



Foundations for measuring IT-outsourcing success and failure

G.P.A.J. Delen^a, R.J. Peters^b, C. Verhoef^b, S.F.M. van Vlijmen^{c,*}

^a Informatics Institute, University of Amsterdam, Science Park 904, Amsterdam 1098XH, the Netherlands

^b Department of Information Management and Software Engineering, VU Amsterdam, De Boelelaan 1081a, Amsterdam 1081HV, the Netherlands

^c Informatics Institute, University of Amsterdam, Science Park 904, Amsterdam 1098XH, the Netherlands

ARTICLE INFO

Article history:

Received 20 March 2018

Revised 13 June 2019

Accepted 16 June 2019

Available online 17 June 2019

Keywords:

IT-outsourcing

Success determinants

Failure determinants

Outcome-prediction

Odds-improvement

ABSTRACT

We implemented five easy-to-complete questionnaires in Excel, which could serve as early warning signals for practitioners interested in the odds of their IT-outsourcing deals and could serve to redirect their course when still possible. The questionnaires are based on our earlier published longitudinal, observational study on 30 representative ITO-deals in the Netherlands, of which we know whether they failed or not. Our questionnaires predicted their outcome correctly. To help redirect the course of a dubious deal, we developed a questionnaire estimating the odds in relation to boosting strongly significant critical success determinants. Another questionnaire guides practitioners how to further improve on less critical factors. There are no specific reasons that limit our results to the Dutch situation, which makes it promising, therefore, to apply the Excel as an aid in improving ITO deals in other contexts.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

In 2016 we published an article in *Science of Computer Programming* which dealt with the research findings from a longitudinal study on IT-outsourcing (ITO) deals in the Netherlands (Delen et al., 2016). About 60 organisations participated: clients (also called outsourcers), suppliers (also called vendors) and intermediaries (also called sourcing consultants). The research sample is a representative cross-section for 700 IT-outsourcing deals in the Netherlands. Representativeness was statistically proved through validations that the sample reflected the Dutch economic sectors, the duration of the deals and the type of outsourced work of the total Dutch ITO-deal population reasonably well. For more details, see Delen et al. (2016). This is a very important result, for it implies that findings of the sample generalize to the entire population. So our Excel-tool can be used for the entire population. Not everybody will immediately realize what this actually means, so we shall elaborate on this fundamental statistical rule below.

When statistical tests are used we usually accept a 5% chance that although the null hypothesis is true we still reject it. The probability of making that (type I) error is often called α . Vice versa when the null is false and we fail to reject it, this is called a type II error, and the probability for such errors is often called β and we usually accept a 20% chance of making a type II error. So

the probability of rejecting the null hypothesis when it is false is $1 - \beta$ (i.e. 80%), which is called the power of the test. Only if a sample is large enough for the effects we want to detect will you have enough power. In the longitudinal study we carried out all the required power tests. This implies that the results of our sample generalize to the entire population. This type of research is, to the best of our knowledge, not very common in our field.

Since the 1960s power analysis has been an established field (Cohen, 1988), but in our field applications are not yet manifold. As the founder of power analysis stated in the second edition of his seminal book: “it is clear that power analysis has not had the impact on behavioral research that I (and other right-thinking methodologists) had expected. But we are convinced that it is just a matter of time.”

Cohen wrote these words in 1988, and in medical science power analysis is at present more the norm than in our field. When we test the effectiveness of a medicine on a representative sample and the tests have sufficient power, this is a prominent reason to release a medicine onto the market for the entire population. New testing on new cases is not necessary. Of course, this is not 100% foolproof but these days it is rare to have drugs admitted that have severe adverse effects such as Softenon had in the 1950s when clinical trials weren't as complete (Kelsey, 1965). Likewise, if you have a representative sample of 30 cases and the effects you want to detect indeed have been detected through sufficiently powerful tests, further testing on more (new) cases is not necessary. Then the detected effects apply to the entire population.

* Corresponding author.

E-mail addresses: x@cs.vu.nl (C. Verhoef), s.f.m.vanvlijmen@uva.nl (S.F.M. van Vlijmen).

Table 1

Overview of the predictive power of all factors tested in our study. Legend: positive = positive influence of the factor on the success; negative = negative influence of the factor on the success; not found = very weak or no noticeable effect found in our study, weak positive = weak effect but not quantifiable in terms of an improved success-chance.

Type	Description	Effect on success
controllable	Working according to the transition plan	positive
controllable	Managing the business case for outsourcing	weak positive
controllable	Managing the business case of the service supplier	not found
controllable	Transfer of staff	weak positive
controllable	Transfer of assets	weak positive
controllable	Demand management	very positive
controllable	Retention of expertise	not found
controllable	Communication inside client organisation	not found
controllable	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 intermediary support	negative

So we do not need extra cases or more research to validate our results: this validation was part of the research design.

The thirty representative deals chosen amounted to a total cost of over 100 million euros, they had a combined contract-span between 1998–2016, and there were eight intermediate contracts. In total 28 thousand data points were collected through 5 questionnaires (516 questions in total) and statistically analyzed with both exploratory data analyses and more classic hypothesis-driven statistics. We measured a 40% failure rate for these representative engagements. Because of the representativeness of the sample the ITO failure rate of the entire population of ITO deals in the Netherlands at that time can be estimated at 40% as well. The study can be characterized as an *observational study*. The main objective was to gain an insight into what factors make ITO deals a success or a failure in the perception of the partners involved, the outsourcer and his service provider(s).

We reported extensively how observations on the 30 ongoing IT outsourcing cases were recorded during 4 years of observation (Delen et al., 2016). Most cases were in progress and had not yet expired during and after the observation period. So their actual outcome was not known, but for determining a success factor we needed an indication whether they would become a success or not in the first place. We therefore designed questionnaires to measure outcome perception en route. We measured outcome perception more than once during the observational period for both outsourcers and service providers. To be on the safe side, we postponed publication of our observational study until 25 cases had been finalized and their actual outcome was known. It turned out that our success-perception model predicted the actual outcome of success or failure 100% accurately. After publication, it occurred to us that such a questionnaire—once robustly developed—could be a useful tool to quickly provide practitioners with a valuable insight into how a deal in progress is doing.

The investigated success determinants. In the ITO literature a large number of potential success determinants shows up, but, to the best of our knowledge, of none of them the power to discriminate between successful deals and failures has been tested through a longitudinal study in which practitioners and academic researchers cooperated in a large research effort. We distinguished between influenceable (controllable) success determinants and success determinants that are a given fact (rigid) at the start of an ITO-deal and cannot (or hardly) be changed en route. We defined a controllable success factor as a factor that can be changed by the outsour-

ing company and/or its service provider during the service delivery phase in order to reduce the chance of failure. Factors that cannot be changed en route are called rigid factors. After much debate with the companies participating in the study, it was decided to focus on 15 factors that, later on, we split into nine controllable and six rigid factors. The selection was a mix of factors discussed in literature studies, input from the participating organisations, and our own very extensive industrial experience with outsourcing deals and their failure and success. Our first author, for instance, worked for many years as an IT-sourcing consultant. For more details, see our observational study (Delen et al., 2016).

In Table 1 we recall an overview of all the success determinants we tested. This table displays the nature of the factors (controllable or rigid), the investigated factors themselves and their significance, if any, to the outcome of a deal. In our observational study we detected three influenceable (controllable) success determinants that can make or break an ITO-deal and three others of which a weak effect on the outcome of the deal could be proved. Three controllable factors were not at all significant. Of the rigid factors three turned out to be significant and three were not at all significant. For instance, *The motive of the outsourcer to engage in an IT-outsourcing deal* is a rigid factor: this motive does not change easily. It turned out that whatever the motive (cheaper, better service, cash, etc), the deal-outcome was independent of it. So not one particular (type of) motive of the outsourcer did stand out as a successful or failing motive. We will not discuss the insignificant factors at length since they are not going to help in improving the deal-outcome. For more information on insignificant factors, readers are referred to our observational study (Delen et al., 2016).

The significant factors in Table 1 are defined in our Excel. Per factor we have specific questions, and their answers are on a (numerical) Likert-scale. From the numerical data, an algorithm is used to give an outcome. All questions and calculations are implemented in the accompanying Excel (Delen et al., 2017). For instance, *Demand Management* is about managing the service provider taking over the outsourced IT functions. Although all kinds of different meanings can be attached to *Demand management*, in our case it is defined via questions scored on a Likert-scale and an algorithm giving an end-score between 0–100, the higher the score the better the performance of the factor. The score of *Demand Management* is based on thirteen questions about demand management. Six questions for the outsourcer and seven for the service provider. For all the other significant factors the same holds: they are defined similarly in the accompanying Excel (Delen et al., 2017).

Three significant influenceable success determinants. In total we tested nine potential influenceable/controllable success determinants (see Table 1). For three of them a strong power to discriminate between successful cases and cases that failed was found. They are:

- Working according to the transition plan;
- Demand management;
- Communication inside supplier organization (i.e., not between clients and suppliers).

