

DETC2012-71362

THE EFFECTIVENESS OF DESIGN AND INNOVATION CONSULTING

William J. Palm IV

Department of Engineering
Roger Williams University
Bristol, Rhode Island 02809
wpalm@rwu.edu

Daniel E. Whitney

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
dwhitney@mit.edu

ABSTRACT

Although the risks and rewards of outsourcing product design have been argued extensively in the literature, little hard data on project outcomes exist to inform the discussion, and even these are methodologically suspect. To address this gap, this paper uses novel random sampling techniques to locate projects and measure the distribution of project outcomes in one particular type of design outsourcing, domestic design consulting. The results suggest that design consulting outcomes are generally good but vary significantly between projects and consultancies. Overall rates of product commercialization and market success compare favorably to results previously reported for in-house product design, and client satisfaction levels are comparable to those of other service industries.

1. INTRODUCTION

As product design outsourcing becomes increasingly common, a flurry of articles has appeared offering managerial guidance. Many provide well-reasoned analyses of the pros and cons of outsourcing but little empirical data upon which to test their deductions [1-5]. Some ground their recommendations using case studies [6-10] or expert opinions [11-13]. Only a handful report quantitative data on project outcomes [14-19], typically under foreboding headlines such as “Design Outsourcing: Beware of Pitfalls” [17], and “Prevent Disasters in Design Outsourcing” [14]. The former found that a third of outsourcing managers at large firms view outsourcing as a liability, while the latter suggests that two-thirds of outsourced design projects at *Fortune* 1,000 firms struggle or fail.

These results, as well as the skeptical tenor of the entire genre [20, 21], seem to suggest that firms should limit design outsourcing, and yet other studies find that outsourcing continues to grow [19, 22, 23]. One possible explanation of the

apparent paradox is that the high failure rates reported for design outsourcing have been exaggerated.

Indeed, close inspection of these reports reveals several methodological limitations that may compromise their findings. Some studies did not use a random sample [15], and those that did achieved very low [16] or unstated [18, 19, 24, 25] response rates. All the studies collected data from only the client’s perspective, often from mid-level managers. Such respondents might be excessively critical of outsourcing if they fear it threatens their jobs. The anonymous format of many studies may facilitate such criticism by shielding the respondents from rebuttal. Another potential source of bias is that some surveys ask respondents for cumulative impressions of outsourcing rather than appraisals of specific projects [17, 19], opening the door to hearsay or parroting of conventional wisdom. In addition, few studies use standard measurement items that have been tested for validity and reliability. Lastly, none of the reports provide baseline information about the success or failure of non-outsourced projects. Studies of traditional in-house product development place the failure rate at about 41% [26], which makes the two-thirds figure for outsourcing, if true, less severe than it would first appear.

Because of these methodological issues, very little high-quality data on outsourcing project outcomes exist to inform academic research or practical decision-making. To fill this gap, this paper surveys project outcomes in one particular type of outsourcing – design and innovation consulting – using rigorous methods adapted from the management and social sciences to avoid the pitfalls that have befallen past research. A random, representative sample was drawn from a well-defined population and data collected from multiple respondents, both consultant and client, associated with each project. The survey instruments were developed based on in-depth interviews with expert practitioners, used standard measurement items from the product development and customer satisfaction literature, and

were pre-tested extensively prior to use. To corroborate the results, two additional near-random samples were drawn using different techniques. Collectively, the three sets of results provide a more complete picture of outsourcing outcomes than has been presented to date.

2. LITERATURE REVIEW

Only a handful of studies have reported quantitative data on design outsourcing outcomes. Roy and Potter evaluated the impact of a 1980s British program providing subsidies for small manufacturers to hire design consultants [18, 25]. They found that 65% of the projects were commercialized, with 89% of these generating a positive return on investment for the client firm [18]. 27% of the firms experienced problems with their consultants. The occurrence of problems did not impact commercialization rates but slightly fewer problem-ridden projects made a profit, and those that did had a 20% longer payback period [25]. Many of the problems stemmed from poor design management by the client firms, 70% of whom were using designers for the first time, and 30% only because subsidies were provided [25]. Only 4% of the firms in the sample employed more than 500 people, and none employed more than 1000. The study's focus on relatively novice, small client firms may compromise the generalizability of the findings to larger, more experienced firms. In addition, the outcome measures used may not be the most appropriate. Quantitative financial data could be obtained for only 91 of the 221 projects in the sample [18]. The study focused heavily on the occurrence of problems, but the results of the present study indicate that problems are ubiquitous in consulting and do not factor strongly in clients' appraisals of project success.

The only other academic data on outsourcing outcomes come from an ongoing study by Anderson, Parker, and Davis-Blake [27]. Unfortunately, the results of this work have not yet been fully documented. A brief blurb in the *Harvard Business Review* states that nearly two-thirds of the projects "struggled or failed," but did not provide the data supporting that claim [14]. A 2008 working paper reports mean values and ranges for project managers' assessments of 55 outsourced product or process development projects [15]. The managers were somewhat satisfied with their working relationships with the suppliers (mean of 3.8 on a 5-point scale) but less so with the performance, quality, cost, and time-to-market impact of the project deliverables (3.3, 3.2, 3.3, and 3.0, respectively). Project performance varied widely, with each measure using the full scale range, but the measures were not strongly correlated – a project might perform well on one measure but poorly on another. Methodological drawbacks to this study include the use of a convenience sample and the pooling together of projects from dramatically different industries such as aerospace and consumer products.

A third source of quantitative data on outsourcing outcomes is the surveys conducted by trade journals and professional associations. These have the benefit of using quasi-random samples but the response rates are so low as to

cast doubts on their representativeness. For example, a study conducted by *EE Times* and *Electronics Supply & Manufacturing* invited 11,000 readers to participate and received 303 responses, or just 2.75% [16]. It is possible that only those with an axe to grind took the time to reply. Nonetheless, the results of these studies offer some insights. The *EE Times* study found that 65% of outsourcing managers had at some point experienced schedule delays on a project, 59% had experienced communication problems, 46% cost overruns, and 21% a project that failed to meet performance requirements [16]. Nearly one-third of those at large companies (>\$500 million annual revenue) viewed outsourcing as a "net liability," compared to just 22% of those at mid-size firms and 13% of those at small companies (< \$10M) [17]. The author presents the one-third figure as a cause for concern, but does not state whether the remaining two-thirds viewed outsourcing as a net asset, were indifferent, or skipped the question altogether. An ASME survey of engineering executives found that 44% agreed that "outsourcing generally improves my company's overall financial performance," but only 15% agreed that "outsourcing to overseas/non-U.S. firms offers significant advantages" [19].

