Measuring the Impact of Data Quality on an EDI Enabled Interorganizational Business Process

Bas H.P.J. Vermeer

b.h.p.j.vermeer@tm.tue.nl / bvermeer@bakkenist.nl

Abstract

The Data quality Impact (DQI) method is a method that quantifies the impact of poor data quality on a business process. This paper describes the application of this method in a case study where we measured the impact of Data Quality on an interorganizational EDI enabled ordering process. We compared the interorganizational ordering processes between a technical materials wholesaler and two customers, one using EDI and one not. Using the DQI method we could demonstrate that poor Data Quality resulted in more checking activities and a higher level of resulting errors in the EDI enabled interorganizational business process. The contribution of this paper is that it demonstrates the applicability of the DQI method to quantify the impact of data quality. However, the way of measuring impact on a business process should be improved using Activity Based Costing. Furthermore, more empirical data using the method should be collected to compare outcomes with a 'normal' situation. This enhances the usability of the method.

Introduction

Electronic Data Interchange (EDI) is an important technology in redesigning interorganizational business processes (Venkatraman 1994, Clark & Stoddard 1996). Using EDI results in more efficient business processes through reduction in paper handling related activities, less data-entry errors, elimination of the data-entry function, a reduction of throughput time and hence in a reduction of inventory costs (Emmelhainz 1993). Furthermore, EDI results in more effective business processes. For instance, Davenport (1993) demonstrates in his IT-Process-Productivity relationship model that IT initiatives (such as EDI) generate new process design options, which lead to more process productivity.

Although several studies regarding operational costs and benefits of EDI confirm these positive effects, (Mackay 1993, Wrighly 1994, Mukhopadhyay et. al. 1995, Jelassi and Figon 1994, Reekers and Smithson 1996), several others do not (Benjamin et. al. 1990, Carter 1990, McCusker 1994, Riggins and Mukhopadyay 1994).

In a field study in the food sector we investigated problems of using scanning and EDI technology to explain these contradictory results. This study revealed that many of these problems emerged from insufficient data quality between the databases of suppliers and retailers (Vermeer 1996-b en -c). The negative impact of poor data quality was confirmed in another case study of EDI partners in the pharmaceutical sector. This case study also revealed EDI problems due to major inconsistencies in the article data of the participants' databases (Vermeer 1996-a).

The negative impact of data quality on an automated process can be explained as follows. The complete automation of processes, such as using EDI in the ordering process, requires high quality of product data in the supply chain. However, the actual quality of the information in the databases does not match these high requirements because it is still based on human use. Human use requires lower data quality because humans can interpret inconsistent or even not available information and then correct it. For instance, when a purchaser uses the computer system to order a product, and the delivery time is not available, the order can still be processed. The purchaser simply contacts the supplier and asks for the delivery time. Then he processes a general order, which means that he orders the product using a dummy product number. In this way, the order is processed, although the quality of the product data in the database is insufficient for automatic processing.

Thus, when evaluating the use of EDI in interorganizational business processes, two counteracting effects appear. Firstly, using EDI leads to more efficient and more effective interorganizational business processes, resulting in an increase of process performance. Secondly, using EDI leads to higher data quality requirements. Combined with poor actual data quality in the chain this leads to more errors in the EDI enabled interorganizational business process leading to a decrease in process performance. These opposite effects are shown in figure 1.

