Business Intelligence: An Analysis of the Literature

Zack Jourdan, R. Kelly Rainer, and Thomas E. Marshall
Department of Management, College of Business, Auburn University, Auburn, Alabama, USA

Abstract This research collects, synthesizes, and analyzes 167 articles on a variety of topics closely related to business intelligence (BI) published from 1997 to 2006 in ten leading Information Systems (IS) journals. We found a generally increasing level of activity during the 10-year period and a focus on exploratory research methodologies. We noted that several methodologies were either underrepresented or absent from the pool of BI research. We also identified several subject areas that need further exploration.

Keywords business analytics, business intelligence, competitive intelligence, data mining, data warehousing, literature review, content analysis

Managers and researchers alike have been working to develop IS that provide business intelligence (BI). BI is “both a process and a product.” The process is composed of methods that organizations use to develop useful information, or intelligence, that can help organizations survive and thrive in the global economy. The product is information that will allow organizations to predict the behavior of their “competitors, suppliers, customers, technologies, acquisitions, markets, products and services, and the general business environment” with a degree of certainty (Vedder, Vanecek, Guynes, & Cappel, 1999).

The stakes are high for organizations to develop successful BI implementations. Winning companies, such as Continental Airlines, have seen investments in BI generate increases in revenue and produce cost savings equivalent to a 1000% return on investment (ROI) (Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006). On the other hand, losing companies have spent more resources than their competitors with a smaller ROI, all while watching their market share and customer base continuously shrink (Gessner & Volonino, 2005).

Two related needs provide the motivation for this paper. First, the growing body of BI research necessitates a review of this literature with the intent of “identifying critical knowledge gaps and thus motivate researchers to close this breach” (Webster and Watson, 2002). Second, as noted by Scandura & Williams (2000), in order for research to advance, the methods used by researchers must periodically be evaluated to provide insights into the methods utilized and thus the methods which should also be used in a given research field. This study analyzes the BI literature and then proposes an agenda for future research efforts.

The remainder of the paper is organized as follows. Section 2 describes the approach to the analysis of BI research, Section 3 contains the results of the research, and Section 4 discusses the limitations of the project and possible future research efforts.

Research Study

We examined the number and distribution of BI articles published in leading journals, the methodologies employed in BI research, and the research topics being addressed in BI research. During our analysis, we identified gaps in the research which would allow us to propose and discuss a research agenda that will facilitate the progression of BI research (Webster & Watson, 2002). We hope to paint a representative landscape of the current BI literature base in order to influence the direction of future research efforts in this important field.

In order to examine the current state of research on BI, we conducted a literature review and analysis in three phases. First, we accumulated a representative pool of
articles, then classified the articles by research method, and finally classified the articles by research topic.

Accumulation of Article Pool

Because BI research is typically published in many IS journals, we searched through a ten year period (1997–2006) of ten leading IS journals. In order to decide which journals to search, we chose four recent rankings for IS journals (Rainer & Miller, 2005; Lowry, Romans, & Curtis, 2004; Katerattanakul, Han, & Hong, 2003; Peffers & Ya, 2003). By selecting the top journals in each of these four rankings, we created a list of ten journals (see Table 1). We deliberately did not search journals devoted to one particular area of business intelligence because we wanted general, mainstream journals, rather than specialized journals.

We then used the ABI/INFORM database to search for the research articles by searching the titles and abstracts of each of the ten journals using phrases such as “business analytics,” “business intelligence,” “data mining,” and “data warehousing.” Because we were looking for research articles in the topic area of BI, we eliminated any result that was a book review or an editorial.

Classification by Research Strategy

Once we identified the articles, we examined the research strategy used in each article and categorized it according to that strategy. Due to the subjective nature of method classification, we performed a content analysis of the articles. Figure 1 shows the process we followed, which was adapted from Neuendorf (2002). First, we defined the research method categories utilizing those presented in Scandura and Williams (2000), who extended the research strategies initially described by McGrath (1982). Specifically, we used nine categories of research strategies: Formal theory/literature reviews (FT/LR), sample survey, laboratory experiment, experimental simulation, field study (primary data), field study (secondary data), field experiment, judgment task, and computer simulation. To guard against the threats to reliability (Neuendorf, 2002), we performed a pilot on unused articles, discussed the results, and refined the definitions.

Each research strategy is defined by a specific design approach and each is also associated with certain trade-offs that researchers must make when designing a study. These trade-offs are inherent flaws that limit the conclusions that can be drawn from a particular research strategy. These trade-offs refer to three aspects of a study that can vary depending on the research strategy employed. These variable aspects include: generalizability from the sample to the target population (external validity); precision in measurement and control of behavioral variables (internal and construct validity); and the issue of realism of context (Scandura & Williams, 2000).

Cook and Campbell (1976) stated that a study has generalizability when the study has external validity across times, settings, and individuals. Formal theory/literature reviews and sample surveys have a high degree of generalizability because they establish the relationship between two constructs and illustrate that this relationship has external validity. A research strategy that has low external validity but high internal validity is likely to be generalizable for other contexts.