In sum, a small amount of data has been reported in the literature, but its quality is compromised by the use of non-random, non-representative, and/or incongruous samples, limited respondent perspectives, non-standard and/or inappropriate outcome measures, as well as low response rates and incomplete reporting of results.

3. METHODS

3.1. Questionnaire Development

In-depth interviews were conducted with 36 experienced design consultants and clients to identify seven key dimensions of project success as well as a process model of a prototypical consulting project that formed the organizing structure for the questionnaire [28]. Measures for the various dimensions were identified from the interviews as well as the product development and customer satisfaction literature. Wherever possible, standard survey items were used to allow comparisons to past results.

3.2. Pilot Testing

Following the interviews, the draft survey instrument was used to collect data on thirty projects from one consultancy. The project manager or a project engineer from each project were identified by the consultancy's study coordinator and asked to complete the questionnaire, mimicking the method that would be used in the full-scale survey. As each questionnaire was returned, the respondent was de-briefed by the investigator via phone or in-person interview to identify misunderstandings and other problems with the instrument. Relatively few changes were required, so the resulting data are generally comparable to those gathered in the full-scale survey. However, data were

collected only from consultant respondents, as the consultancy was not comfortable extending the pilot survey to its clients.

3.3. Benchmarking and Patent Surveys

3.3.1. Sample selection. The study population for the survey phase consisted of all projects completed within the last four years by U.S. consultancies specializing in consumer, medical, and industrial product development [29]. Two separate samples were used to maximize coverage of the population. The first sample was implemented as a client satisfaction benchmarking study. Nineteen consultancies were randomly selected from the consultancy population and invited to participate. Five agreed to take part. A study coordinator at each consultancy compiled a list of all projects completed within the prior two years, from which a total of 126 projects (~25%) were randomly drawn for study. The coordinators identified one to four client and one to three consultant participants of each project to solicit feedback from. The coordinators then mailed the questionnaires to 184 client and 104 consultant participants. 82 client and 74 consultant questionnaires were returned directly to the investigator, yielding response rates of 44.6 and 71.1%, respectively. These groups are referred to as the “Benchmarking Client” and “Benchmarking Consultant” subsamples in the results. A total of 62 projects were assessed by at least one client respondent, 52 by at least one consultant, and 46 by at least one of each type.

The second sample emerged from an attempt to counter the potential biases in the benchmarking study. The worst of these would be if the participating consultancies attempted to skew their results by excluding poor projects from the sampling frame or encouraging respondents to provide positive feedback. Several defenses against such behavior were deployed. First, mutual non-disclosure agreements were signed to prevent either the consultancy or the investigator from revealing consultancy-specific results publicly. Second, the consultancies were reminded that use of the sponsoring university’s name without approval would constitute a trademark violation. Third, they were encouraged to view the study as a learning opportunity, with the lessons only as valuable as the representativeness of the results. Finally, they were allowed to add key clients that had not been randomly selected onto the respondent list. Feedback was gathered from these clients and reported back to the consultancy but not included in these study results.

Nonetheless, at least two other biases remain that could positively skew the results of the benchmarking sample. First, the consultancies that declined to participate may have done so because they anticipated poor results. Second, the client respondents may have been charitable in their responses because they knew that the consultancies would see them. To minimize such bias the respondents were instructed that all responses would be anonymized and aggregated before being reported to the consultancies.

These last two biases could be largely eliminated by removing the consultancies from the sampling process and collecting data directly from the clients. The challenge is

locating them. Design consulting, while growing, is still a relatively small industry, so few practicing product developers have direct experience with a recent consulting project. In addition, many client firms prefer not to publicize the fact that they do not design their own products. To randomly locate consulting clients would require a very large sample, with pre-screening to separate the customers from the non-customers.

Instead, consulting clients were identified directly from public traces left behind by their projects. Such approaches are commonly used by researchers studying rare or sensitive subjects [30]. For the present purpose, the best trace evidence was government records, specifically, patent applications. The pilot data suggested that nearly 45% of projects yielded a U.S. patent or patent application, which identifies both client and consultant inventors and usually the client firm (the assignee). Better still, statistical analysis of both the pilot and benchmarking samples showed that projects which generated a patent record did not significantly differ on any other dimension from projects that did not generate a patent record. Thus, patent records can provide a reasonably representative subset of all consulting projects.

To generate the patent sample, five consultancies were randomly selected from the fourteen that declined to participate in the benchmarking study, so as to counter non-response bias. For each consultancy, one or more seed employees were identified from the consultancy’s own web site or press releases. These names were used to search U.S. patent records for projects the employee had worked on. Each patent provided names of client inventors as well as additional consultants, who were used in subsequent searches. Proceeding through the patent network in this fashion produced a total of 230 projects having patent applications filed within the past four years. A total of 784 client inventors were identified, of whom the whereabouts of 310 could be confirmed from public records. From these, 262 were randomly selected and mailed the questionnaire, care of their present employer. Consultant inventors were not included in the patent sample out of respect for their employers’ decisions not to participate in the study. 38 completed questionnaires and 18 non-deliverables were returned after one postcard reminder, yielding an effective return rate of 15.6%.