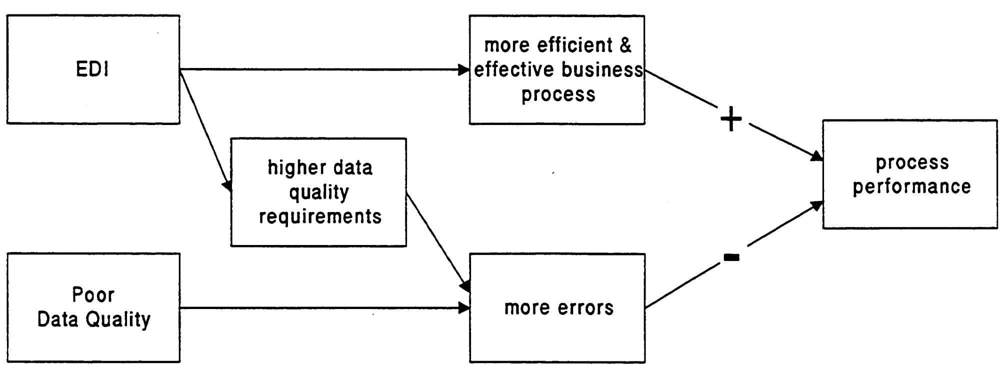

Figure 1. Influence of EDI & Data quality on process performance

When researchers want to evaluate the impact of EDI on interorganizational business process performance, they need to incorporate the impact of poor data quality. The question is how? In an earlier paper (Vermeer 1998) we introduced the Data Quality Impact (DQI) method. This method helps researchers to quantify the impact of data quality on a business process. This paper describes a case study where we used the DQI method to quantify the impact of data quality on an EDI enabled interorganizational business process.

In section 2 we will introduce the DQI method we have developed to quantify the impact of data quality on a business process. Section 3 describes the case study settings and methodology. In section 4 we will discuss the results of defining data quality in the case, after which in section 5 the processing time of the data handling process is quantified. In section 6 we will present the results of assessing the impact on the EDI enabled interorganizational business process. In section 7 conclusions are presented and future research is discussed.

The Data Quality Impact Method

To measure the impact of data quality on a business process we need to answer two questions: How do we measure data quality and how do we measure the effect of this quality on a business process?

Measuring data quality

Wang and Strong (1996) define data quality as data that are fit for use by data consumers. In the DQI method we developed three levels of testing, based on this users' perspective (see figure 2).

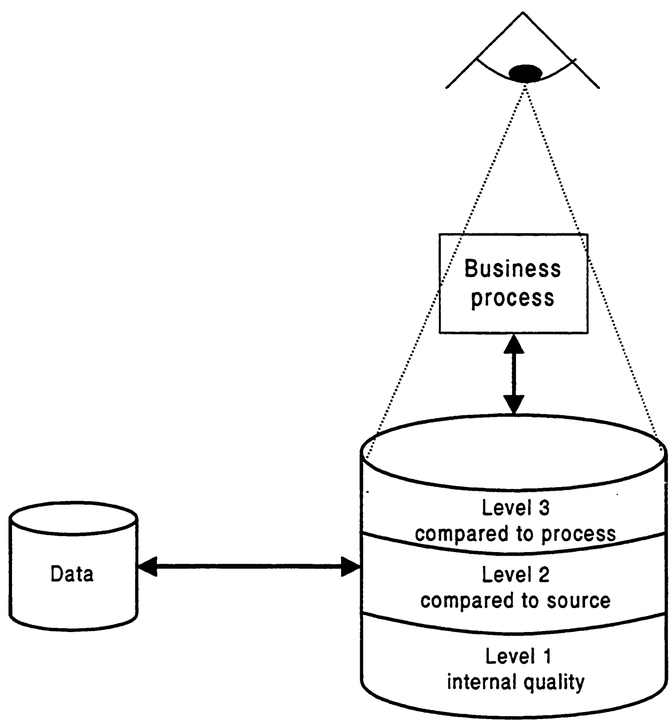

Figure 2. Data quality tests in the DQI method

The first level tests measure the internal data quality of the data, which means that the conformance of the data with the data model is tested. Typical checks are: the percentage of the primary key field values that are identical, or the percentage of mandatory fields that contains NULL values.

The second level tests compare the data in the database with the source data. This means that data in the receivers' database is compared with the corresponding data in the senders' database. Here, accuracy, depth completeness and to some extent, timeliness are tested. With depth completeness we mean the degree that facts (for instance, articles) in the senders' database for the receiver are indeed available in the receivers' database and vice-versa. This measure is also an indirect measure of the timeliness of the data. Opposed to depth completeness we define breadth completeness, which is the degree that data fields (or attributes) in the receivers' database are indeed available in the senders' database. Accuracy is tested through testing for instance whether the price in the receivers' database is the same as in the senders' database.

