equals 5 with a p-value of 0.06. We provide a last example that our sample size is acceptable. We calculate the actual effect size, which is 0.41, a bit lower than Cohen's suggested 0.5 for large(r) effects of which you do not know the effect size yet. Given this smaller effect size, the actually achieved power of our test turns out to be 61% at the 5% significance level. Since the significance level turned out to be 6%, the achieved power is 64%. In a ceteris paribus situation with 15 to 17 cases more the power would be at 80%.

So there is a strong statistical indication that hiring sourcing consultancy support does not increase the chance on a successful outcome of the outsourcing deal. Even stronger, the opposite is at hand: the chance on a failure increases when a sourcing consultant is hired.

From Table 19 we infer that $\chi^2 = 2.738$, with a p-value of 0.143. So there is no convincing statistical evidence that past experience with outsourcing increases the chance on a successful outcome of the deal. This is surprising since you would assume that being experienced in outsourcing helps in future success.

Finally, from Table 20 we calculate $\chi^2 = 3.394$ with a p-value of 0.137. Hence, there is a very slight statistical indication that inexperienced outsourcers more frequently hire sourcing consultants than experienced outsourcers. This finding is not surprising.

From the additional contingency table tests, we conclude that the major cause of the mutual dependency of the three classification criteria A, B, and C is the absence or presence of a sourcing consultant. If we separate the effect of hiring a sourcing consultant from the experience of the outsourcer and test whether there is a relationship with the outcome of the case then there is a strong statistical indication that hiring consultancy support works contra-productive, whereas the relationship between having experience or not and success or failure is only weakly present.

6. Discussion

In this study we investigated the discriminative power of nine controllable success determinants during the service delivery phase of the IT-outsourcing lifecycle and six rigid success factors. The results of our tests have been presented in Section 5. In this section we will elaborate on the interpretation and meaning of our findings. First we briefly summarise them in Table 21 and give their effect on success or failure if any. Below, we discuss them in more detail.

Table 21
Overview of the predictive power of all factors tested in our study.

Type	Short name	Description	Effect on success
Controllable	D1	Working according to the transition plan	positive
Controllable	D2	Managing the business case for outsourcing	not found
Controllable	D3	Managing the business case of the service supplier	not found
Controllable	D4	Transfer of staff	not found
Controllable	D5	Transfer of assets	not found
Controllable	D6	Demand management	very positive
Controllable	D7	Retention of expertise	not found
Controllable	D8	Communication inside client organisation	not found
Controllable	D9	Communication inside supplier organisation	very positive
Rigid	–	The motive of the outsourcer to engage in an IT-outsourcing deal	not found
Rigid	–	Long-term motives of the service provider to engage in an IT-outsourcing deal	positive
Rigid	–	Short-term motives of the service provider to engage in an IT-outsourcing deal	negative
Rigid	–	The match of the organisation cultures of the service provider and his client	not found
Rigid	–	The type of outsourced work	not found
Rigid	–	The capability of the service provider to put himself in the position of his client (vendor's empathy capability)	positive
Rigid	–	Hiring consultancy support	negative

Legend: positive = positive influence of the factor on the success; negative = negative influence of the factor on the success; not found = weak or no noticible effect found in our study.

6.1. Controllable success factors

Three controllable factors significant We found strong statistical evidence for three controllable determinants that they are good predictors of success or failure. First, there is the communication inside the vendor's organisation, next there is demand management, and then comes strictly working according to the transition plan. Particularly, the finding that the quality of the communication within the supplier organisation is an important controllable success factor is remarkable. To the best of our knowledge, there is almost no reported research that addresses this issue. We found only one publication by Urbach in 2012 in which the author indicates the importance of good communication inside the outsourcer's and vendor's organisation for a successful IT-outsourcing project [53]. Our research validates Urbach's suggestion of good communication inside the organisation of the vendor; our research did not show that the quality of the communication inside the organisation of the outsourcer is critical for success, however.