3.3.2. Non-response analysis. Three forms of non-response analysis were performed. First, the benchmarking projects that produced a client response were contrasted to those that did not on the basis of project demographic data provided by the consultancies. Next, intra-project client response rates were calculated for each of the benchmarking projects and regressed against all other variables. Finally, all variables from the benchmarking and patent responses were regressed against response promptness (number of days from survey mailing to return postmark), under the assumption that non-respondents are more similar to tardy respondents than to prompt ones. Together, the non-response analyses suggest that the benchmarking sample is likely biased towards clients with whom the consultancies have extensive working relationships,

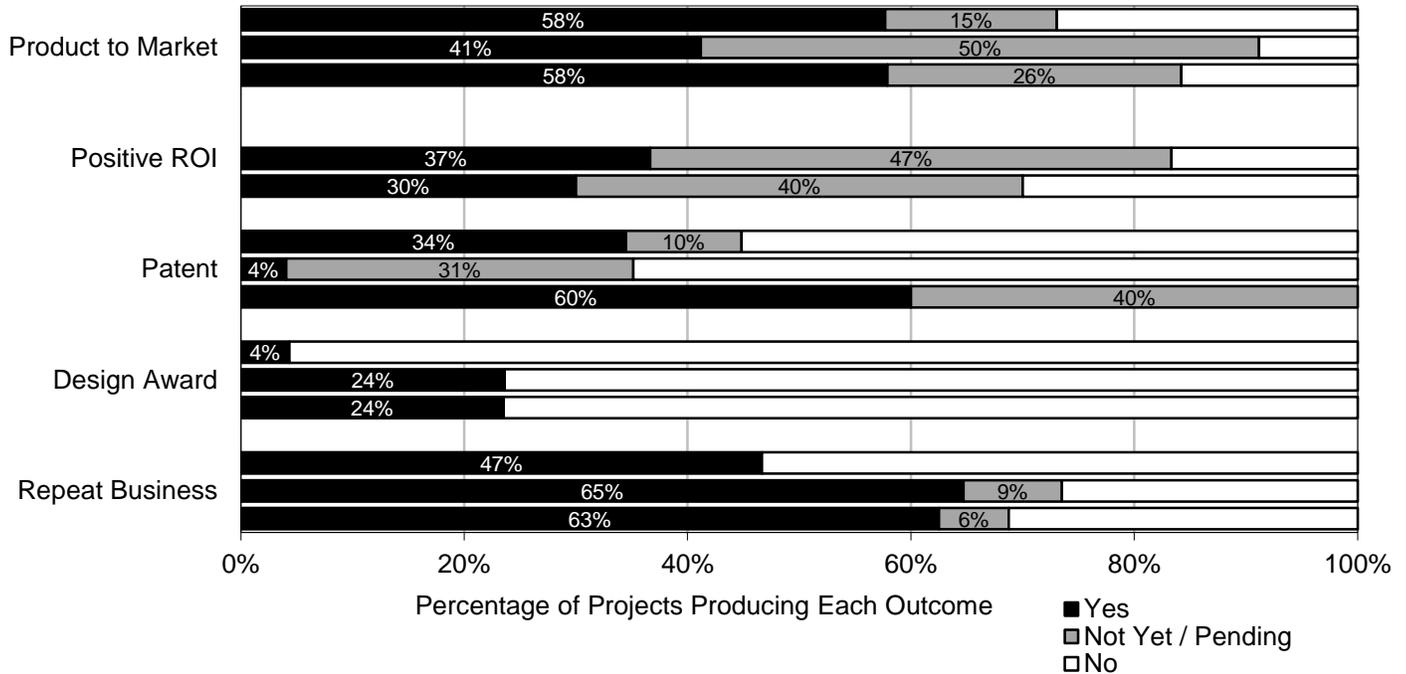


Figure 1. OBJECTIVE MEASURES OF PROJECT SUCCESS. THE TOP BAR IN EACH GROUP REPRESENTS THE PILOT SAMPLE, THE MIDDLE BAR THE BENCHMARKING SAMPLE, AND THE BOTTOM BAR THE PATENT SAMPLE.

is likely biased towards longer projects, may be biased towards client respondents having more experience with consulting, and may be biased towards more successful projects. The one form of non-response analysis that could be performed on the patent sample did not find evidence of bias.

4. RESULTS

4.1. Business Impact Measures

4.1.1. Product commercialization. Figure 1 presents the commercialization rates for the three samples. In this and other figures, percentages are based on the meaningful responses only, disregarding “I don’t know” and other non-applicable responses. 41 to 58% of projects have already commercialized one or more products, with another 15 to 50% still in development at the time of the survey. The benchmarking sample contained fairly recent projects (surveyed zero to two years post-project, versus one to four years for the pilot sample) so it is not surprising that many are still in development. Including these pending projects, the three samples are significantly different (χ^2 test, $p = 0.040$), but excluding them they are not ($p = 0.33$).

To put these results in context, Roy and Potter found that 65% of the consulting projects in their study commercialized a product (projects surveyed three to six years after completion) [18]. Similarly, the Product Development and Management Association (PDMA) Best Practices Study estimates that 70% of all development projects (outsourced or not) yield a product

[26]. To achieve similar commercialization rates, approximately half of the pending projects in the present study would have to yield a product, which seems plausible.

4.1.2. Commercial success. For the products that did go to market, client respondents were asked how well the product performed relative to their original objectives. A 9-point scale was used to provide greater resolution than traditional binary measures [26, 31-33]. Commercial success varied widely, as shown in Fig. 2, and was significantly higher for the benchmarking sample than the patent sample (Wilcoxon rank-sum test, $p = 0.019$). The pilot sample was not assessed as the consultant respondents could not accurately judge the products’ performance or the clients’ expectations for them.

For context, the PDMA Best Practices Study found that 58% of all commercialized projects are deemed successful by the firm launching them [26]. Assuming that scores of 5 and up on the 9-point scale constitute success, 92% of the benchmarking and 64% of the patent sample were successful. If scores of 6 and up constitute success, 65 and 45% were successful.

4.1.3. Return on project investment. Client respondents were also asked whether the project generated a positive return on its cost. As shown in Fig. 1, 37% of benchmarking and 30% of patent sample projects had already paid for themselves by the time of the survey, while 47 and 40% had not yet done so but might in the future. For comparison, Roy and Potter found that 58% of the projects in their sample generated a positive return [18].

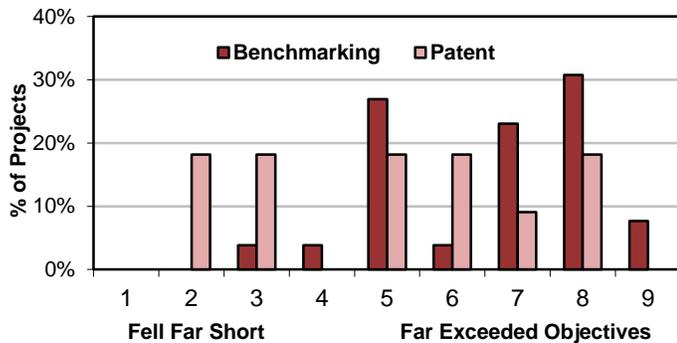


Figure 2. SUCCESS OF COMMERCIALIZED PRODUCTS IN THE MARKET. THE PILOT SAMPLE WAS NOT ASSESSED.