As we can see from the first- and second level tests, they are still largely independent from the business process. The third level tests explicitly focus on assessing data-quality in respect to the business process. For the interorganizational ordering process we developed two tests. Firstly, it is important to check whether the available data fields in the receivers' database are relevant for the EDI ordering application (=breadth completeness). For instance, in a case study in the Electrotechnical sector (Vermeer 1998) we found that the delivery time was not available in the database, although this field was indispensable for placing an order. Secondly, measuring the actual use of the data source is important. This can be tested through measuring how many order lines were processed using the database. In terms of Wang & Strongs' framework of data quality dimensions (Wang & Strong 1996), this tested the believability and reputation of the article data.

Measuring impact

To measure the impact of data quality on a business process, the DQI method firstly separates the business process from the data handling process (see figure 3). The data handling process is the process where data is collected, checked, entered in the database, stored and distributed. The business process uses data from the database to produce business transactions or physical goods. Some examples of business processes are: ordering, invoicing, goods reception in the warehouse or production. Secondly, the performance of each process is measured. The performance of the data handling process is measured as the processing time of the data (e.g. the time of collecting, checking, entering, storing and distributing data) and the resulting data quality to the business process. The performance of the business process is measured as the processing time (of an order, an invoice or a good) and the resulting number of errors.

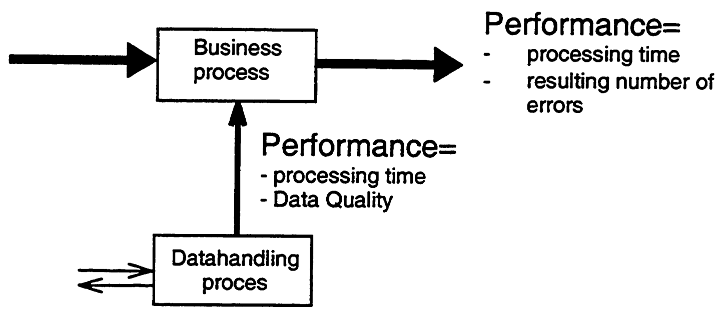

Figure 3. Measuring impact of data quality on a business process

Finally, the impact can be measured through varying the data quality of the data handling process, or through comparing differently implemented business processes, based on the same data quality. In the case study in this paper, where we measured the impact of data quality on EDI enabled processes, we used the second option.

Costs of prevention and correction

In the Electrotechnical case study where the DQI method was developed (Vermeer 1998), we found that a low level of data quality could still result in few resulting errors in the business process. However, the processing time of the business process was seriously increased. We found that this was the result of an increase in preventive checking activities in the business process. Therefore, we have added an extra step to the method, where we measure the change in checking activities compared to the change in processing time resulting from a change in the data quality. This will not change the results, but will clarify how the change in data quality is absorbed in the business process: more prevention (extra checking) or more correction (more errors).

The Schwartz Case

E. Schwartz BV is a Dutch wholesaler that supplies ball bearings and other technical materials to the industrial market. E. Schwartz BV is a typical medium sized company with 40 employees and a yearly turnover of 13 million US dollars. In the Dutch technical materials sector, where EDI is virtually non-existent, E. Schwartz BV successfully implemented EDI with one supplier and two customers.

Case methodology

The objective of the case study was to quantify the impact of data quality on an EDI enabled business process using the DQI method. Important research questions were: Is the DQI method applicable in practice? What are important shortcomings? How should the method be improved? Therefore, the focus of the case study was on directions for further improvement.

The interorganizational ordering process between E. Schwartz BV and the customers was chosen as the EDI enabled business process. To measure the impact of data quality on an EDI enabled business process, we developed two mini-cases: One case with an EDI enabled customer and one case with a non-EDI customer. Each process used the same database for processing the customer orders. Through measuring the difference in processing time and the resulting number of errors, we wanted to quantify the impact of data quality on an EDI enabled business process.

Validity

To test the validity of the case study we checked two things. Firstly, is the EDI enabled ordering process with the first customer comparable with the non-EDI ordering process with the second customer? Secondly, are the differences in processing time and number of errors the result of the quality of the data in the database?

Comparable processes?