Six controllable factors not significant The other six controllable success factors did neither show discriminative nor predictive power. From this it does not follow, however, that these success determinants do not matter. It certainly makes sense to keep track on them for other reasons.

Predictive model Next to discriminating between failure and success we were also interested in the odds for such outcomes. We used logistic modelling to answer that question. The logistic regression showed that the three already found significant controllable factors turned out to be significant as well in estimating how much the chance of success will increase if one succeeds in improving the scores on these factors. It turned out that improving demand management and internal communication within the supplier organisation increased the odds the most. So a key-lesson is that excellent demand management and seamless internal communication at the supplier are both important factors to assess an IT-service provider, and vice versa that service providers will add value by superior management of the customer's demand and effective internal communication. We would have expected that sticking to the transition plan was key. For instance, since requirements volatility can seriously jeopardize a successful outcome [35]. However, although sticking to the transition plan also turned out to increase the odds in our study, it did not have that much an effect. Other controllable factors did not show predictive power in the logistic regression equation.

Retention of expertise One would have expected that retention of expertise within the client organisation would have been key in increasing the odds. See for example the article of Willcocks et al. [58]. Providing more control by internal experts who know the drill of the IT-systems is likely to aid in increasing a successful sourcing. However, the impact on the chance of success of paying more attention to retention of expertise could not be estimated from the available data of our study. In spite of that, we do think that having in-house expertise is crucial for the user organisation. First of all, the expertise is needed to be able to manage the service provider. Next, if the user organisation lacks sufficient expertise, perception of success or failure and factual reality can be far apart, which can turn perceived success into failure once the facts are known. We give a small example of such a situation. Suppose A outsources to B, and B informs A that they found a lot of problems in one of A's systems. They propose a quality improvement at a cost. The perception of A is favourable: they indeed want to improve the quality of the IT-landscape, so they agree. The perception of B is favourable: they get the money to improve the quality. So the perception at both parties is favourable. But hired expertise on behalf of A assessed the system and its evolution history. It turned out that the quality problems were due to changes by B, and as soon as this was known

to *A* their positive perception flipped to negative. Likewise, the perception of *B* turned to negative as well, since they had to solve the problems at their own cost. Key lesson from this example is that one should be aware that it is perception that we measure, which can flip sides once new information has become available. We recall that in our study for none of the cases there were indications that the perception we measured were far from the factual situation, as substantial time passed between the first interview and the concluding interview and we did not observe any dramatic change of mind-set. Additionally, we validated our success measure by comparing our outcomes with four cases that obtained attention in the press. Our outcomes coincided with the actual outcomes.

Business case Also the impact on the chance of success of paying more attention to managing the business case could not estimated from the available data, although a solid business case that is properly managed throughout the project is of course a useful addition.

Transfer of staff/assets Transfer of staff or assets is something that can be part of a deal. An argument that is sometimes heard is that the local IT-personnel will benefit from the new environment of the IT-service provider whose core business is obviously IT. Although this is presumably beneficial for the personnel of the client organisation, the available data of our study did not indicate transfer of staff or assets as a factor that significantly increased a positive course of action for the sourcing. Remarkably, transfer of staff appeared to be a risky endeavour. There were eight cases in which no transfer of staff occurred and seven of these cases turned out to be successful. The number of observations was too small, however, to provide statistical evidence, at least at the 5% level of significance, that transfer of staff would lower the chance on success of the deal.

Client communication Finally, communication within the client organisation turned out not to be a factor with predictive power. It makes sense that effective internal communication would make organisations more successful. So maybe that is true in general, but the impact on increasing the odds for a favourable sourcing deal could not be estimated from the available data of our study. Surprisingly, communication within the supplier organisation did turn out to be a significant factor with predictive power.

6.2. Rigid success factors

The available data did also allow us to test the impact of a number of case characteristics that are rigid on the outcome of a deal. Rigid, in the sense that they cannot be altered anymore, once the deal has been closed. A number of the tests yielded surprising results.