Beyond the percentages, a notable finding is how many respondents could not answer the question, either because their companies do not track return-on-investment (ROI) or because they did not personally have the data. ROI results are therefore not available for fully one-quarter and one-third of the benchmarking and patent projects, respectively. What's more, responses from multiple respondents on the same project often differed. Majority rule was used to produce the percentages in Fig. 1, with ties decided by the most senior respondent's answer. As a result, these data should be viewed as approximations. They are presented primarily to illustrate the challenges involved in measuring a project's financial impact. This finding confirms the experience of Roy and Potter, who could only obtain quantitative financial data for 41% of their sample and qualitative estimates for 40% more [18].

4.2 Design Quality Measures

4.2.1. Meeting of project requirements. Most projects were judged to have exceeded the requirements stated in the project contract, by both consultants and clients (Fig. 3). In contrast to the results of prior studies [16, 25], no projects were rated below 3 on the 5-point scale. This may reflect the sophistication of the contemporary domestic consulting industry. Indeed, the interviews suggested that gross dereliction of duty by the consultant is quite rare.

4.2.2. Implementation and rework of consultant's deliverable. In approximately 86% of projects the consultant's work was implemented in the client's broader development project. Non-implementation occurred primarily in exploratory projects or when the broader development project was canceled, halted, or re-scoped due to a market shift. Only in four projects (all in the benchmarking sample) was non-implementation due to the quality of the work itself.

Most projects required at least minor rework but projects in the patent sample required significantly more (Wilcoxon rank-sum, $p = 0.028$).

4.2.3. Patents and design awards. 45% of pilot and 35% of benchmarking sample projects generated a patent record

(Fig. 1). The pilot sample had more granted patents at the time of the survey because they were older projects. About one quarter of both the benchmarking and patent samples won design awards versus just 4% of the pilot sample, a significant difference (χ^2 test, $p = 0.044$). A possible explanation is that the projects in the pilot sample were more engineering-focused than those in the other samples, and design awards tend to recognize industrial design achievement. Also, the pilot consultancy tends not to invest heavily in applying for awards. Lastly, the client respondents of the benchmarking and patent samples were in a better position to observe awards than the consultant respondents of the pilot sample.

4.3. Process Efficiency Measures

4.3.1. Schedule and budget performance. Both consultant and client respondents were asked to assess when the consulting project was completed relative to the schedule specified in the contract and how much it cost relative to the specified budget. The results are shown in Fig. 3. 67 to 86% of projects were rated 3 or higher on the 5-point schedule scale, suggesting that they were completed on time or ahead of schedule. For budget performance, 73 to 92% of projects were rated 3 or higher, suggesting they came in at or under budget. The high fraction of "3" responses is not surprising given that nearly two-thirds of the projects used fixed-fee contracts.

4.3.2. Perceived project value. Respondents were asked to judge the value of the consulting project to the client firm, considering both the cost of the project and the benefits obtained from it. Because this question is somewhat subjective, the results are presented on a respondent basis rather than a project basis. An "unfamiliar with project contract" option was provided for those who did not know the project cost, which roughly 10% utilized. Although the consultants would not be expected to know the full benefits of the project to the client firm, they were asked to estimate project value anyway to allow comparison of perspectives.

Figure 3 shows the distributions. Overall, 79 to 91% of consultants and 63 to 84% of clients rated their project's value 4 or 5. Patent sample respondents rated their projects significantly lower than the others (Kruskal-Wallis test, $p = 0.0005$). Benchmarking consultants rated project value higher than benchmarking clients did, though the difference is not significant. When their ratings are compared on matched projects, the difference does become significant (Wilcoxon signed rank test, two-sided, $p = 0.002$). The pilot sample consultants rated their projects slightly lower than both the benchmarking clients and consultants, but this effect was also not significant.

4.3.3. Occurrence and resolution of problems. 36 to 70% of respondents experienced "Some Problems" on their projects and an additional 0 to 6% reported "Serious Problems." While these categories are somewhat vague, they were used to allow comparison to Roy and Potter's study, which found that 16.5%