The finding that these three particular controllable success determinants can make or break a deal is of both theoretical and practical interest. It allows outsourcers and service providers to focus on improving the performance of these three success determinants. Because of this, the joint efforts to increase the chance of success can be reduced to a minimum.

Three significant rigid success determinants. In our observational study we also detected three significant factors that are a given fact at the start of an ITO-deal and which lower or enlarge the chance of a successful outcome of the deal (see Table 1). We will explain them below:

- The motive of the supplier to close the deal: success chances are increased in case of strategic motives (increasing brand awareness, increasing market share, etc.) that will pay off in the long run, and decreased in case of non-strategic (financial) motives that mostly pay off in the short run and can easily be monetized. Of course, you can have more than one motive, and in our Excel we measure the dominant motive with a dominance metric, based on answers to questions.
- The empathy capability of the supplier (the capability to place himself in the position of his client). More empathy increases the chance of success and low empathy decreases the chance of success. This, too, was implemented with questions for both parties as well as by measuring the distance of the answers between the two parties.
- Hiring a sourcing consultant (intermediary) or not. Intermediaries turned out to be counterproductive: not hiring them increased success chances.

The first two factors are defined in the accompanying Excel (Delen et al., 2017) as discussed earlier. The third is trivial to measure: either you have or you have not hired an intermediary. These critical rigid factors are a firm warning signal for the outsourcer before closing ITO deals.

It might be expected that suppliers want to conclude long-term contracts and customers short ones. But in our observational study we did not find that suppliers usually favor long-term contracts and clients short ones. For instance, if an outsourcer has opted for IT-outsourcing in order to be able to focus more on his core business processes, he will aim at concluding a long-term contract rather than a short one. Suppliers, too, will not always strive for long-term contracts. For reasons of risk diversification or short-term profit, suppliers may opt to conclude more short-term contracts with different clients, as dependency on one client (e.g. high asset specificity) may be too risky for them. All in all, be aware of a larger chance of failure if the dominant motive of the service provider to close the deal is short-term financial instead of long-term strategic and/or his empathy capability is assessed as low. Also the service provider should appreciate the fact that short-term financial motives exert pressure on the cooperative success.

Finally, in half of the cases a consultant was involved in the decision phase of the outsourcing process to support outsourcers in negotiating the best deal with their vendors (Delen et al., 2016).

The observational study has proved that it is a bad idea to out-source outsourcing. Hiring sourcing consultants (intermediaries) turned out to be counterproductive: it lowered the chance of success.

Practical implications. We published the fundamentals of the observational study for an academic audience, statistics included (Delen et al., 2016). Although the findings are challenging and on occasion surprising, the article did not focus on practical applications of the results. The observational study does, however, enable the development of suitable measurement and prediction tools, so that practitioners can measure outcome perception themselves and redirect the course of the deal if necessary. For this article we justified and developed these tools, and implemented them in an accompanying Excel. For instance, the likely outcome of an ITO-deal in progress can be predicted using our Excel. It contains a readme for all the developed and discussed tools in this article.

Of course, learning what makes or breaks a deal excluded intervening during the observation period, so in our previous research we did not intervene, just observe. The focus of this article is the opposite: we provide tools to measure and, if necessary, intervene and redirect the course of a deal in case there is a serious chance of its derailing.

Although being aware of the three particular success determinants for improvement is already of value to the practitioner, just knowing about them through our earlier study does not suffice. We must be able to measure these discriminants as well as the impact of an intervention on the performance of these critical risk drivers. The only way to gain the desired insight is to measure the performance before and after intervention. So, for the present article we developed questionnaires for precisely this purpose, using the results of our observational study. Practitioners will now be able to easily calculate how much the chance of a successful deal has increased as a consequence of their interventions. Just to be sure, the measurements might be repeated after some time.

It may not be obvious how to measure the empathy capability of the service provider. Nor may it be simple to determine what the main motive was of the service provider to close an ITO deal. In the present article we provide practitioners with tools that they can use to accomplish this. To keep things simple we hid all calculations in the background of the Excel: all the user is requested to do is answer a number of questions by ticking the listed options, and the outcome is presented.

In this article we provide a set of tools that practitioners may use to further improve their insights into the success chances of ITO-deals. All the relevant measurement tools are implemented in a ready-to-use Excel spreadsheet that accompanies this article. Each tool consists of a small subset of the 516 different questions used to collect the 28 thousand data points. We required only 12% of the 516 questions to develop the Excel. We must point out once again that this tooling was not present in the observational study, in which we analysed the vast amount of data using statistical handwork. The Excel helps in predicting the outcome of a deal in progress and in intervening in it if necessary.

It is advisable to add some observations to the tool. The tool has been developed to enable practitioners to profit from the results of the observational study we carried out. Think of the tool as a ready-to-use risk analysis tool. We used IT-qualified interviewers to complete our much more extensive lists of questions. When necessary, the interviewers clarified the objectives of the questions to the persons being interviewed. We suggest that the role of intermediaries should be reshaped by the use of our tool in order to help both parties improve on their cooperative success. After all, this is the main reason why you would want to hire an intermediary. And, of course, intermediaries are usually very knowledgeable about the questions in our Excel-tool.

In this article we also unveil characteristic patterns in the behavior of outsourcers and service providers in cases of success or failure. For example, outsourcers are inclined to be significantly more negative than suppliers in cases of failure, whereas in successful cases clients are significantly more positive than suppliers. All this is of significant value to anyone interested in putting into practice the findings published in the observational study. Also, the timing of the interviews with the outsourcer and the service provider is an issue of great practical relevance. Two rounds of interviews took place and a significant difference between the first and second round of answers was detected and reported on in this article. This issue and many more issues for practical usage were neither the objective nor did they fall within the scope of the observational study described in our earlier publication (Delen et al., 2016), but are clearly of great interest to practitioners who want to put into practice the findings published in that article.

To sum up, the focus of the present article is on the development of tools for practical use based on our findings published in our observational study. In this article we also focus on the analysis of typical answer patterns for further guidance so that practitioners know what type of answers to expect. Many years of observing deals have led to results that can now help in lowering their failure chances. If we had to use statistics to justify our arguments, we offered them separately whenever possible, so that the practical nature of this article will not get overshadowed.

2. What is ITO success anyway

There is no sense in discussing success and failure determinants when the notion of success has not been properly defined. For example, Standish's Chaos reports are often referenced for their success rates and determinants despite the fact that these rates have turned out to be meaningless (Eveleens and Verhoef, 2010). We will briefly explain: when Standish's (plan accuracy) definitions of successful and challenged projects were used, good projects of a best-in-class company were rated unsuccessful. A different company with disastrous plan accuracy was rated highly successful according to their definitions. The Standish determinants are presumably inferred from their meaningless numbers, so the success-factors for the first company were taken as failure-factors and the failure-factors for the second company interpreted as success-factors. It is not a good idea to use success determinants which potentially have opposite effects.

Similar work by others suffers from the same problems as Standish's work: although many researchers recognize the presence of subjective components in forecasting important IT-related Key Performance Indicators, no one has incorporated bias into comparisons between (initial) forecasts and actuals (Eveleens and Verhoef, 2009). Such comparisons are often part of their definitions of success, although bias can dangerously influence research results, up to the level of utter nonsense as Standish shows.

Bapna et al. (2010) defined the success of an IT-outsourcing deal as a contract that is extended or expanded, and failure as a contract that is renegotiated or cancelled. But this is still not the entire story: you can rescue a derailing deal by renegotiating it, like in cases of winner's curse. Also, failure does not always mean cancellation. In one of our cases we predicted failure, and it did turn out to be an outright failure, and yet it had not been cancelled due to denial on the part of the client. Parts of the system were put into production (to save faces). Or suppose that the client is locked-in; swapping the supplier or back-sourcing would only be possible at great expense and/or much loss of time. In this context a continuation would mean success according to the definition of Bapna et al. (2010), which is not in line with reality. Our definition does take such situations into account: if one of the parties

perceives the deal as a failure the deal is considered a failure. In our Excel the decision rule takes the minimal value of the success-perception score of the outsourcer and that of the supplier.

Barthélemy (2003) asked managers who were responsible for an outsourcing deal to rate their satisfaction on a five-point Likert scale, and partitioned success as very/totally satisfied and as failure otherwise. However, the suppliers' relevant opinion was not taken into account. Limiting ourselves to asking the management's perception only of the outsourcer's organisation will lead to a one-sided view as is illustrated by the examples we just mentioned.

The approach we adopted measures the outcome perception of both parties involved (Delen et al., 2016), rather than invent some quantitative measure of success based on plan accuracy or something of that sort. Our approach is in line with Barthélemy and other literature that favors the perception measurement approach, with the proviso that the outcome perception of both parties involved is measured and avoids the pitfalls of a one-sided view.