Type of work First, the finding that the type of outsourced work did not turn out to be correlated with success or failure was surprising, considering that outsourcing high-routine work in contrast with work that demands intimate business knowledge would be expected to increase the chance of success. See for example the article of Grover et al. [26] in which the finding of a strong relationship between IT-outsourcing success and outsourcing of systems operations and telecommunication is reported.

Organisation culture Also the (mis)match between the organisation cultures of the service provider and his client was no significant factor in decreasing or increasing the chance of success of a deal. As discussed in Section 2.2 there is an abundance of papers focused on the effect of cultural differences between vendor and outsourcer on the success of an IT-outsourcing deal. See, for example, the literature surveys by Dibbern et al. [17] and Lacity et al. [36]. The conclusions from the literature are not uniform. Generally speaking the conclusion is that cultural differences may matter in the case of offshore outsourcing. In case of domestic outsourcing the effect of cultural differences between outsourcer and vendor usually is of minor importance. We took organisational culture into account by the perceived levels of formality within client and supplier: formal, informal and in between. We found statistically significant evidence that there is no relationship between matching of organisation cultures of outsourcer and vendor and the success of the deal: whether they were similar in formality, or not.

Client motive to outsource Finally, the outsourcer's motive to outsource did not show up as a significant factor that made the difference between failure and success. So short-term cost reduction or guaranteeing operational continuity, access to knowledge and experience of the IT-service provider, focus on core activities, and more such factors did not show a significant influence on whether the deal was perceived to be successful or not. Since we could not find any publication on research into the relationship between the motivation and IT-outsourcing success, we are not able to compare our finding to those of others.

Three rigid factors significant Three case characteristics did show discriminative power: the motivation for insourcing of the service provider, the service provider's empathy, and hiring consultancy support.

Supplier motive to insource The motive of the service provider to obtain the outsourcing deal turned out to be a discriminating factor. There were motives that lowered chances of success and vice versa. For instance, short-term motivations like increasing profit margins or business volume decreased the chance of success, while long-term (strategic) motivations like increasing market share or becoming a player increased the success rate. Our finding is also in line with the finding reported from literature that a strategic partnership between outsourcer and vendor increases the chance on a successful deal [5,26,45,39]. The decision to enter into a strategic partnership of the client can be considered as a long-term motivation of the vendor to obtain the outsourcing deal. All in all, a key-lesson for user organisations is to investigate the true motive of the supplier before signing a deal.

Supplier empathy There were clear indications that the capability of the service provider to imagine himself in the position of his client increases the chance of success of the deal. This finding is in line with the findings of the few articles on this issue we found in our literature scan (see Section 2.2). We are not aware of any publication that contradicts our finding about vendor's empathy capability. Although we decided to classify vendor's empathy capability as a rigid factor one can also argue that it is a controllable success factor. Empathy is an attribute and therefore it should be possible to improve empathy capability. However, we think it takes a long time to change a quality that is a particular part of an organisation and we therefore decided to consider empathy as a rigid factor in our study.

Hiring consultancy support Maybe the most surprising finding was that hiring sourcing consultants is contra-productive: it lowered the chance on success. We found this by investigating the apparent dependency between three variables: (in)experience of the outsourcer, hiring external expertise, and perceived outcome of the deal. It turned out that the degree of experience with sourcing was no convincing factor of success. One would expect that having more experience with sourcing would increase the odds, but this was only weakly present. But, there is a strong statistical indication that hiring consultancy support works contra-productive.

At first glance the finding that hiring consultancy support works contra-productive appears to be surprising. Possible explanations can be as follows. When sourcing consultants take over the driving seat of the sourcing process, the company is not learning and not taking enough responsibility for the endeavour, which can easily lead to failure. Also when there is resistance to change, an outside player is more easily blamed in case of (staged) failure.