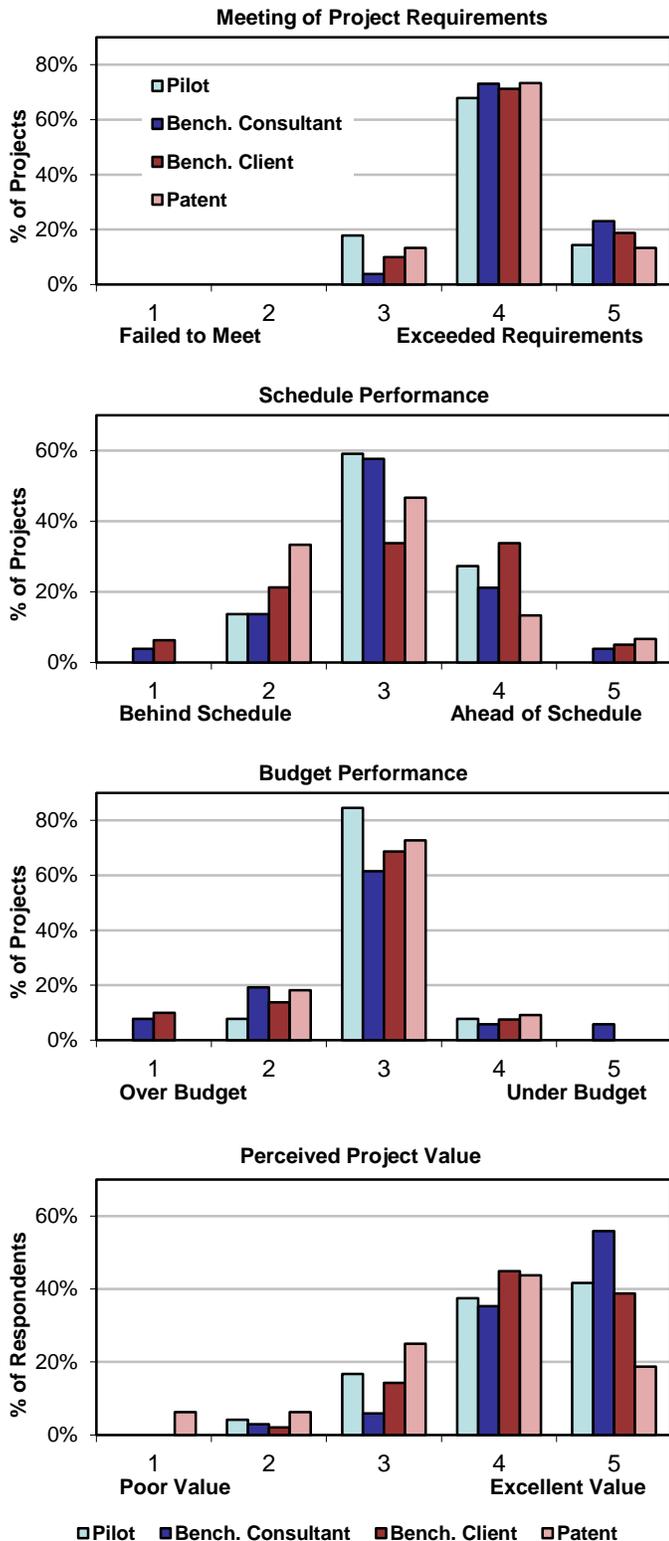


Figure 3. CONTRACTUAL PERFORMANCE MEASURES.

of projects encountered some problems and an additional 10% serious problems [25]. It is not immediately clear why the

respondents in the present study reported more problems but less severe ones. One possible explanation is that the projects in Roy and Potter's study were smaller in scale (the government subsidy only paid for 15 to 30 days of consulting service) but the clients were quite novice (over two-thirds had never used a consultant before).

Interestingly, consultant respondents in the benchmarking study reported more problems on those projects than did their clients (57% vs 42%), though the difference is not significant. Likewise, a significantly higher percentage of the patent sample reported problems than the benchmarking clients (70% vs. 42%, $p = 0.020$). This may indicate biases in the benchmarking study. About half and three-quarters of all consultant- and client-reported problems were eventually resolved.

Problems occurred in nearly all phases of the consulting process. Communication problems during development topped the list, confirming the need for study of interfirm coordination mechanisms [5]. Patent sample respondents were far more likely than others to flag problems communicating with the manufacturer, despite the fact that 43% of these products were produced by the client firm itself. The greater incidence of manufacturing coordination problems in the patent sample is likely due to these projects being further along in development.

Besides communication, the most problem-ridden phases were those occurring after the formal completion of the project, when the client firm must accept the consultant's work into its organization and refine it. These issues have received little attention in past empirical research. The issues that have been identified, such as poor specification of requirements and development problems by the consultant, persist but at lower levels than previously reported. This may reflect a maturation of the industry, or simply the fact that the present study included more experienced client firms (versus [25]) and more sophisticated design suppliers (versus, e.g., [16]).

4.4. Relationship Quality Measures

4.4.1. Quality of working relationship. Most respondents rate their working relationship with the other party well, with 80 to 90% of each group scoring it 4 or a 5 on a 5-point scale. The mean scores for the pilot, benchmarking consultant, benchmarking client, and patent groups are 4.27, 4.36, 4.56, and 4.25, each higher than the 3.8 value reported by Anderson et al [15]. What's more, no respondent in the present study assigned a score of 1, whereas that did occur in their study. Although demographic details were not provided, it appears that their study included more international collaborations, in which the greater physical and cultural distances may have adversely impacted the working relationships. Likewise, their study appears to have focused on integrated design-and-manufacturing suppliers rather than design consultants. This may have led to a more arms-length relationship than the partner-like relationships typical in the present study. The benchmarking clients rated their relationships significantly better than both the patent clients ($p = 0.010$) and the

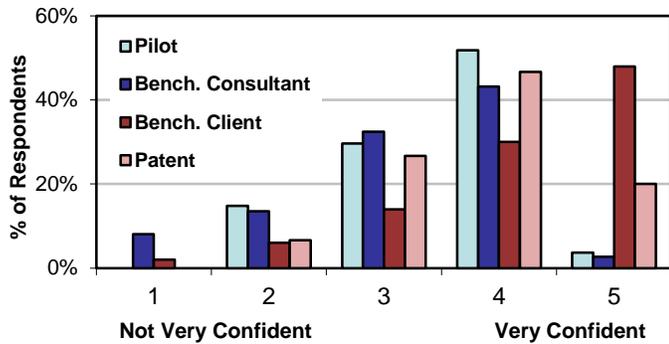


Figure 4. CONFIDENCE THAT OTHER PARTY WILL ALWAYS ACT IN RESPONDENT'S BEST INTEREST.

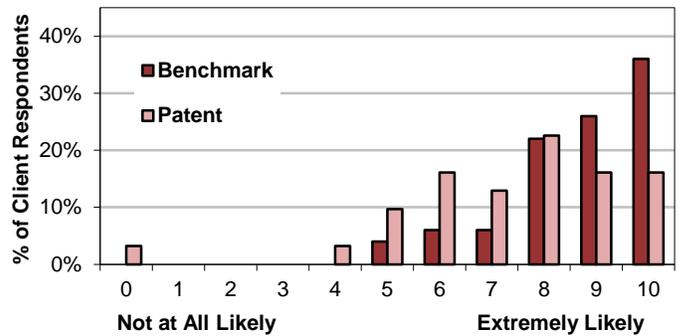


Figure 5. CLIENT'S LIKELINESS TO RECOMMEND CONSULTANCY TO A FRIEND OR COLLEAGUE.

benchmarking consultants ($p = 0.050$), suggesting that the benchmarking sample may provide a non-conservative estimate.

4.4.2. Goodwill trust. Consultant and client respondents were asked to evaluate the likelihood that the other party would always act in the respondent's best interest [34], producing the striking results shown in Fig. 4. Client respondents generally trusted their consultants, though the patent sample did so significantly less than the benchmarking sample (Wilcoxon rank-sum, $p = 0.019$). Consultants, on the other hand, were dramatically less trusting than the benchmarking clients ($p < 0.001$ for both benchmarking and pilot consultants) and somewhat less trusting than the patent clients ($p = 0.005$ for the benchmarking consultants, $p = 0.093$ for the pilot consultants). This finding contradicts an oft-repeated theme from the outsourcing strategy literature, that clients should be wary of suppliers "holding up" the deliverable until a ransom (in the form of increased prices or a contract extension) is paid. It appears that in the consulting industry the client firms hold the upper hand, which is not surprising given that design consultancies are small, relatively abundant, and somewhat interchangeable (in terms of services offered, if not value or quality). Consultant interviewees told anecdotes of client firms refusing to pay invoices for services rendered, or scapegoating the consultancy for problems in development. Such egregious acts were relatively rare. More common were instances of client firms pressuring the consultancies for more work than the project contract warranted.