3. Outcome prediction

We designed two perception-based questionnaires to predict the outcome for an ongoing deal, one for outsourcers (10 questions) and one for suppliers (9 questions). We used intuitive statements ranging from strongly disagree to strongly agree on a classical 5-point Likert scale, which provides six answer options of which only one can be selected: strongly disagree, disagree, neutral, agree, strongly agree, or not applicable (encoded 0, 25, 50, 75, 100, NA). We implemented the perception-based questionnaires in a simple-to-complete Excel (Delen et al., 2017). Below we provide a few questions to give the reader an idea.

Some questions for clients.

- I trust my supplier
- I am satisfied with the financial results of the sourcing
- We receive the services as agreed upon in the current contract and/or SLA
- I am satisfied with the quality of the service

Some questions for suppliers.

- I trust my customer
- I am satisfied with the results of the sourcing
- I am satisfied with the quality of the demand management of the outsourcer

Remember that possible answers are ranging from strongly disagree to strongly agree. We designed the questions with a particular goal in mind and validated their consistency by computing the Cronbach alpha reliability coefficient (Delen et al., 2016), which turned out to be high enough (Gliem and Gliem, 2003) to trust the consistency.

Our model predicts a failure if the minimum of both scores (clients and suppliers) is below 50, and a success if otherwise. So, to predict the outcome of a deal the minimum of both scores is taken, and not the average. Suppliers might be tempted to deem this decision rule unfair when confronted with a failure prediction if the client's score is too low for success but theirs is not. The minimum value corrects for potentially overly optimistic answers by suppliers in cases of failure and by clients in cases of success. We predicted 12 failures and 18 successes, so a failure rate of 40% in the Netherlands since the sample is representative (Delen et al., 2016). Table 2 summarizes the scores for our 30 cases: its identifier, both calculated scores, the predicted outcome, and the actual status at the time of writing this article. Note that the calculated scores are not simple averages of answers to multiple questions.

Table 2
Empirically found client and supplier scores, predicted outcomes and present status of the deals.

Case	Supplier	Client	Prediction	Outcome	Case	Supplier	Client	Prediction	Outcome
2145	75.00	80.99	success	success	2161	38.89	32.14	failure	failure
2146	63.89	40.44	failure	unknown	2162	83.33	80.97	success	success
2147	44.44	67.45	failure	failure	2163	61.11	68.68	success	success
2149	55.56	66.29	success	success	2164	72.22	42.55	failure	failure
2150	66.67	76.49	success	success	2165	69.44	70.60	success	success
2151	69.44	78.35	success	success	2166	61.11	69.23	success	success
2152	69.44	71.61	success	success	2167	69.44	63.10	success	success
2153	72.22	49.26	failure	failure	2168	55.56	54.69	success	success
2154	77.78	65.48	success	unknown	2169	50.00	40.36	failure	failure
2155	46.43	79.09	failure	failure	2170	66.67	35.71	failure	failure
2156	65.62	75.74	success	unknown	2172	52.78	38.10	failure	failure
2157	66.67	56.25	success	success	2200	53.57	23.74	failure	unknown
2158	22.22	26.93	failure	failure	2201	52.78	66.22	success	success
2159	55.56	29.84	failure	failure	2203	72.22	96.88	success	success
2160	55.56	59.90	success	success	2204	61.11	66.15	success	unknown

Related questions are averaged, and these averages are in turn also averaged. The exact algorithms can be fully appreciated by inspecting the hidden calculations in the accompanying Excel (Delen et al., 2017).

Knowing whether a prediction model predicts actual outcomes, means following the IT-outsourcing deals until their actual outcome is known, which can take long. Our research started in the 1990s. This led to an elaborate research program with about 60 carefully selected industrial partners and a research grant of 1.5 million euro in 2006. The 30-deal contract-span started in 1998 and ended in 2016, with eight deals without due-date but with a reasonable notice period. The status in 2016 was that we knew the actual outcomes of the 25 deals, and that our model had predicted them all correctly (10 cases of failures, 15 cases of success, which means an actual failure rate of 40% as well). For three cases the deal ended before the final interviews. They were predicted correctly (but obviously post-hoc).

We implemented our two success-perception questionnaires in Excel, which in total contained 19 questions, 10 for the client and 9 for the supplier (Delen et al., 2017). One of the questions for the outsourcer is about the quality of the service of the supplier. In case clients need extra guidance on how to determine this, we provided 14 sub-questions and used their average as outcome.

Characteristic patterns. When using our two perception-based questionnaires, practitioners may expect to find different opinions between parties. In Fig. 1 we depict the difference between the scores of client and supplier perception on the vertical axis, and their minimum on the horizontal axis. The minimum is of course the outcome prediction. The vertical red line at 50 separates between (predicted) success and failure: if the minimum of both scores is below 50 we call it a failure, and otherwise a success. The horizontal red line at zero shows who is more negative; the client or the supplier. Two patterns become visible: in cases of

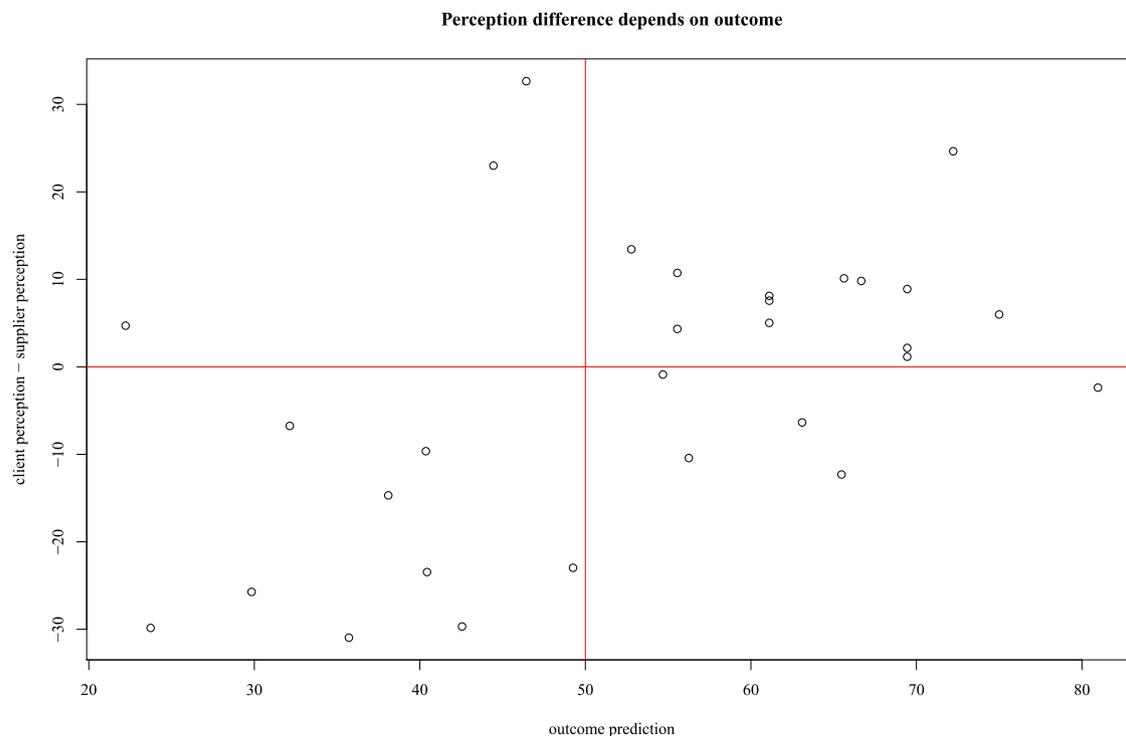


Fig. 1. Perception difference depends on outcome.