We shared our finding with problem owners in large outsourcing deals in the private sector (billions of Euros per annum). They were not surprised at all. They stated that flying in the best external expertise was indeed a reflex upon first outsourcing. But that turned out to be contra-productive. They characterised the expertise of sourcing consultants as overrated. They mentioned lack of insight into the constitutional idiosyncrasies of the organisation that is in transition and lack of understanding of the underlying motives of the outsourcing decision. This led to insufficient understanding of the customer needs and thus limited support in aiding transitions to external IT-service providers. Most of the times rigid models and work processes were adhered to by the junior and medior sourcing consultants who did the actual work. So the blind were leading the blind. This is a typical problem with consulting companies where the business model is that the top-consultants (often partners) bring in an army of well-educated but experience-lacking consultants that follow a relatively simple, coarse-grained and too rigid process. This creates a multiplier for the managing partners, but does not necessarily add enough value for their customers. As a consequence, later deals were mainly engineered by in-house teams that intimately knew the organisation. To bring them up to speed external expertise was used to close the knowledge gaps such as contracting and for reflections to the transition team.

We also shared our finding with an executive involved in a 15–20 million per annum government-to-government outsourcing deal. That hiring external sourcing consultancy is contra-productive in our case studies confirmed their own anecdotal inquiries. They interviewed parties with experience in outsourcing in both the private and public sector. These inquiries gave them the intuition that hiring sourcing consultants was probably not the best idea. This resulted in the decision to keep control in-house. They formed a team to prepare and carry out the transition to another governmental body more equipped for the IT-tasks. For specific aspects such as legal issues, technical knowledge gaps and coaching/reflection, external expertise was hired. This turned out to be successful, and the transition to the other party worked out properly.

We conclude that the motive of the outsourcing company to hire consultancy sourcing expertise is crucial for the effectiveness of the support the company gets. If the reason is that the company does not want to take the responsibility for the sourcing process itself and therefore invites an external consultant, explicitly or implicitly, to effectuate the taken outsourcing decision then the seeds have been sown for later failure. If the motive for hiring external sourcing expertise is to fill specific knowledge and experience gaps in managing the sourcing process by in-house teams, presumably the conditions for an effective consultancy support are created. The lesson for sourcing consultants is that taking over the driving seat of the client's sourcing process may cause a short-term profit but apparently does not lead to good results in the long-term.

7. Conclusions

In this paper we reported about the analyses of data that was collected from 30 pairs of outsourcing and service delivery companies in the Netherlands, that had an IT outsourcing agreement and had entered the service delivery phase.

The participants in our longitudinal multi-client study were interested in the options to take corrective actions during the service delivery phase of the IT-outsourcing life-cycle in order to control the risk of an upcoming failure. Out of nine

potential controllable success factors we found three factors which gave a clear warning signal whether a failure is impending: the communication inside the vendor's organisation; demand management, and strictly working according to the transition plan. Particularly, the finding that the quality of the communication within the supplier organisation is an important controllable success factor is remarkable as opposed to client–supplier communication. Paying more attention to these critical controllable success factors can decrease the chance on failure significantly. Logistic regression was used to estimate the increase of the chance of success due to improving the scores of critical controllable success factors.

Also the tests on the discriminative power of the rigid success factors yielded interesting and surprising results. We found strong statistical evidence that if the motive of the service provider to enter the outsourcing deal is long-term strategically, such as access to the market or increasing their market share, then the chance that the deal will be successful is notably larger than in case of a short-term motive. Also the finding that the type of work that is being outsourced turned out to be insignificant, was a surprise to us.

Surprisingly, the degree of experience of the outsourcing company built up in the past with sourcing did not show to be a factor that positively influenced the chance of success. At first glance the finding that hiring sourcing consultants works out contra-productive appeared to be counterintuitive, but there are plausible explanations for this, so you better not outsource outsourcing.

Acknowledgements

We thank the anonymous reviewers for their valuable comments. This research received partial support by the Dutch *Joint Academic and Commercial Quality Research & Development (Jacquard)* program on Software Engineering Research via contract 638.004.405 *Equity: Exploring Quantifiable Information Technology Yields* and contract 638.003.611 *Symbiosis: Synergy of managing business-IT-alignment, IT-sourcing and offshoring success in society*.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.scico.2016.04.001.