The first customer is a cigarette manufacturer, who uses EDI to process their orders. They order about 4000 different articles (such as maintenance and spare parts) using a reorderpoint strategy. This means that orders are completely electronically generated. The second customer, who is not-EDI enabled, is a ship handler. They order about 2000 different articles that they supply to ships that enter the harbor and have to reload. Since every purchasing order of the ship handler is preceded by a quotation, all information for placing the order is already available in the ordering process. Therefore, the ship handler has started to implement EDI in its ordering process, which should result in direct orders that are driven by the quotation process. Since both customers have used (or will use) EDI to eliminate rekeying of information by the purchasers, we conclude that the ordering processes of both customers are comparable.

Are differences in costs and resulting errors the result of data quality?

We developed the following strategy to check whether differences in costs and resulting errors are the result of data quality. Firstly, we reviewed all the activities in the two ordering processes. Then we checked which activities were especially influenced by the EDI initiative, such as the data-entry activity. These activities should explain the difference in processing time due to the EDI initiative. Secondly, we reviewed all activities that are influenced by data quality e.g. the checking activities. These activities should explain the difference in processing time due to data quality. Finally, we checked all errors that resulted after the ordering process and checked which were caused by data quality problems. Only those errors were taken into account to explain the difference in numbers of errors. This strategy should guarantee that we measure only the effects of data quality on an EDI enabled business process.

Assessment of Data Quality

The quality of the data was tested through comparing the article data in the database of E. Schwartz BV with a subset of the article data of the EDI customer (the cigarette manufacturer) and the non-EDI customer (the ship handler). The overall results of the data quality tests on the three quality levels are displayed in table 1.

Level	Important characteristics	Schwartz & EDI customer	Schwartz & non-EDI customer
1	duplicates in key	0,0%	0,1%
2	not matched	18,0%	4,0%
	net price differences	5,8%	6,8%
	% of quality problems	23,9%	10,9%
	Data Quality	**76,1%**	**89,1%**
3	Breadth incompleteness	0,0%	0,0%
	actual non-use of database	<9,8%	<0,3%

Table 1. Data quality for the ordering process with 2 Schwartz customers

For the internal data quality tests (level 1 tests) the two customers provided the article data they normally requested from E. Schwartz BV on floppy disk. The requested data consisted of: *article number customer, description, description E. Schwartz BV, article number E. Schwartz BV, gross price, discount percentage, packing unit, price unit, delivery time, minimum order size and net price.* Firstly, we tested if duplicates existed in the article numbers for E. Schwartz BV, or the two customers. Table 1 shows that the EDI customer had no duplicates on 4293 records. The non-EDI customer file contained two duplicates on 2082 records. The article database of E. Schwartz BV also had no duplicates, although the most important article search field contained 87 duplicates on 54.954 article numbers. Next we checked whether the field contents complied with the field constraints. For E. Schwartz BV and both customers no important deviations were found.

In the level 2 tests, we compared 11 fields (see previous paragraph) with the corresponding fields in the Schwartz article database. Although sometimes considerable differences were found (especially in field descriptions) we have concentrated on the fields that are indispensable for order processing. These were the article numbers and the net prices. In the net price check we considered differences due to price unit errors and gross price errors (the non-EDI customer did not use a net price, but calculated the net price through applying a discount percentage to the gross price). Table 1 shows that 18% of the article numbers in the database of the EDI customer were not available in the Schwartz database. This was considerably more than the 4% in the database of the non-EDI customer. Finally, 5.8% of the net prices of the EDI customer were different, compared to 6.8% of the net prices of the non-EDI customer.

Finally, in the level 3 tests we checked breadth completeness and the actual use of the article database by the EDI and non-EDI ordering processes. Through interviewing the ordering staff of E. Schwartz BV and both customers we learned that the information in the database was sufficient to process an order (see table 1). Furthermore, we found that 9.8% of the order lines with the EDI customer were processed with product numbers that were not available in the database of E. Schwartz BV. This could indicate that the article database was not used to process the order (which is possible, if the order was manually processed). However, further research revealed that many of these order lines contained product codes that were removed during the year, but possibly available, when the order was placed. For the non-EDI customer we found that only 0.3% of such order lines existed.