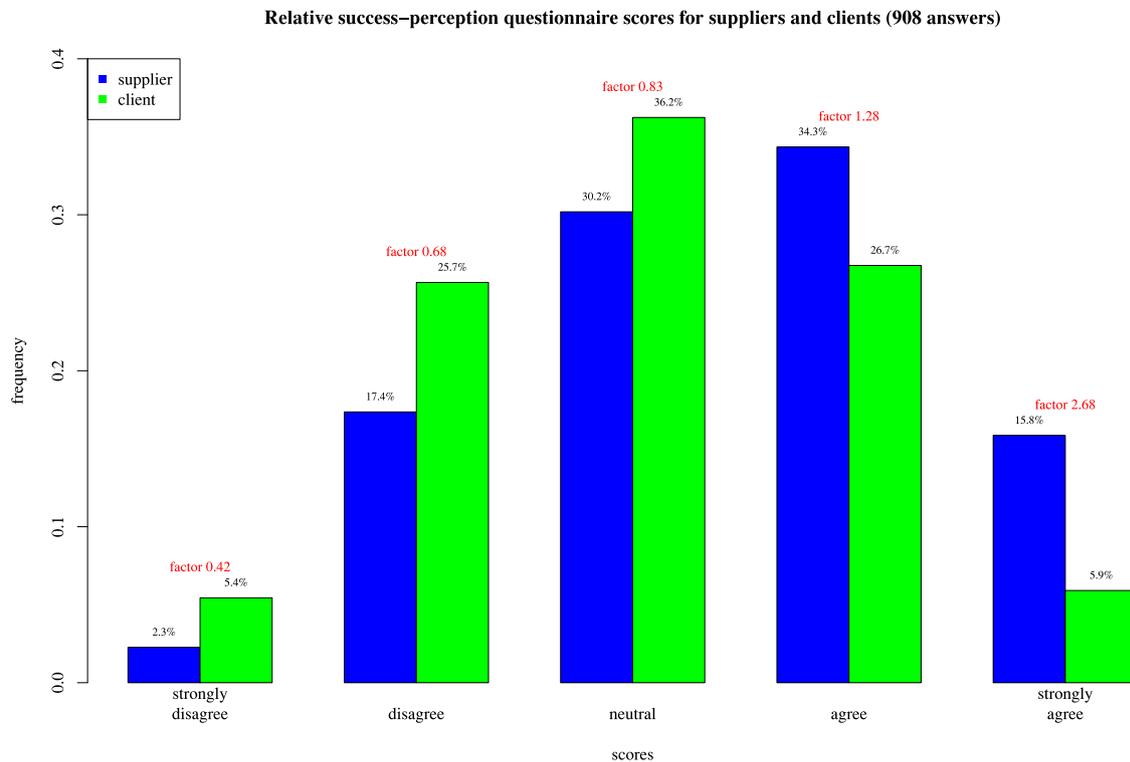


Fig. 2. Relative success-perception scores for suppliers and clients.

failure clients are significantly more negative than suppliers, and in successful cases clients are significantly more positive than suppliers.¹ A reasonable explanation for these effects may be that in the case of predicted failure, clients perceive they are paying for potentially nothing, whereas their suppliers are still earning and don't perceive any problems. In the case of predicted success the supplier needs to work harder, and may be earning less: no pain no gain as the saying goes. That is why asking the perception of the outsourcer organisation only as in (Barthélemy, 2003), leads to a misleading view.

How much both parties differ in perception is split into the five possible answer-categories. Fig. 2 aggregates almost a thousand success-perception questionnaire answers given by both suppliers and clients during our observational study. As can be seen in the figure, suppliers answered *strongly disagree* in only 2.3% of the cases, whereas clients answered *strongly disagree* twice as often (factor 0.42). This reverses for the two most positive answers, e.g., suppliers answer *strongly agree* a factor 2.7 more often than clients. Suppliers are overall more positive than clients: in cases of negative scores clients are much more negative and in positive cases suppliers are much more positive.

Although suppliers are more positive than clients, this does not imply that they always answer on median higher than 50. There is

Table 3
Descriptive statistics answers by suppliers.

Question	Min	1st qu.	Median	3rd qu.	Max
1	25	75	75	100	100
2	25	50	75	100	100
3	25	50	75	100	100
4	25	75	75	87.5	100
5	25	75	75	75	100
6	0	25	25	50	75
7	25	50	50	75	100
8	25	50	50	50	75
9	0	50	50	50	75
all	0	50	75	75	100

an exception. Table 3 splits supplier answers per question. The abbreviations stand for the lowest value (minimum), the lowest 25% (first quartile), 50% of the data (cut in half, hence median), then up to 75% (third quartile) and up to the highest value (maximum). On median all answers score at least 50: neutral, agree or strongly agree—except for one at 25, which means suppliers on median disagree with the following statement: I am satisfied with the quality of the demand management of the outsourcer. The overall median score is 75 (agree) so disagree is a very low median score. Demand management had already been found to be a strongly discriminating factor between failure and success, and now the supplier's perception adds to the finding that there are ample opportunities for improvement on the client-side.

Practical guidelines. Decision makers will naturally react to predicted failure by assuming that the solution lies in regaining trust between clients and suppliers. A good conversation will supposedly clear the air and working on the ITO-deal can be resumed as before. This is neither a solution nor very helpful. This is just a case of the *deaf effect* (Robey and Keil, 2001): the failure of decision makers to act properly on very clear risk warnings. See Nuijten et al. (2016) for better understanding of this effect and how

¹ We tested this with a paired *t*- and Mann-Whitney test, because of a borderline *p*-value for normality in one case near $\alpha = 0.05$. For the four partitions of clients/suppliers and failing/succeeding projects, the Shapiro-Wilk test for normality provides *p*-values: 0.0592, 0.643, 0.397 and 0.635. Client scores for failing projects are just above α . By taking 0.05, all partitions are normally distributed, but at $\alpha = 0.1$ one partition is not, so we can use the Mann-Whitney test that does not assume normality as the *t*-test does. Pooling all data both effects cancel each other out using *t*- and Mann-Whitney tests respectively: $p = 0.562, 0.715$. Restricted to cases of failure, the one-sided *t*-test is $p = 0.049$. Given that the normality check for this case is just above α , we carried out a Mann-Whitney test too, which is significant at 0.1 ($p = 0.065$). So, accepting at 0.05, the *t*-test confirms our hypotheses, and at 0.1, the Mann-Whitney test supports the hypotheses. For the success cases the significance is below α either way: $p = 0.025, 0.033$.

to deal with it. What should be done is improvement of significant success factors. We shall discuss them next.

4. Intervention

A negative outcome-prediction en route does not mean inevitable failure at the end. Boosting controllable significant factors increases the odds. Of course, this is easier said than done. How good is the deal now, and how much it must be improved is discussed below.

4.1. Boosting controllable significant factors

We inferred a logistic model predicting the success-chance for which the same three controllable factors turned out to be significant as in the success-perception questionnaire (not a surprise from a statistical viewpoint). The model calculates how much each factor contributes to the success of a deal. The success-chance questionnaire is implemented in our Excel (Delen et al., 2017). We hid the mathematics from the user whom we simply will ask to answer all the questions by ticking one of the given possibilities, and the success-chance will be calculated in the background and subsequently presented. We recall the logistic formula below for explanatory reasons (Delen et al., 2016).

$$\ln\left(\frac{p}{1-p}\right) = -36.7664 + 0.06806W + 0.22886D + 0.22926C$$

(13.953) (0.050) (0.101) (0.102)

Here, p is the success-chance (\ln is the natural logarithm). In total, 26 answers score the three factors: working according to plan (W), demand management (D), and communication inside the supplier organisation (C). The standard errors of the estimated regression coefficients are in brackets. They are not small, so the outcome is a signal and not the absolute truth as could be used in a

court case. We have not implemented error margins in Excel for simplicity's sake. The coefficients show that the first factor has the least impact: so a full focus on working according to plan is not the best strategy. The latter two factors play the most important role in contributing to success upon improvement.

To give you an idea of outcomes, we provide in Fig. 3 the success-chance p for $W = 80$ and varying D and C (all between 0–100). All the possible demand management scores are displayed horizontally, the scores for the internal communication in the supplier-organisation are on the depth axis, and the chance of success, denoted by p (between 0–1), on the vertical axis. The surface of Fig. 3 is mainly red, which means a low success-chance. The grey surface is relatively small and means high success-chances. A low score on demand management lowers success-chances considerably, which might explain a lot of the high failure rate of 40% that we measured. At the heart of proper demand management lie strong requirements engineering and, of course, streamlining user requests and client projects.

We advise improvement upon all three critical success determinants to the higher ranges. As our research suggests, it will pay off. Leading predicted failure out of the danger zone requires significant effort: it needs improvements on both the client-side and supplier-side, as well as a plan how to cooperate. It is not merely a matter of regaining trust between parties.

Table 4 shows what scores to expect per question based on 780 answers where sometimes averages were taken from multiple answers to the same question by different persons in different roles, for we know that they can rank differently (van Genuchten, 1991). Both outsourcers and suppliers answer on median around 75 (agree) to almost all questions. One exception is formed by their answer to the question:

- The SLA has been adjusted appropriately to the changing business needs.