References

[1] Ray Hackney, Akshay Bhagwatwar, Kevin C. Desouza, Considerations for information systems "backsourcing": a framework for knowledge re-integration, Inf. Syst. Manag. 28 (2) (2011) 165–173.
[2] S. Ali, S.U. Khan, Critical success factors for software outsourcing partnership (SOP): a systematic literature review, in: Global Software Engineering (ICGSE), 9th International Conference, IEEE, 2014, pp. 153–162.
[3] Muhammad Ali Babar, June M. Verner, Phong Thanh Nguyen, Establishing and maintaining trust in software outsourcing relationships: an empirical investigation, J. Syst. Softw. 80 (9) (2007) 1438–1449.
[4] L.P. Baldwin, Z. Irani, P.E.D. Love, Outsourcing information systems: drawing lessons from a banking case study, Eur. J. Inf. Syst. 10 (2001) 15–24.
[5] E. Baraldi, J. Proenca, T. Proenca, L. de Castro, The supplier's side of outsourcing: taking over activities and blurring organizational boundaries, Ind. Mark. Manage. 43 (4) (2014) 553–563.
[6] J.A. Bergstra, G.P.A.J. Delen, S.F.M. van Vlijmen, Introducing sourcements, arXiv:1107.4684, 2011.
[7] J.A. Bergstra, G.P.A.J. Delen, S.F.M. van Vlijmen, Stratified outsourcing theory, arXiv:1110.1957, 2011.
[8] Jan Bergstra, Guus Delen, Diana Hoogeveen, Karl de Leeuw, Rob Peters, Ise Vermeire, Bas van Vlijmen, Outsourcing in NL (Op Zoek Naar Outsourcing-succes), Verdonck, Klooster & Associates, 2011.
[9] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd edition, L. Erlbaum Associates, 1988.
[10] Jacob Cohen, A power primer, Psychol. Bull. 112 (1) (1992) 155–159.
[11] W. Alec Cram, Success factors for information systems outsourcing: a meta-analysis, in: AMCIS 2009 Proceedings, 2009, number Paper 554. De proceedingsversie is de enige versie van het stuk, er is geen journal versie. Zie mail van de auteur op 30.11.2011, http://aisel.aisnet.org/amcis2009/554.
[12] Sara Cullen, Peter B. Seddon, Leslie P. Willcocks, IT outsourcing success: a multi-dimensional, contextual perspective on outsourcing outcomes, in: Proceedings of the Second Information Systems Workshop on Global Sourcing: Service, Knowledge and Innovation, January 2008.
[13] Tomi Dahlberg, Mari Nyrhinen, A new instrument to measure the success of IT outsourcing, in: 39th Hawaii International Conference on System Sciences, 2006.
[14] Guus Delen, Decision- en Controlfactoren voor IT-Sourcing, Van Haren Publishing, 2005 (in Dutch).
[15] Guus Delen, 7.5 – decision and control factors for IT-sourcing, in: Jan Bergstra, Mark Burgess (Eds.), Handbook of Network and System Administration, Elsevier, Amsterdam, 2008, pp. 929–946.
[16] S. Deshpande, I. Richardson, V. Casey, S. Beecham, Culture in global software development – a weakness or strength?, in: Global Software Engineering (ICGSE), 5th International Conference, IEEE, 2010, pp. 67–76.
[17] Jens Dibbern, Tim Goles, Rudy Hirschheim, Bandula Jayatilaka, Information systems outsourcing: a survey and analysis of the literature, Data Base Adv. Inf. Syst. 35 (4) (Fall 2004) 99–107.
[18] W.J. Dixon, F.J. Massey, Introduction to Statistical Analysis, international student edition, McGraw–Hill, 1957.
[19] J. Laurenz Eveleens, Peter Kampstra, Chris Verhoef, Quantifying fair payment after outsourcing – a case study, J. Softw. Evol. Process 27 (3) (2015) 147–165.
[20] J.L. Eveleens, C. Verhoef, The rise and fall of the chaos report figures, IEEE Softw. 27 (1) (2010) 30–36.
[21] David Fitoussi, Vijay Gurbaxani, IT outsourcing contracts and performance measurement, Inf. Syst. Res. 23 (1) (2012) 129–143.
[22] G. Gable, A multidimensional model of client success when engaging external consultants, Manag. Sci. 42 (8) (1996) 1175–1198.
[23] D. George, P. Mallery, SPSS for Windows Step by Step: A Simple Guide and Reference, 4th edition, Allyn & Bacon, 2002, SPSS 11.0 update.
[24] Joseph A. Gliem, Rosemary R. Gliem, Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Presented at the Midwest Research-to-practice Conference in Adult, and Community Education, The Ohio State University, 2003.
[25] Jahyun Goo, C.D. Huang, Facilitating relational governance through service agreements in it outsourcing: an application of the commitment trust theory, Decis. Support Syst. 46 (1) (June 2008) 216–232.