Since we are not able to determine the difference between phantom orders and valid orders with already removed article numbers, and since breadth completeness was 100%, we decided to define the overall data quality up to level 2. Table 1 shows that in total 23.9% of the shared article numbers with the EDI customer contained quality problems. This means that the data quality with the EDI customer is 76.1%. In a similar way the data quality with the non-EDI customer is 89.1%. There are several reasons to explain the difference in data quality between the EDI and non-EDI ordering process. We will elaborate on that in the following section.

Processing Time of Data Handling

The following step in the DQI method is to quantify the processing time in the data handling process. To do this, we first made a detailed description of the procedure for transferring data between E. Schwartz BV and the EDI and non-EDI customer respectively. Next we measured the processing time through measuring the time that the data administration staff spent on the different steps in the procedure. Finally, we compared the total processing times for exchanging article data of the customers. The results are shown in table 2.

EDI customer			
company	stepnr	description	# minutes
EDI-customer	1	Production of article-price disk & send to Schwartz	5
Schwartz	2	Run query with new prices of customer in own database and check prices	960
Schwartz	3	update special (POA) prices	480
Schwartz	4	Enter delivery times per article and send to EDI customer	480
EDI-customer	5	Reading updated floppy disk and rechecking prices	240
		Total	**2165**
Non-EDI customer			
company	stepnr	description	# minutes
non-EDI-cust	1	Request for article-price diskette to Schwartz	5
Schwartz	2	Production of diskette and sending to non-EDI customer	20
Schwartz	3	Reception and reading at computer department non-EDI cust.	5
non-EDI-cust	4	Production of list and final check at Purchasing	480
		Total	**510**

Table 2. Processing costs for the EDI- and non-EDI customer

As we can see from the table, the data processing time of the data handling process at the EDI customer is much higher than the processing time at the non-EDI customer. The main reason for this difference is that E. Schwartz BV does not check the prices before they send them to the non-EDI customer.

Despite these lower processing costs at the non-EDI customer, the data quality with the non-EDI customer is higher. There are two reasons two explain this. Firstly, the data-administration process between E. Schwartz BV and the EDI customer is much more complex because all prices are based on a net price contract. Since these contracts are not available in the information system, all prices have to be checked manually. This is especially important since several net prices co-exist, based on order quantities. The net price for the non-EDI customer is calculated via a discount on the gross prices. Since the gross prices and the discount percentage are already available in the Schwartz system, no extra checks are necessary.

Secondly, the low data quality percentage for the EDI customer is mainly the result of non-matched articles (18%). Further research showed that this was caused because E. Schwartz BV already had removed a number of articles, which appeared to be still in use. This problem was resolved through better agreements on how to handle expired product numbers.

Impact on the Ordering Process

The final step in the DQI method is to measure the changes in processing time and the resulting number of errors.

Changes in processing time

To quantify the processing time in the ordering process, we first described the complete ordering processes between E. Schwartz BV and the EDI customer and non-EDI customer respectively. The ordering process was defined as follows: the ordering process starts from the moment that an order is generated at the customer (through exceeding the order level at the EDI customer or through the decision of accepting one quotation at the ship handler). The ordering process stops after entering the order in the information system of E. Schwartz BV.

The total processing time was measured using time registration forms at E. Schwartz BV and through time estimations provided by the customers. The results are shown in table 3.

EDI customer		EDI	checking	total
EDI customer	Generate order advise and checking	2	26	28
Schwartz	Manual transport from PC to AS/400	3	0	3
Schwartz	Repairing missing order lines in order	0	7	7
Schwartz	Checking & repairing prices	0	11	11
	Total time per order line (sec)	**5**	**44**	**49**

Non EDI customer		non EDI	checking	total
non-EDI customer	Manual entry quotation in ordersystem	30	0	30
Schwartz	Answer phone & checking order	1	22	23
Schwartz	Data-entry Sales	22	0	22
	Total time per order line (sec)	**53**	**22**	**75**

Effects				
	positive EDI effect	-64%		
	negative extra prevention effect		+30%	
	total effect			-34%

Table 3. Positive effect of EDI and negative effect of preventive checking

Table 3 shows a processing time difference of 26 seconds (=34%) between the EDI enabled and traditional ordering processes. As we can see from table 3, the positive effect of using EDI leads to a decrease of 64% in processing time. However, since extra checking activities are necessary, an increase of 30% is measured in the EDI enabled process. This result clearly shows that the measured benefits from using EDI decrease because of extra checking activities in the EDI enabled process.