4.5. Client Satisfaction Measures

4.5.1. Net Promoter Score. Clients were asked how likely they were to recommend their consultancy to a friend or colleague (Fig. 5). The benchmarking sample scores are significantly higher than the patent sample (Wilcoxon rank-sum, $p < 0.001$). The strong negative skew of the benchmarking distribution is typical of customer satisfaction in competitive markets [35]. The more uniform distribution of the patent sample suggests that dissatisfied individuals may have been more likely to return the survey.

62% of the benchmarking sample and 32% of the patent sample rated their consultancy a 9 or 10 and would be considered "Promoters" under the Net Promoter rubric. 10 and 32% provided scores of 6 or lower and would be considered "Detractors." Subtracting the percentage of Detractors from the percentage of Promoters yields the Net Promoter Score (NPS), which is +52% for the benchmarking sample and 0% for the patent sample. These values are very good and fairly poor, respectively, compared to those of other industries. For example, the brokerage and investments industry has an NPS of +35%, whereas the health insurance industry trails most others with a score of -5% [36].

4.5.2. American Customer Satisfaction Index measures. Data were also collected using the three measures that comprise the American Customer Satisfaction Index (ACSI): the extent to which the client's expectations were met, how well the consulting service compared to a hypothetical ideal, and the client's overall satisfaction with the consulting service. The patent sample significantly lags the benchmarking sample in all three measures ($p = 0.0004, 0.0014, 0.0007$).

Although the exact weighting factors used for averaging the three measures to form the ACSI are industry-specific and proprietary to the survey's developer, they differ minimally from equal weighting. Using equal weighting, the resulting scores for the benchmarking and patent samples are 80.7 and 66.7, respectively. Like the Net Promoter scores, these two values differ significantly ($t = 3.35, p = 0.0015$, two-sided) and are very good and fairly poor, respectively, relative to published benchmarks [37]. For example, recent ACSI surveys place express delivery services at 83, wireless carriers at 71, the airline industry and cable television at 66, and the U.S. federal government at 65. High performing companies such as FedEx and Apple score in the mid to upper 80s, while low performers such as Comcast and United airlines score in the low 60s. Unfortunately, very little data is publicly available for business-to-business services, and even less for highly customized knowledge-based services such as innovation consulting.

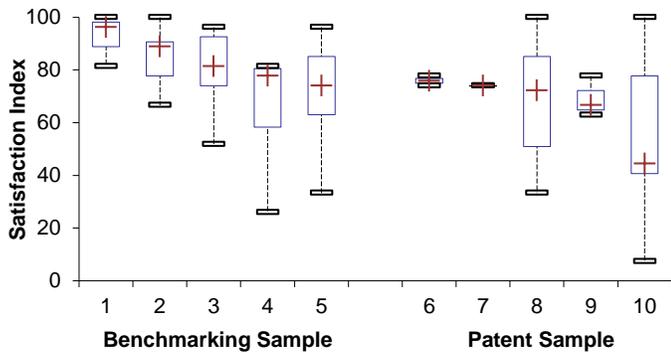


Figure 6. VARIATION IN CLIENT SATISFACTION BETWEEN CONSULTANCIES.

Figure 6 presents score distributions for the consultancies participating in the benchmarking study as well as the consultancies used for the patent sample. Disregarding consultancy 1 for small sample size, the median scores generally range from the upper 60s to the upper 80s.

4.5.3. Repeat business. Although not a perfect measure of client satisfaction, repeat business is symptomatic of it [38]. 47 to 65% of the projects were followed by additional projects, with an additional 7 to 9% pending at the time of the survey (Fig. 1). The pilot sample had less repeat business than the other two but the difference is not significant.

5. DISCUSSION

The results suggest several important findings. First, despite implications to the contrary in the popular press, outsourced product development is not necessarily a “disaster.” This study examined one particular type of outsourcing – domestic design consulting – and found that rates of product commercialization and market success compared favorably to benchmarks from in-house product development. Even the patent sample, which likely underestimated satisfaction due to non-response bias, produced scores comparable to those of major industries such as the airlines and cable television. Customers may grumble about the airlines, but they continue to fly, and usually get to their destinations in one piece. Few would characterize the airline industry as a disaster, and neither is the domestic design consulting industry.

Having said that, the data do indicate large and significant variation in outcomes between projects. Even within projects, some respondents were far less satisfied than others. Prior reports have generally obscured this variation by presenting limited summary data. The variation is significant because it suggests that poor project outcomes can be improved upon, by better project planning and management. We provide suggestions for doing so in a separate paper [39].

One aspect of project management is responding to problems. Prior research has focused on problems as a success measure but may have overstated their impact. The present

results suggest that problems are abundant and not necessarily detrimental. Indeed, a project that is entirely absent of problems is either solving a trivial need or is over-resourced. While the product development literature has disparaged over-use of “fire-fighting” [40], limited use of fire-fighting is generally more cost-effective than complete fire prevention. The relative scarcity of “serious problems” suggests that the developers were able to suppress most of the problems that did ignite. Likewise, the fact that consultants reported more problems than clients may indicate that the consultants resolved them before the clients noticed. Nonetheless, about one-third of problems went unresolved, suggesting opportunity for improvement.

The relatively small size of this study makes it difficult to pinpoint precise values for any of the measurements, but the benchmarking and patent samples likely represent upper and lower bounds, respectively. All the biases in the benchmarking study push it towards over-estimating success, while most of those in the patent study lead to under-estimation.