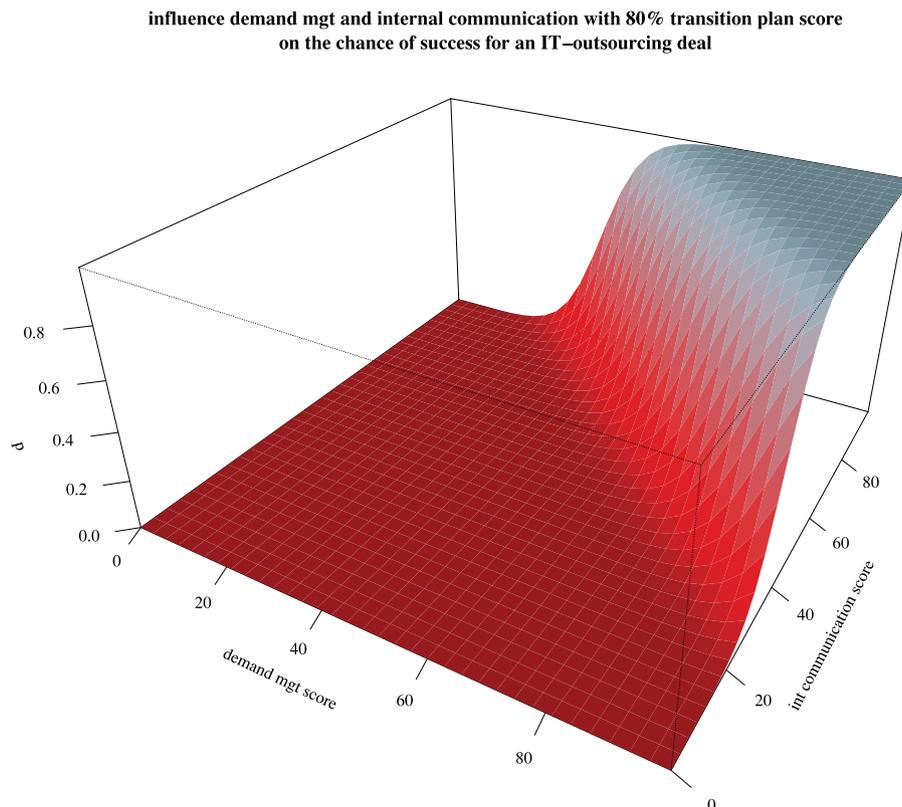


Fig. 3. Impression of the model (see www.cs.vu.nl/~x/pred/pred.mp4 for an animation where W also varies from 0–100).

Table 4

Descriptive statistics of success-chance scores per question (qu. = quartile, #obs = number of observations).

Question	Who	Min	1st qu.	Median	3rd qu.	Max	#obs
w1	out	25	75	75	75	100	66
w2	out	25	75	75	75	100	61
w3	out	25	75	75	75	100	61
w4	sup	0	68.75	75	91.67	100	62
w5	sup	25	50	75	87.5	100	63
w6	sup	50	75	75	83.33	100	57
w7	sup	50	71.875	75	82.29	100	57
w8	sup	50	62.5	75	87.5	100	54
d1	out	25	70	75	91.67	100	50
d2	out	12.5	75	75	85	100	50
d3	out	25	68.75	75	83.33	100	50
d4	out	41.67	75	75	75	100	51
d5	out	41.67	50	75	75	100	50
d6	out	0	25	37.5	75	100	54
d7	sup	25	66.67	75	81.25	100	50
d8	sup	0	58.33	75	79.17	100	50
d9	sup	0	75	75	87.5	100	50
d10	sup	33.33	59.375	75	82.735	100	52
d11	sup	33.33	57.29	75	82.29	100	52
d12	sup	25	25	50	75	100	53
d13	sup	25	75	75	75	100	50
c1	sup	60	75	75	91.67	100	50
c2	sup	0	50	75	75	100	51
c3	sup	41.67	75	75	87.5	100	50
c4	sup	33.33	54.17	71.875	83.33	100	50
c5	sup	50	75	81.25	90.625	100	52

Clients answer this question on median with 37.5 (question d6 in the Excel) and suppliers answer this question on median with 50 (question d12 in the Excel). Again, this is a strong signal that demand management is problematic. Adjusting agreements is an essential part of contract management, which in turn is part of demand management. Apparently, once made, agreements are not updated when necessary. The other questions on demand management are rated more positively by both parties. In the success-perception questionnaire satisfaction with demand management has been rated, and the outcome is a low score on median. In the success-chance questionnaire not satisfaction, but how demand management is implemented is rated through questions like: are agreements routed via demand management, is there any supervision, are the SLAs SMART, are requests sufficiently detailed with respect to functionality and quality. So most of it is in place, but overall satisfaction is low.

4.2. Boosting controllable weak factors

Six of the nine regression coefficients in the logistic model underlying the success-chance questionnaire did not appear significantly different from zero (see Table 1). However, for three of the non-significant determinants a weak (one-sided) positive effect was measured.² This indicates some non-negligible power to discriminate between cases of success and cases of failure in our observational study.

The first factor is managing the business case for outsourcers (that is to say not for suppliers), which is often assumed to be vital for success. Although it was not found to be vital for success, it was not found to be irrelevant either. The second factor is the transfer of staff, and the third factor is the transfer of assets. As for the latter two factors, we assume that it will help to handle them with care, although they are but secondary issues. It is not unreasonable to assume that these factors positively influence the success-chances: high scores are better than low scores. However,

we just cannot tell how much better, as they are not significant in the logistic formula we illustrated earlier. In the Excel (Delen et al., 2017) we implemented a questionnaire that scores those three factors.

Practical guidelines. Besides the success-perception questionnaire, we advise that you also complete the questionnaire that estimates the chance of success of the deal. It provides an insight into how the ITO-deal scores on the three significant factors. According to our data, the three weakly positive factors do not make or break a deal, but our advice is to improve on these as well, particularly in the case of low scores, so as to further improve on a successful deal. This should be done for example when the other three much stronger factors have significantly been improved upon. When a repeated success-perception questionnaire and success-chance questionnaire show sufficient improvements on the main factors, then it is time to also improve on low scoring weakly positive ones.

Do not forget to pay attention to answers in the low regions, for they are strong early-warning signals. Our advice is to delve deeper and search for root causes. We shall give two examples.

Example 1. Suppose the client did not select the most appropriate type of contract. Then it is highly likely that the client or supplier or both will score in the low range on the question:

- The SLA has been adjusted appropriately to the changing business needs.

It may well be that agreements once made cannot be updated when necessary because the wrong type of contract has been concluded. The contract has to be renegotiated to prevent a further escalation of problems in the future.

Example 2. Let us assume that the supplier delivers a highly specialized service. This situation is often characterized as a lock-in situation: the outsourcer is locked-in by the supplier due to high asset specificity. Does this imply that both outsourcers and service providers will answer the questions of the success perception questionnaire positively? Suppliers because they are content with their dominant position and outsourcers because they have no other choice than to continue their collaboration with service suppliers? We do not think this will be the case. It is important to both outsourcers and suppliers to obtain a candid opinion on their perceptions of the deal. If the outsourcer is not happy, this is a strong signal for the supplier to change his politics. No matter what the specialized service is, outsourcers can always find an alternative. Backsourcing may be an option or else swapping suppliers in spite of the fact that this may be costly, risky and/or time consuming.

In our experience backsourcing is often the chosen option. For example, ING, the large Dutch-based bank, outsourced all IT to India-based IT-service provider NIIT. ING bought many of their shares, yet decided to backsource their IT when the results turned out to be disappointing.

In their turn suppliers do not have to be satisfied with an ITO-deal in spite of their dominant position. Or take KPN, for example, a major Dutch landline and mobile telecommunications company, which, together with Motorola, for two decades now has delivered the highly specialized service which maintains C2000: a digital communication system for emergency services used by the police, fire brigade, ambulance services, the Ministry of Defence and others. At the time of writing this article the renewal of the new C2000-network was delayed so there was need for prolonged support. Earlier still, KPN/Motorola had challenged the new

² No two-sided effects, but for the one-sided Mann-Whitney tests p -values between 0.07 and 0.08 were found, which is significant at the 0.1 significance level.