Changes in error correction

Finally we measured the number of errors that remained after the ordering process. To do this, we looked up all invoices of the passed year and analyzed all invoice errors that had occurred. Then we classified these in errors that had surfaced in the ordering process and errors that occurred in other processes, such as the delivery process, or the invoicing process itself. The results of this analysis are shown in table 4.

EDI customer			
	Coupling error customer	0,29%	
	Wrong net price	0,21%	
	Return of inventory	0,25%	
	Coupling error Schwartz	0,14%	
	Data-entry error Schwartz	0,07%	
	Total % errors in ordering process		**0,97%**
	Errors in other processes	0,57%	
Non-EDI customer			
	Wrong product ordered	0,05%	
	Wrong net price	0,05%	
	Wrong delivery address	0,16%	
	Data-entry error	0,11%	
	Total % errors in ordering process		**0,38%**
	Errors in other processes	0,33%	

Table 4. Analysis of remaining errors

Typical errors with the EDI enabled customer were: *Coupling errors,* which are the result of article number matching problems between the article database of the customer and E. Schwartz BV. *Wrong net prices,* because of price differences. *Return of inventory errors,* which resulted because the customer stored two equal products under different private product numbers. These errors are especially expensive. *Data entry errors,* when the Sales department of E. Schwartz BV accidentally miscorrected some prices. The typical errors for the non-EDI customer speak for themselves.

Table 4 shows first of all that the number of errors in both ordering processes is lower than 2% of all order lines, which is very low. Furthermore, the number of errors without EDI is more than twice as low as the number of errors with EDI. The main reason is that net prices, coupling errors at the customer and return of inventory errors are not detected through preventive checking. However, these errors do not occur in the non-EDI process, since in that process the buyer and the seller talk with each other on the phone. We learned from interviews that in this conversation, problems with prices and mismatches in product numbers are detected before they can really occur.

Conclusions on impact

Comparing the two mini-cases, we see that the EDI enabled ordering process had a processing time that was 34% less than the traditional ordering process. If both processes are indeed comparable, we may conclude that implementing EDI resulted in a decrease in processing time of 34%. However, we also found that two negative effects occur due to the impact of data quality. Firstly, if we consider only the positive effect of using EDI, the decrease in processing time could have amounted up to 64%. This positive effect is negatively compensated with an increase of 30% in processing time due to extra checking activities. These activities are directly resulting from data quality problems. Secondly, the number of errors resulting from data quality problems in the ordering process increased from 0,38% to 0,97%.

Conclusions and Further Research

The objective of the case study was to test whether the DQI method is applicable in practice and how it could be further improved. The case study showed that the method produces meaningful results. However, there are several points of attention. Firstly, the DQI method uses processing time as the main performance indicator. In the case, processing time was operationalized as manual labor time. The question is whether this does not underestimate the time and costs of electronic order processing. Also, other costs, such as accommodation costs of staff or overhead costs, are not incorporated. Since these costs may be substantial compared to direct labor time (and costs) we think that all these costs should be included. Therefore we propose to include Activity Based Costing (Cooper and Kaplan 1988) into the method. Secondly, it is still unclear how to interpret the different levels of data quality in relation to the decrease in processing time and the increase in remaining errors. This is because at this moment no empirical data is available about this relationship besides this case. However, such data would greatly enhance the usability of the DQI method, since it enables a comparison of new cases with a 'normal' situation.