[26] Varun Grover, Myun Joong Cheon, James T.C. Teng, The effect of service quality and partnership on the outsourcing of information system functions, J. Manag. Inf. Syst. 12 (4) (1996) 89–116.
[27] Rudy Hirschheim, Mary Lacity, The myths and realities of information technology insourcing, Commun. ACM 43 (2) (February 2000) 99–107.
[28] J.L. Hodges, E.L. Lehmann, The efficiency of some nonparametric competitors of the t-test, Ann. Math. Stat. 27 (2) (1956) 324–335.
[29] G. Hodosi, L. Rusu, How do critical success factors contribute to a successful it outsourcing: a study of large multinational companies, J. Inf. Technol. Theory Appl. 14 (1) (2013) 17–43.
[30] R.V. Hogg, E. Tanis, D. Zimmerman, Probability and Statistical Inference, ninth edition, Pearson Education, 2014.
[31] T.L. Huber, T.A. Fischer, L. Kirsch, J. Dibbern, Explaining emergence and consequences of specific formal controls in is outsourcing – a process-view, in: 47th Hawaii International Conference on System Sciences, IEEE, 2014, pp. 4276–4285.
[32] K. Jablokow, M. Myers, Managing cognitive and cultural diversity in global it teams, in: Global Software Engineering (ICGSE), 5th International Conference, IEEE, 2010, pp. 77–86.
[33] Peter Kampstra, Beanplot: a boxplot alternative for visual comparison of distributions, J. Stat. Softw. Code Snippets 28 (1) (2008) 1–9.
[34] Sung Kim, Chung Young-Soo, Critical success factors for is outsourcing implementation from an interorganizational relationship perspective, J. Comput. Inf. Syst. 43 (4) (2003) 81.
[35] G.P. Kulk, C. Verhoef, Quantifying requirements volatility effects, Sci. Comput. Program. 72 (2008) 136–175.
[36] Mary C. Lacity, Shaji A. Khan, Leslie P. Willcocks, A review of the IT outsourcing literature: insights for practice, J. Strateg. Inf. Syst. 18 (3) (2009) 130–146.
[37] Mary C. Lacity, Leslie P. Willcocks, An empirical investigation of information technology sourcing practices: lessons from experience, MIS Q. 22 (3) (1998) 363–408.
[38] J.-N. Lee, Y.-G. Kim, Effect of partnership quality on is outsourcing success: conceptual framework and empirical validation, J. Manag. Inf. Syst. 15 (4) (1999) 29–61.
[39] J.N. Lee, The impact of knowledge sharing, organizational capability, and partnership quality on is outsourcing success, Inf. Manag. 38 (2001) 323–335.
[40] J.N. Lee, S.M. Miranda, Y.M. Kim, IT outsourcing strategies: universalistic, contingency, and configurational explanations of success, Inf. Syst. Res. 15 (2) (2004) 110–131.
[41] Natalia Levina, Jeanne W. Ross, From the vendor's perspective: exploring the value proposition in information technology outsourcing, MIS Q. 27 (3) (2003) 331–364.
[42] E. Lioliou, A. Zimmermann, L. Willcocks, L. Gao, Formal and relational governance in it outsourcing: substitution, complementarity and the role of the psychological contract, Inf. Syst. J. 24 (2014) 503–535.
[43] E. Méndez, L. Mendoza, M. Pérez, Critical success factors as a strategy for risk mitigation in it outsourcing projects, in: Americas Conference on Information Systems (AMCIS) 2006 Proceedings, 2006, number Paper 396.