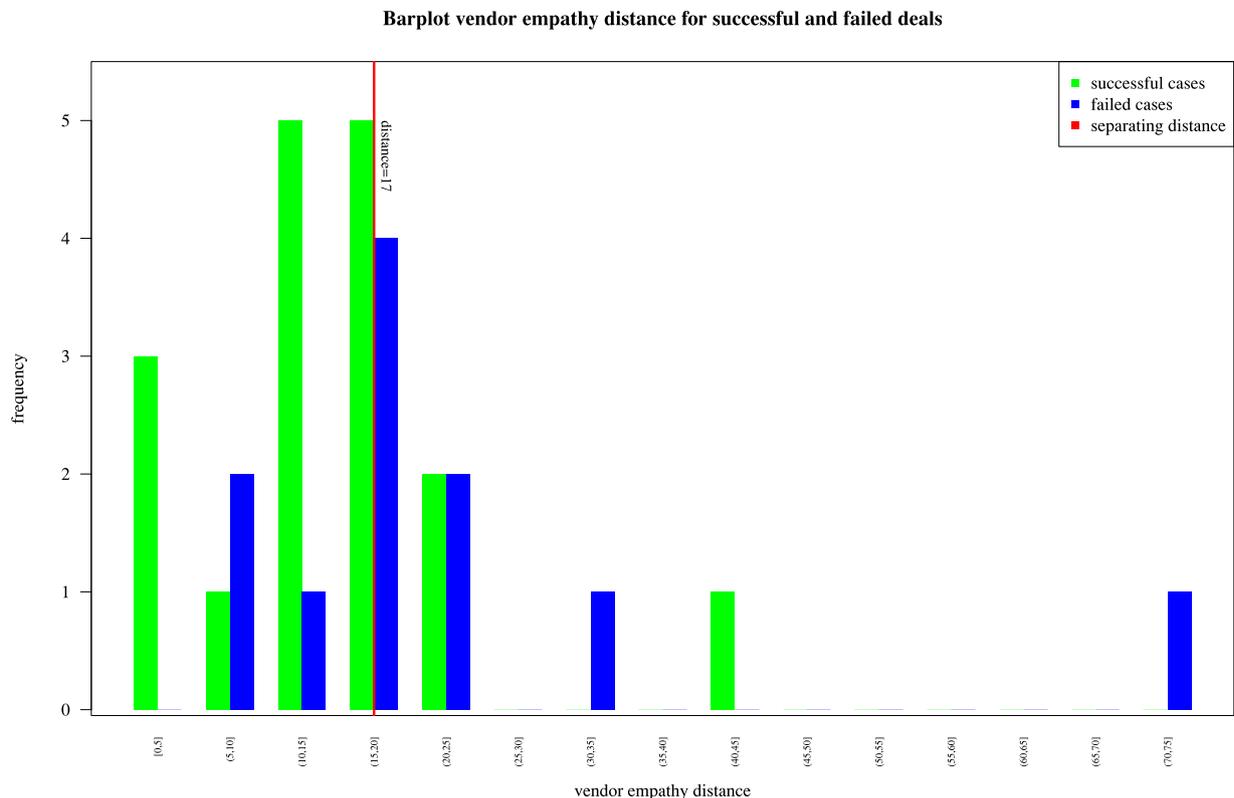


Fig. 4. Barplot of vendor empathy distance for successful and failed deals.

provider in court,³ trying to keep their position as a new supplier. KPN/Motorola, however, is neither the new supplier nor the party for the last year of prolonged support, in spite of the fact that a locked-in situation with respect to network, hardware and software is a certainty. Nevertheless, another consortium consisting of Nokia, Koning & Hartman and Kolibri Systems was contracted for the last year of support.⁴ So a way out of the lock-in was found.

It is very easy to see that suppliers are not (or no longer) making the profit they had expected to make. In such cases the success-perception questionnaire will presumably be negative. In all cases the contract has to be renegotiated to prevent further escalation of dissatisfaction in the future. Renegotiation can also lead to an ITO-deal with other parties, even if there is a general perception that there is a lock-in.

5. Rigid factors

We will now turn our attention to the significant rigid or hardly changeable factors (see Table 1). For these factors we found two positives: namely, long-term (strategic) supplier motives and vendor empathy, i.e., a vendor that correctly images the client's point of view. Two other factors are negative: non-strategic supplier motives and, contrary to common belief, hiring outsourcing intermediaries. Measuring the use of intermediaries is trivial, but how to measure vendor empathy or supplier motives is a very different matter.

Vendor empathy. First of all, it is simply a matter of common sense. If the simplest things are difficult or nonintuitive between clients and suppliers, then this can be considered as a strong signal

of low empathy capability of the service provider. If possible, obtain an intuition of the vendor's empathy *before* closing a deal, e.g., by paying a working visit to another company that struck a deal with this vendor. We investigated ongoing deals and measured vendor empathy by asking vendors to guess supplier answers on three statements: the present agreements meet the expectations of their client's management, their middle management and their end-users. Answers were pairwise compared and this showed a non-negligible dependence.⁵ If we restrict to failed deals, the dependency is roughly gone. For successful deals the first two comparisons show very significant dependencies, but the third shows none.⁶ Apparently, suppliers find it inherently difficult to estimate the client's end-user expectations. For practical purposes we defined a straightforward vendor empathy distance, which measures the average absolute deviation from the vendor's estimates of the client's expectations.⁷ It measures how empathic a vendor is by means of a simple questionnaire and ditto calculation in our Excel Delen et al. (2017). In our observational study the mean vendor empathy distance is 14.6 for successful deals and 23.1 for failed ones. As the distributions are overlapping, we chose 17 as limit value in our Excel. In Fig. 4 we plot both distributions via a barplot. Successful deals are colored green and failures are blue. The red line is at vendor empathy distance 17 switching between empathic or not. The metric is indicative: there are three blue failed deals to the left of the red line, and three green cases of success to the right. In case of doubt our advice is to have a statistician carry

⁵ Mann-Whitney tests showed p -values between 0.07–0.09.

⁶ Mann-Whitney tests show very low two-sided p -values for the first two questions and a high one for the third: $p = 0.0048, 0.0477, 0.2117$, and in the one-sided case similar results: $p = 0.0024, 0.0239, 0.1059$.

⁷ This simple distance is significant at 0.08 (Delen et al., 2016). We can improve on that using the Median Absolute Deviation as implemented in the statistical package R: this distance is significant at 0.05, but it's nonintuitive and difficult to explain, e.g. since a magic constant 1.4826 is involved.

³ <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHDHA:2015:2609>.

⁴ <https://zoek.officielebekendmakingen.nl/kst-25124-90.html>.

Table 5
Descriptive statistics of ranks per vendor motive (qu. = quartile, #obs = number of observations).

Motive	Type	Min	1st qu.	Median	3rd qu.	Max	#obs
Access to the market	S	1	1	2	2	6	49
Visibility as supplier	S	1	2	2	3	6	48
Short term turnover	NS	1	2	3	5	8	32
Long term turnover	NS	1	1	2	2	6	59
Attractive profit margin	NS	1	2	3	4	7	37
Knowledge acquisition	NS	1	2	3	5	8	35
Synergy with existing activities	S	1	2	3	4	8	44
Acquisition of specialised staff	NS	1	3	4	8	9	20
Strengthening market position	S	1	1	1	1	1	2
External pressure	S	1	1	1	1	1	5
Reciprocal business opportunity	S	1	1	1	2.5	4	3
Professionalization of services	S	1	1	3	5	5	2
Growth	S	2	2	2	2	2	2

out multiple measurements and the appropriate statistical tests for more assurance (Delen et al., 2016).

Vendor motives. We distinguish between strategic (S) and non-strategic (NS) vendor motives to close the deal. Strategic motives are intended to improve or maintain the strategically important position of the vendor. For instance, by increasing brand awareness or by increasing market share. Another objective may be to increase visibility of the company as an expert in a particular niche of the ITO field. In all cases the purpose is to ensure turnover in the long-term. For this reason we designated motives that imply long-term turnover, such as reciprocal business relationships, access to the market and visibility as supplier, as strategic. Note that aiming at long-term turnover with a particular client is classified as non-strategic (NS). As *dominant* motive it is not necessarily beneficial to the customer. For example, suppliers dominantly motivated by long-term turnover may not have any interest in finalizing a project.

It is usually hard to monetize the benefits of a strategic motive. Mostly the benefits are indirect or inferred and will pay off only in the long run. In the case of non-strategic motives the objective is usually short-term profit seeking. The payback period of an investment should be as short as possible. In contrast to the benefits of strategic motives, short-term motives can usually be measured much easier.

If the vendor is money-driven this should be considered as a strong signal of non-strategic motivations. We measured the dominant motive by asking the vendor to rank the motives as listed in Table 5. Some motives concern (in)direct benefits, and others inferred benefits.

We developed a dominance metric measuring the main vendor motive to decide whether the vendor is of strategic intent or not. We asked between 1–4 persons from the supplier-side to rank the predefined motives in Table 5. We counted strategic motives and non-strategic ones. The first-ranked motives discriminate unless there is a tie. If that is the case, we take second-ranked motives to decide, unless there's another tie, in which case we resort to third-ranked motives. In our experience, more than 3 ranks are not necessary (but the Excel can be trivially extended to more ranks). In the rare case of a tie at rank 3 our advice is to ask a few additional persons to rank their motives. There is a strong dependence between the dominant motive and the outcome: strategic vendors are more often successful.⁸ The dominant vendor motive metric is implemented in the accompanying Excel (Delen et al., 2017). The possible answers are 1, 2, 3, 4, ... ranking the predefined motives.

⁸ Fisher's exact test shows a very low p -value on the null hypothesis that the dominant vendor motive is independent from the deal-outcome, indicating that there is a strong dependence ($p = 0.008$).

Not every motive needs to be ranked, and equally important motives can be equally ranked.

Intermediaries. Hiring sourcing consultants (intermediaries) turned out to be counterproductive: it lowered the chance of success. We ruled out the potential causality of hiring intermediaries *on the grounds of inexperience* and failing on the grounds of inexperience (Delen et al., 2016). Our intermediaries were mainly local, so that might restrict our results to the Netherlands. But this is unlikely since US-based research had similar findings (Bapna et al., 2010):

Contrary to the received view of the positive role of intermediaries, our analysis using 700 large IT outsourcing contracts from 1989–2009 suggests that the likelihood of IT outsourcing contract failures is higher in the presence of intermediaries, as intermediaries make vendor selection overly competitive. Our conjecture is that in the process of getting the best deal for clients, increase in competitive intensity causes the winning vendor to suffer from winner's curse or that high-quality vendors are dissuaded from participating in the bidding process.