We propose three directions for further research. Firstly, more case studies must be conducted using the DQI method to collect empirical data. As we stated in our second conclusion, this will enhance the usability of the method substantially. Secondly, the method can be used to test hypotheses about the impact of data quality on business processes. One drawback however, is that many case studies must be conducted to test the external validity of these hypotheses. Thirdly, although the method provides evidence of the impact of data quality, research is needed to provide solutions to improve data quality. In our research we are developing a method named Information Logistics that aims at synchronizing product data across an interorganizational business network. The result is a high degree of data quality across the interorganizational network. In practice this is referred to as Data Alignment. We use the DQI method in a preliminary phase to analyze the seriousness of data quality problems in the interorganizational business network.

Acknowledgements

We would like to thank E. Schwartz BV for their support in conducting this case study and EAN Nederland for supporting this research.

References

Benjamin R.I., D.W. de Long and M.S. Scott Morton, "Electronic Data Interchange: How Much Competitive Advantage?", Long Range Planning, 23:1 (1990) pp. 29-40.

Carter J.R., "The Dollars and Sense of Electronic Data Interchange", Production & Inventory Management Journal, 31:2 (1990) pp. 22-26.

Clark T.H. and D.B. Stoddard, "Interorganizational Business Process Redesign: Merging Technological and Process Innovation", Journal of Management Information Systems 13:2, (fall 1996), pp. 9-28.

Cooper R. and R.S. Kaplan, "Measure Costs Right, Make the Right Decisions", Harvard Business Review, sept/oct (1988), 96-103.

Davenport T.H., "Process Innovation, Reengineering Work through Information Technology", Boston: Harvard Business Press (1993).

Emmelhainz M.A., "EDI: A Total Management Guide", Van Nostrand Reinhold, USA (1993).

Jelassi T. and O. Figon, "Competing through EDI at Brun Passot: Achievements in France and Ambitions for the Single European Market", MIS Quarterley, 18:4 (1994), pp. 337-352.

Mackay D.R., "The Impact of EDI on the Components Sector of the Australian Automotive Industry", Journal of Strategic Information Systems, 2:3 (1993), pp. 243-263.

McCusker T., "How to Get More Value from EDI", Datamation 40:9 (May 1994), pp. 56-60

Mukhopadhyay T., S. Kekre and S. Kalathur, "Business Value of Information Technology: a Study of Electronic Data Interchange", MIS Quarterly, 19:2 (1995), pp. 137-155.

Reekers N. and S. Smithson, "The Role of EDI in Interorganizational Coordination in the European Automotive Industry", European Journal of Information Systems vol. 5, (1996) pp. 120-130.

Riggins and Mukhopadhyay. "Interdependent Benefits from Interorganizational Systems: Opportunities for Business Partner Reengineering", Journal of Management Information Systems, 11:2 (fall 1994), pp. 37-57

Venkatraman N., "IT-enabled Business Transformation: from Automation to Business Scope Redefinition", Sloan Management Review (Winter 1994), pp. 73-87.

Vermeer B.H.P.J., "The Information Distribution Center: Distribution Mechanism for Distributing Article Master Data in the Pharmaceutical Sector", Utrecht, The Netherlands: Neprofarm In Dutch (1996-a).

Vermeer B.H.P.J., "Articles Are Not Alike! An Investigation of Problems due to Poor Synchronization of Article Master Data in the Grocery Sector", Eindhoven, The Netherlands: TU Eindhoven Report EUT/BDK/77, In Dutch (1996-b).

Vermeer B.H.P.J., "EDI: Electronic Distribution of Information products", In proceedings of the Second EDIspuut Workshop 'Bridging Worlds', M. Hoogeweegen and B. Vermeer (eds.), 's Hertogenbosch, The Netherlands: Edispuut (1996-c).

Vermeer B.H.P.J., "Impact of Data Quality on the Ordering Process: How to Assess Data Quality Impact", accepted paper to be presented at the EDISPUUT conference in Oct 1998, Utrecht, The Netherlands

Wang R.Y. and D.M. Strong, "Beyond Data Accuracy: What Data Quality Means to Data Consumers", Journal of Management Information Systems, 12:4 (Spring 1996), pp. 5-34.

Wrighly C.D., R.W. Wagenaar and R.A. Clarke, "Electronic Data Interchange in the International Trade: Frameworks for the Strategic Analysis of Ocean Port Communities", Journal of Strategic Information Systems, 3:3 (1994), pp. 211-234.