The magnitude of the difference between the benchmarking and patent samples is large relative to the variation within the samples and represents a significant methodological finding. Despite examining ostensibly the same class of projects, the two samples produced client satisfaction results at the upper and lower edges of what is normally observed in the ACSI for companies and industries. For explanatory research, this is something of a boon. Use of the two samples enabled collection of data with much greater variation than would have been gathered from either one alone, making regression and other explanatory analyses much more powerful. For descriptive research, the power of sampling design is more troubling. Results depend heavily on the sampling method and must be interpreted accordingly. Design service providers should be aware that audits of client satisfaction will likely produce optimistic estimates, particularly if inadequate incentives and protections are offered to the dissatisfied.

Overall, the results highlight the many challenges inherent in measuring success and failure. Despite its growth, design consulting remains relatively rare and its customers hard to locate, making it difficult to achieve the scale needed for highly accurate results. Rigorous customer satisfaction measures such as the ACSI strive for 250 respondents for each rated company [41], but a typical consultancy might only perform forty projects per year. Even when customers can be located, data collection is often frustrated by client firms’ failure to track project outcomes and reluctance to disclose what information they do have. Another challenge is that success and failure is a sensitive topic whose measure has consequences for the research informants. A concern in the benchmarking sample was that consultant project managers might view the study as a means for their employer to evaluate their performance, and therefore be tempted to influence the results. To mitigate this risk, project manager identities were not collected, even though this has been shown to be a strong explanatory factor in other research, and the results were thoroughly anonymized before being reported back to the consultancies.

6. CONCLUSIONS

This paper presented original empirical data on project outcomes in design and innovation consulting. The work contributes to scholarship by providing unique descriptive data for both the product development community, which has not measured outsourced product design success extensively, as well as the customer satisfaction community, which has only begun to investigate knowledge-based business-to-business professional services. The study also developed a novel indirect method of identifying informants via trace evidence in public records. The results contribute to practice by providing baseline project outcome data, and by demonstrating the impact of sampling methods when measuring customer satisfaction.

The chief limitation of the study is its small size. A second limitation is that the patent sample, while not significantly different from the benchmarking sample on demographic measures, is not truly representative of all consulting projects. Sampling from patent applications excludes projects that did not yield anything worth patenting, as well as those whose scope did not include technical or aesthetic invention. More and more, design consultancies focus on front-end user research and design strategy. While these may generate insights that lead to a patent, the detailed design is so far downstream that the consultant researchers are rarely listed as inventors, so these projects would not have been located. A third limitation is the modest response rate of the patent sample. Two points should be made. First, the addresses used for the respondents were identified from sales directories and similar sources and not confirmed before mailing. While some of the invalidly addressed ones were returned as undeliverable, others probably were not, so the actual response rate may have been higher. Second, due to limited funds and a desire not to unduly harass the respondents, only one postcard reminder was used after the initial questionnaire mailing. Best practices recommend at least three reminders, the second of which should include a duplicate copy of the questionnaire [42]. Nonetheless, the 15.6% response rate obtained compares reasonably with other research. Hart achieved 18.7% [43] and the PDMA just 2.7% [31]. Response rates of 10-25% are common in business-to-business satisfaction surveys [44]. A final limitation of the study is that the client satisfaction measures used were developed primarily for business-to-consumer products and services, and may not be ideal for business-to-business. Unfortunately, few standards exist in this area.

The study suggests several avenues of future research. A larger dataset would improve descriptive accuracy. More rigorous testing should be performed to ensure the reliability of the survey instrument. Factor analysis might be used to examine the relationships between the various success measures and to determine which have the strongest impact on client satisfaction and other key outcomes. The present study focused on design consulting, which is only one form of outsourced product development. Useful comparisons could be made to other types, such as integrated design-and-manufacturing outsourcing or low-cost offshoring. Comparisons also need be made to in-house product

development. A useful question might be to ask client respondents if the project would have been more or less successful if it had been performed in-house.

Most importantly, need exists to explain the causes of the substantial variation in outcomes between projects. Numerous theories have been advanced in the literature, but few have been empirically tested [39]. During the survey phase of this project, we collected data on several dozen potential explanatory variables, such as the scale and scope of the project, the capabilities of the consultant and client, and the distance between them. We now turn our attention to developing statistical models to explain and perhaps even predict the outcomes of particular outsourced design projects.

ACKNOWLEDGMENTS

Funding for this project was provided by the MIT-Portugal Program. We thank the design consultancies and survey respondents who generously donated their time to make the survey possible, as well as the three anonymous reviewers whose suggestions significantly improved the manuscript.

REFERENCES

- [1] Gasvoda, E. L., 2003, "Outsourced Design: Strategic Implications and Effective Management of Dispersed, Inter-Firm Product Development," M.S. thesis, Massachusetts Institute of Technology, Cambridge.
- [2] Grant, P. L., 2000, "Outsourced Knowledge: Knowledge Transfer and Strategic Implications from Design Outsourcing," M.S. thesis, Massachusetts Institute of Technology, Cambridge.
- [3] Walsh, V., 1996, "Design, Innovation and the Boundaries of the Firm," *Research Policy*, **25**(4), pp. 509-529.
- [4] Piachaud, B., 2005, "Outsourcing Technology: Here Are Guidelines for Establishing Sourcing Partnerships That Are Effective and Truly Strategic," *Research Technology Management*, **48**(3), pp. 40-46.
- [5] Anderson, E. G., Jr, Davis-Blake, A., Erzurumlu, S. S., Joglekar, N. R., and Parker, G. G., 2007, "The Effects of Outsourcing, Offshoring, and Distributed Product Development Organization on Coordinating the N.P.D. Process," *Handbook of New Product Development Management*, C. Loch, and S. Kavadias, eds., Butterworth-Heinemann, Amsterdam, pp. 259-290.
- [6] Amaral, J., Anderson Jr, E. G., and Parker, G. G., 2011, "Putting It Together: How to Succeed in Distributed Product Development," *Sloan Management Review*, **52**(2), pp. 51-58.
- [7] Barragan, S., Cappellino, C., Dempsey, N., and Rothenberg, S., 2003, "A Framework for Sourcing Product Development Services," *Supply Chain Management*, **8**(3/4), pp. 271-280.
- [8] Bruce, M., and Morris, B., 1994, "Managing External Design Professionals in the Product Development Process," *Technovation*, **14**(9), pp. 585-599.