[44] M. Nasir, S. Sahibuddin, Critical success factors for software projects: a comparative study, Sci. Res. Essays 6 (10) (2011) 2174–2186.
[45] Nilay V. Oza, Tracy Hall, Austen Rainer, Susan Grey, Trust in software outsourcing relationships: an empirical investigation of Indian software companies, Inf. Softw. Technol. 48 (2006) 345–354.
[46] Z. Perunović, M. Christoffersen, R.N. Mefford, Deployment of vendor capabilities and competences throughout the outsourcing process, Int. J. Oper. Prod. Manag. 32 (3) (2012) 351–374.
[47] A. Plugge, H. Bouwman, F. Molina-Castillo, Outsourcing capabilities, organizational structure and performance quality monitoring: toward a fit model, Inf. Manag. 50 (6) (2013) 275–284.
[48] Mark J. Power, Kevin C. Desouza, Carlo Bonifazi, The Outsourcing Handbook: How to Implement a Successful Outsourcing Process, Kogan Page, London and Philadelphia, 2006.
[49] Symbiosis Project Team, Symbiosis database, http://easy.dans.knaw.nl, 2012.
[50] A.C. Rouse, Explaining i.t. outsourcing purchasers' dissatisfaction, in: PACIS 2006 Proceedings, 2006, Paper 1.
[51] C. Schwarz, Toward an understanding of the nature and conceptualization of outsourcing success, Inf. Manag. 51 (1) (2014) 152–164.
[52] H. Smuts, A. van der Merwe Kotzé, M. Loock, Critical success factors for information systems outsourcing management: a software development lifecycle view, in: SAICSIT'10 Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, ACM, 2010, pp. 304–313.
[53] N. Urbach, T. Würz, How to steer the it outsourcing provider. Development and validation of a reference framework of it outsourcing steering processes, Bus. Inf. Syst. Eng. 5 (2012) 247–259.
[54] C. Verhoef, Quantitative aspects of outsourcing deals, Sci. Comput. Program. 56 (3) (2005) 275–313, Available via www.cs.vu.nl/~x/out/out.pdf.
[55] S.F.M. van Vlijmen, R. Peters, Codebook to the symbiosis database, http://easy.dans.knaw.nl, 2012.
[56] S.F.M. van Vlijmen, R. Peters, The cross-reference of symbiosis' questions and database columns, http://easy.dans.knaw.nl, 2012.
[57] S.F.M. van Vlijmen, R. Peters, Symbiosis' original questionaires in Dutch and English, http://easy.dans.knaw.nl, 2012.
[58] Leslie Willcocks, John Hindle, David Feeny, Mary Lacity, IT and business process outsourcing: the knowledge potential, Inf. Syst. Manag. 21 (3) (2004) 7–15.
[59] J. Winkler, J. Dibbern, A. Heinzl, The impact of cultural differences in offshore outsourcing – case study results from German–Indian application development projects, Inf. Syst. Front. 10 (2) (2008) 243–258.
[60] R. Zaal, Sourcingtrend zet in 2007 onverminderd door – Opnieuw meer meerjarige IT-uitbestedingen in Nederland, Autom. Gids (34) (August 22, 2008).
[61] R. Zaal, Deals die expireren, Autom. Gids (35) (August 29, 2008).
[62] R. Zaal, Deals overzicht 2009, Autom. Gids (19) (May 7, 2010).
[63] R. Zaal, Outsourcings in 2010: Minder, kleiner, korter, Autom. Gids (17) (April 29, 2011).