They spotted two of our results: intermediaries and non-strategic supplier motives are counterproductive. Evidence for their conjectured causality is not present in our data: intermediaries do neither attract vendors with short-term motives nor shun vendors with long-term strategic motives (high-quality vendors). In our representative field study the presence or absence of intermediaries and vendor motives are not dependent.⁹

So the counterproductivity of intermediaries is not a typical Dutch phenomenon. This, too, is an indication that our results are more generally applicable than in the Dutch market alone.

Practical guidelines. Before closing a deal, please take note of the significant rigid factors. It is better to favor suppliers with dominant long-term strategic motives, and to favor empathic suppliers over less understanding organisations. There may be good reasons for bringing in intermediaries, but it is better not to outsource outsourcing. Of course, suppliers understand very well that short-term motives and lack of empathy are not helpful properties, but the suppliers should appreciate that those properties have a significant negative impact.

Although the dominant vendor motive is not easily flipped, sometimes strategic motives do turn into less strategic ones during a deal. For instance, suppose a package was offered (e.g., at a fixed

⁹ We have 15 cases with intermediaries of which 6 suppliers with short-term and 9 with long-term strategic motives, and 15 cases without intermediaries score 7 and 8 respectively. The expected frequencies in case of independence ($13 \cdot 15/30 = 6.5$, $17 \cdot 15/30 = 8.5$) are almost identical with the observed frequencies, making dependence unlikely: both Fisher's exact test and the χ^2 test give $p = 1$.

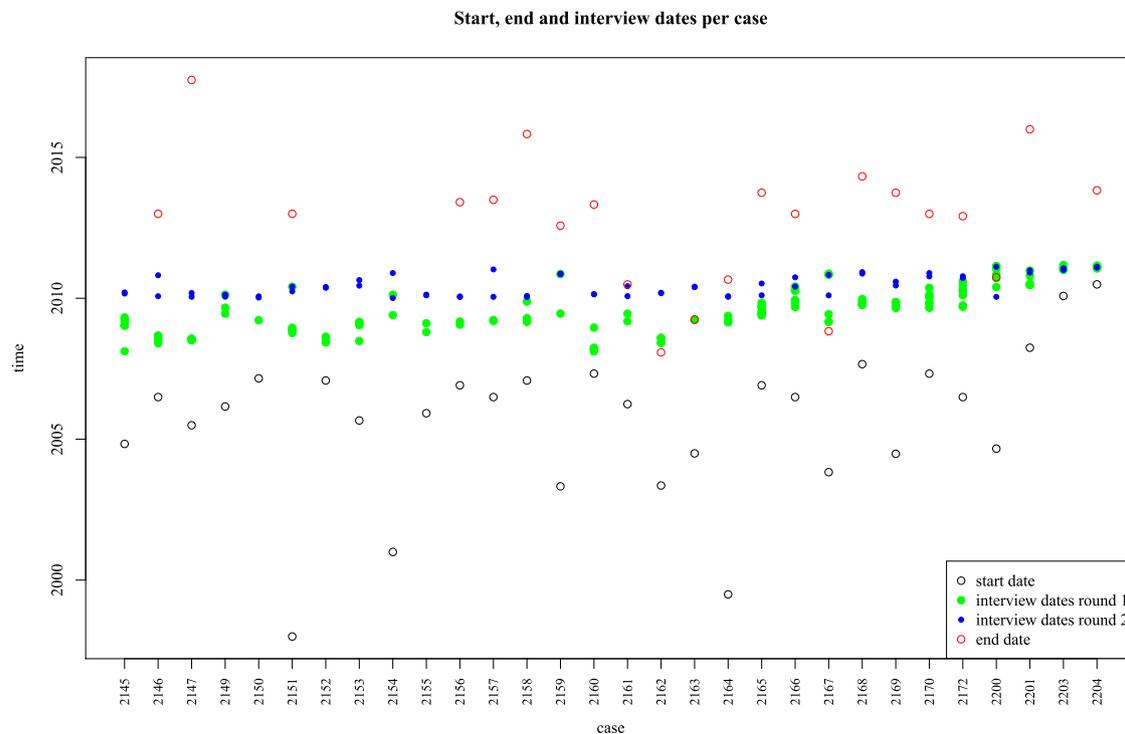


Fig. 5. Start, end and interview dates per case.

price) which then, on a sliding scale, would change into full-blown custom development. The package buyer would turn into a launching customer. The supplier's cost-cutting pressure would start to overshadow their odds-improving strategic motives. In such cases, an option would be to renegotiate the deal, e.g., by paying more and becoming an investor so as to reap additional benefits when selling the (now jointly developed) product to others. Or by negotiating enhancement and maintenance free-of-charge because more than the fixed price is paid for the development. Better acquisition practice might have prevented this, but that is the lesser of two evils.

Measuring vendor empathy with our questionnaire can only be done once the deal is closed, for only after closing a deal does the vendor get to know the customer and can he start estimating the client's expectations. So, initial intuitions concerning empathy can be confirmed or nuanced by our questionnaire. In the case of a negative outcome our advice is to make the topic discussable to enable improvement, for instance by making sure that the vendor is made explicitly aware of the client's expectations.

6. Timing the interviews

Our interviews were carried out between 2008–2011, whereas the contract-span was from 1998–2016, and we could not in all cases carry out the interviews between start-dates and end-dates of deals. Fig. 5 provides proof of this with an impression of start-dates, end-dates and interview-dates on a per case basis with data taken from our observational study.¹⁰ The interview-span did not limit our research. Three cases ended before all interviews were finalized, and one ended in failure just after the second interview round. The others all were in progress until after the interviews. We know the actual outcome of 25 cases at the time of writing this article, which in all cases equals to the predicted outcome of

the success-perception questionnaire. A few were, of course, post-hoc predicted, and the rest ex-ante: some very early and some very late in the project due to the interview time-span.

We had two rounds of interviews. On median there were about 200 calendar days in between, but with a large spread. Due to the time-span a few cases had first and second rounds within a short timeframe, but the rest had a longer period as can be seen from the plot. We detected a significant difference¹¹ between first and second round answers: they went from optimistic to realistic. Figs. 6 and 7 show that both outsourcers and suppliers become less positive in their answers. Apart from the numbers, the qualitative answers that interviewers collected during these rounds indicated this as well (answers translated from Dutch):

- The municipality continues to struggle with governance and demand management, as profit margins are under pressure.
- The customer becomes more and more demanding.
- There are problems with the transition.
- The deal was still young, but I foresaw large problems which emerged indeed.
- Both the relation with the customer and the odds to prolong the contract have worsened.

That is why we used the answers from the second interview round for building the models in the questionnaires, and that is also why we advise you to reiterate the measurements made with our questionnaires in order to determine changes in perception when the parties involved have experienced a somewhat longer collaboration.

Practical guidelines. Our advice is to plan a first round of interviews not too long after the honeymoon period of the project when most is in place and the work has really started. A second

¹⁰ Not every interview-date was captured, not all start-dates and end-dates were exact days, and eight cases had no end-date at baseline because they were indeterminate contracts.

¹¹ Two one-sided Mann Whitney tests give p -values of approximately zero indicating that the null hypotheses that both suppliers and clients do answer more positively in the second round than the first do not hold. So they both are more positive in the first round than in the second ($p = 9.5 \cdot 10^{-10}, 7.7 \cdot 10^{-5}$).