- [9] Munsch, K., 2004, "Outsourcing Design and Innovation," *Research Technology Management*, **47**(1), pp. 27-30.
- [10] Parker, G. G., and Anderson, E. G., Jr, 2002, "From Buyer to Integrator: The Transformation of the Supply-Chain Manager in the Vertically Disintegrating Firm," *Production and Operations Management*, **11**(1), pp. 75-91.
- [11] Goering, R., 2006, "Design Outsourcing: Navigating a Maze of Decisions," *EE Times*, (1405), January 9, p. 1.
- [12] Rassam, C., 1995, *Design and Corporate Success*, Design Council/Gower, Aldershot, England.
- [13] Engardio, P., and Einhorn, B., 2005, "Outsourcing Innovation," *BusinessWeek*, (3925), March 21.
- [14] Amaral, J., and Parker, G., 2008, "Prevent Disasters in Design Outsourcing," *Harvard Business Review*, **86**(9), pp. 30-31.
- [15] Anderson, E. G., Jr, Davis-Blake, A., and Parker, G. G., 2008, "Managing Outsourced Product Design: The Effectiveness of Alternative Integration Mechanisms," Working paper, McCombs School of Business, Austin, TX.
- [16] Mokhoff, N., and Wallace, R., 2005, "Outsourcing Trend Proves Complex by Design," *EE Times*, (1356), January 31, p. 10.
- [17] Rayner, B., 2005, "Design Outsourcing: Beware of Pitfalls," *EE Times*, (1356), January 31, p. 20.
- [18] Roy, R., and Potter, S., 1993, "The Commercial Impacts of Investment in Design," *Design Studies*, **14**(2), pp. 171-193.
- [19] Rothstein, A. J., 1998, "Outsourcing: An Accelerating Global Trend in Engineering," *Engineering Management Journal*, **10**(1), pp. 7-14.
- [20] "Be Careful When Outsourcing Design Work," 2007, *Manufacturing Engineering*, **139**(6), December, p. 30.
- [21] Bursky, D., 2000, "Design Outsourcing: Are Companies Doing Too Much?" *Electronic Design*, December 4, p. 52.
- [22] Carson, S. J., 2007, "When to Give up Control of Outsourced New Product Development," *Journal of Marketing*, **71**(1), pp. 49-66.
- [23] Kimzey, C. H., and Kurokawa, S., 2002, "Technology Outsourcing in the U.S. And Japan," *Research Technology Management*, **45**(4), pp. 36-42.
- [24] Bruce, M., Potter, S., and Roy, R., 1995, "The Risks and Rewards of Design Investment," *Journal of Marketing Management*, **11**(5), pp. 403-417.
- [25] Roy, R., and Potter, S., 1990, "Managing Design Projects in Small and Medium-Sized Firms," *Technology Analysis & Strategic Management*, **2**(3), pp. 321-336.
- [26] Barczak, G., Griffin, A., and Kahn, K. B., 2009, "Trends and Drivers of Success in NPD Practices: Results of the 2003 PDMA Best Practices Study," *Journal of Product Innovation Management*, **26**(1), pp. 3-23.
- [27] Anderson Jr, E. G., Parker, G. G., and Davis-Blake, A., 2011, "Outsourced Product Development Integration Study." <http://pdoutsourcing.org>.
- [28] Palm, W. J., and Whitney, D. E., 2010, "Exploring the Meaning of Success in Outsourced Product Development," *Proceedings of the ASME 2010 IDETC/CIE*, Montreal, Quebec.
- [29] Core77 Inc., 2010, "Core77/Businessweek Design Firm Directory." <http://www.designdirectory.com/>.
- [30] Lee, R. M., 1993, *Doing Research on Sensitive Topics*, SAGE Publications, London.
- [31] Griffin, A., 1997, "PDMA Research on New Product Development Practices: Updating Trends and Benchmarking Best Practices," *Journal of Product Innovation Management*, **14**(6), pp. 429-458.
- [32] Page, A. L., 1993, "Assessing New Product Development Practices and Performance: Establishing Crucial Norms," *Journal of Product Innovation Management*, **10**(4), pp. 273-290.
- [33] Booz Allen and Hamilton, 1982, *New Products Management for the 1980s*, Booz Allen Hamilton, New York.
- [34] Sako, M., 1992, *Prices, Quality, and Trust: Inter-Firm Relations in Britain and Japan*, Cambridge University Press, Cambridge.
- [35] Fornell, C., 1995, "The Quality of Economic Output: Empirical Generalizations About Its Distribution and Relationship to Market Share," *Marketing Science*, **14**(3), pp. G203-G211.
- [36] Satmetrix Systems, Inc., 2011, "2011 NPS Benchmarks for the U.S. And Europe." <http://www.netpromoter.com/resources/benchmarks.jsp>.
- [37] American Customer Satisfaction Index, 2011, "ACSI Results." http://theacsi.org/index.php?option=com_content&view=article&id=12&Itemid=110.
- [38] Yi, Y., 1990, "A Critical Review of Consumer Satisfaction," *Review of Marketing*, V. A. Zeithaml, ed., American Marketing Association, Chicago, pp. 68-123.
- [39] Palm, W. J., and Whitney, D. E., 2010, "Improving Outcomes in Outsourced Product Development: A Joint Consultant-Client Perspective," *Proceedings of the 21st Annual Conference of the Production and Operations Management Society*, M. D. Hanna, ed., Vancouver, BC.
- [40] Repenning, N. P., 2001, "Understanding Fire Fighting in New Product Development," *Journal of Product Innovation Management*, **18**(5), pp. 285-300.
- [41] Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., and Bryant, B. E., 1996, "The American Customer Satisfaction Index: Nature, Purpose, and Findings," *Journal of Marketing*, **60**(4), pp. 7-18.
- [42] Mangione, T. W., 1995, *Mail Surveys: Improving the Quality*, SAGE Publications, Thousand Oaks, CA.
- [43] Hart, S., 1993, "Dimensions of Success in New Product Development: An Exploratory Investigation," *Journal of Marketing Management*, **9**(1), pp. 23-41.
- [44] Infosurv, 2011, "Customer Research FAQs." <http://www.infosurv.com/faqs/customer-research-faqs/>.

effectiveness of the innovation system in which they operate (access to information, to advice, to funds, to potential collaborators etc.).
Separating the effect of an intervention; How to design and manage an evaluation for the type of measure, according to the focus of the evaluation (relevance, value for money, results and impacts). The main methods used to evaluate each type of measure and their pros and cons and possible other approaches. A summary of key "pointers" to keep in mind during the evaluation.