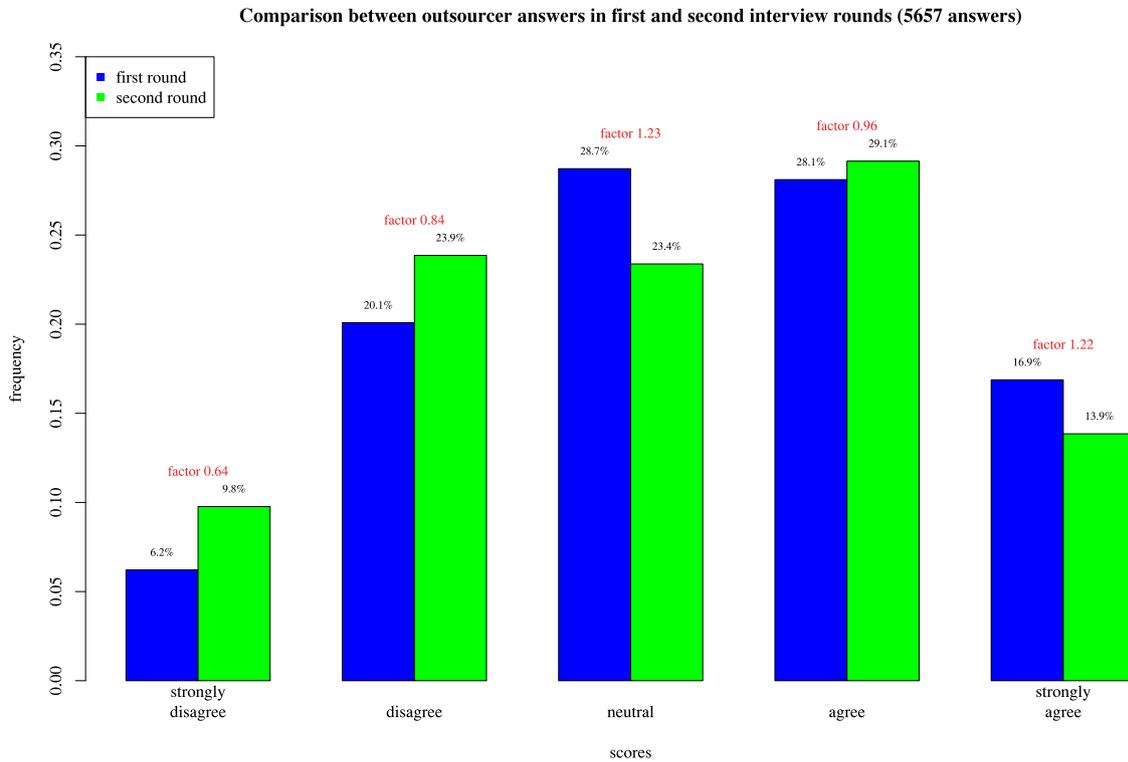


Fig. 6. Comparison between outsourcer answers in first and second interview rounds.

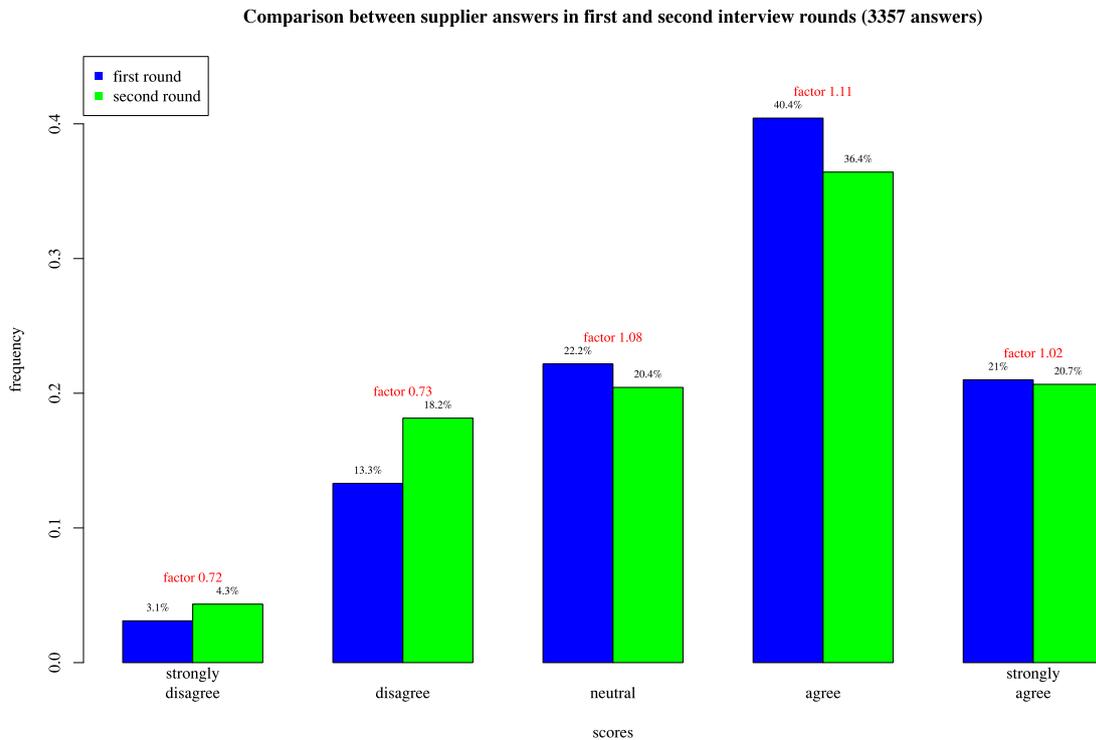


Fig. 7. Comparison between supplier answers in first and second interview rounds.

round could be carried out half a year or a year later, depending on the length of the deal.

7. Conclusions

In an earlier observational study, after a long-term scrutiny of a representative sample of Dutch IT-outsourcing deals, we found

evidence of significant factors that determine success as well as factors that were deemed significant but actually were not. In this article we laid the foundations for various tools to measure and influence the outcome of ITO deals.

Taking the results of the observational study as our starting point, we developed an early-warning system in the form of a success-perception questionnaire as described in this article. Based

on intuitive statements it provides an insight into the perception of both clients and suppliers and has predictive power as regards the actual outcome. Remarkably, we found that in cases of failure clients generally foresee a much darker future than suppliers. In cases of failure their success-perception score is generally much lower. An other remarkable fact we found is that both outsourcers and suppliers become more realistic when time progresses. We detected a significant difference between answers in first and second round interviews: answers were less positive in the second round.

In addition, we developed a success-chance questionnaire that estimates the odds for a deal and provides indications what to emphasize when improving success determinants. Yet another questionnaire provides an additional qualitative insight into weakly significant factors that can be improved upon particularly in the case of low scores on strong determinants. Our advice is to first improve on strong determinants and measure the effects gained on these factors. We also provided practical guidelines regarding when and how to use and interpret the outcomes of the questionnaires.

We also found hardly influenceable (rigid) factors that can make or break a deal. Hiring intermediaries and non-strategic motives of suppliers turned out to be counterproductive rigid factors. High vendor empathy and long-term strategic supplier motives are rigid factors that aid in successful deals. Before closing a deal it is worthwhile to complete our dominant-vendor-motive questionnaire, and to form a first opinion on vendor empathy (e.g., by paying a working visit to a peer already using the vendor's services). When the deal is already in progress, vendor empathy can be measured with the questionnaire we designed for that purpose.

All tools are implemented in a ready-to-use Excel spreadsheet hiding the mathematics (Delen et al. 2017). To avoid experimenting for favorable outcomes and too much variation in interpretation we advise the use of (neutral) interviewers to complete the forms. The Excel supports ten form tabs: so up to ten persons from the client-side as well as ten persons from the supplier-side can be interviewed for each questionnaire, which is more than adequate. Each tab contains all five questionnaires. A summary tab shows the aggregated results of all completed individual forms. We expect the Excel questionnaires to be useful and we hope that our research will help to increase the present state of IT-outsourcing success in general, and in the Netherlands in particular, which has a substantial failure rate of 40%.

Acknowledgements

We thank the reviewers for their valuable comments. This research was partially supported 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*.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jss.2019.06.074](https://doi.org/10.1016/j.jss.2019.06.074).

References

- Bapna, R., Gupta, A., Ray, G., Singh, S., 2010. Analyzing IT outsourcing contract outcomes: the role of intermediaries. In: 11th International Conference on Web Information System Engineering.
- Barthélemy, J., 2003. The seven deadly sins of outsourcing. *Acad. Manage. Exec.* 17 (2), 87–98.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. L. Erlbaum Associates.
- Delen, G., Peters, R., Verhoef, C., van Vlijmen, S., 2016. Lessons from Dutch IT-outsourcing success and failure. *Sci. Comput. Program.* 130 (C), 37–68. Available via www.cs.vu.nl/~x/pred/pred.pdf
- Delen, G., Peters, R., Verhoef, C., van Vlijmen, S., 2017. Outcome-prediction and odds-improvement for IT-outsourcing – accompanying Excel spreadsheet. Available via www.cs.vu.nl/~x/sf/sf.xlsx.
- Eveleens, J., Verhoef, C., 2009. Quantifying IT forecast quality. *Sci. Comput. Program.* 74 (11–12), 934–988.
- Eveleens, J., Verhoef, C., 2010. The rise and fall of the chaos report figures. *IEEE Softw.* 27 (1), 30–36.
- van Genuchten, M., 1991. Why is software late? An empirical study of reasons for delay in software development. *IEEE Trans. Software Eng.* 17 (6), 582–590.
- Gliem, J. A., Gliem, R. R., 2003. 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.
- Kelsey, F.O., 1965. Problems raised for the FDA by the occurrence of thalidomide embryopathy in Germany, 1960–1961. *Am. J. Public Health Nations Health* 55 (5), 703–707.
- Nuijten, A., Keil, M., Commandeur, H., 2016. Collaborative partner or opponent: how the messenger influences the deaf effect in IT projects. *Eur. J. Inf. Syst.* 25 (6), 534–552.
- Robey, D., Keil, M., 2001. Blowing the whistle on troubled software projects. *Commun. ACM* 44 (4), 87–93.