Table 1. Journals in Study

<table>
<thead>
<tr>
<th>#</th>
<th>Journal Title</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIS Quarterly</td>
<td>MISQ</td>
</tr>
<tr>
<td>2</td>
<td>Information Systems Research</td>
<td>ISR</td>
</tr>
<tr>
<td>3</td>
<td>Communications of the ACM</td>
<td>CACM</td>
</tr>
<tr>
<td>4</td>
<td>Journal of Management Information Systems</td>
<td>JMIS</td>
</tr>
<tr>
<td>5</td>
<td>Management Science</td>
<td>MS</td>
</tr>
<tr>
<td>6</td>
<td>Journal of the ACM</td>
<td>JACM</td>
</tr>
<tr>
<td>7</td>
<td>European Journal of Information Systems</td>
<td>EJIS</td>
</tr>
<tr>
<td>8</td>
<td>IEEE Transactions on Software Engineering</td>
<td>IEEESE</td>
</tr>
<tr>
<td>9</td>
<td>Information &amp; Management</td>
<td>I&amp;M</td>
</tr>
<tr>
<td>10</td>
<td>Harvard Business Review</td>
<td>HBR</td>
</tr>
</tbody>
</table>
times, settings, and populations. While the formal theory/literature reviews and sample surveys have a high degree of generalizability and the laboratory experiment has a high degree of precision of measurement, these strategies have low degree of realism of context. The only two strategies that maximize degree of realism of context are field studies using either primary or secondary data because the data is collected in a field setting (Scandura & Williams, 2000).

The other four strategies maximize neither generalizability, nor degree of precision of measurement, nor degree of realism of context. This point illustrates the futility of using only one strategy when conducting IS research. Because no one strategy can maximize all types of validity, it is best for researchers to use a variety of research strategies. Table 2 contains an overview of the nine strategies and their ranking on the three strategy tradeoffs (Scandura & Williams, 2000).

We then classified the articles independently as to research strategy. We coded only a few articles at a time to minimize coder fatigue and thus protect intercoder reliability (Neuendorf, 2002). Upon completion of the independent classification, we tabulated agreements and disagreements, intercoder crude agreement (percent of agreement), and intercoder reliability using Cohen’s kappa (Cohen, 1960). The latter two calculations were well within the acceptable ranges for intercoder crude agreement and intercoder reliability (Neuendorf, 2002). We calculated the reliability measures prior to discussing disagreements as mandated by Weber (1990). If two reviewers did not agree on how a particular article was coded, a third reviewer arbitrated the discussion of how the disputed article was to be coded. This process resolved the disputes in all cases.

### Classification by BI Research Topic

To classify articles by research topic, we held several brainstorming and discussion sessions where we attempted to identify BI topics with the intent to categorize the diverse body of literature. In these discussion sessions, we sought to synthesize the literature and provide a better understanding of the current state of BI research.

Once we established the category definitions, we independently placed each article in one BI category. As before, we placed only a few articles at a time to minimize coder fatigue and thus protect intercoder reliability (Neuendorf, 2002). Upon completion of the classification process, we tabulated agreements and disagreements, intercoder crude agreement (percent of agreement), and intercoder reliability using Cohen’s kappa (Cohen, 1960).

---

**Table 2. Research Strategies**

<table>
<thead>
<tr>
<th>Research Strategy</th>
<th>Description</th>
<th>Degree of Precision of Measurement</th>
<th>Degree of Realism of Context</th>
<th>Generalizability to Target Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal Theory/Literature Reviews</td>
<td>Summarization of the literature in an area of research in order to conceptualize models for empirical testing.</td>
<td>Low</td>
<td>Low</td>
<td>Maximizes</td>
</tr>
<tr>
<td>Sample Survey</td>
<td>The investigator tries to neutralize context by asking for behaviors that are unrelated to the context in which they are elicited.</td>
<td>Low</td>
<td>Low</td>
<td>Maximizes</td>
</tr>
<tr>
<td>Laboratory Experiment</td>
<td>Participants are brought into an artificial setting, usually one that will not significantly impact the results.</td>
<td>Maximizes</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Experimental Simulation</td>
<td>A situation contrived by a researcher in which there is an attempt to retain some realism of context through use of simulated situations or scenarios.</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Field study: Primary Data</td>
<td>Investigates behavior in its natural setting. Involves collection of data by researchers.</td>
<td>Low</td>
<td>Maximizes</td>
<td>Low</td>
</tr>
<tr>
<td>Field Study: Secondary Data</td>
<td>Involves studies that use secondary data (data collected by a person, agency, or organization other than the researchers.</td>
<td>Low</td>
<td>Maximizes</td>
<td>Low</td>
</tr>
<tr>
<td>Field Experiment</td>
<td>Collecting data in a field setting but manipulating behavior variables.</td>
<td>Moderately high</td>
<td>Moderately high</td>
<td>Low</td>
</tr>
<tr>
<td>Judgment Task</td>
<td>Participants judge or rate behaviors. Sampling is systematic vs. representative, and the setting is contrived.</td>
<td>Moderately high</td>
<td>Low</td>
<td>Moderately high</td>
</tr>
<tr>
<td>Computer Simulation</td>
<td>Involves artificial data creation or simulation of a process.</td>
<td>Low</td>
<td>Moderately high</td>
<td>Moderately high</td>
</tr>
</tbody>
</table>

*Source: Scandura & Williams, 2000.*
for each category. Again, the latter two calculations were well within the acceptable ranges (Neuendorf, 2002). We again calculated the reliability measures prior to discussing disagreements as mandated by Weber (1999). If two reviewers did not agree on how a particular article was coded, a third reviewer arbitrated the discussion of how the disputed article was to be coded. This process also resolved the disputes in all cases.

Results

Using the described search criteria within the selected journals, we collected a total of 167 articles. (For the complete list of articles in the sample, see Appendix A.) We then analyzed the articles’ year of publication, journal, and author.

Table 3 shows the number of articles per year in our sample. With BI issues becoming more important to researchers and practitioners, we see a generally increasing trend in the number of articles per year.

In order to determine the degree to which leading IS journals publish BI research, we calculated the percentage of BI articles based on the total number of articles published in each journal in the ten-year period. Calculating this percentage involved several steps. First, three sample issues of each journal were examined to determine the average total number of articles per issue. We counted any article within the journal over five pages in length to eliminate editors’ notes, book reviews, and editorial pages. Second, the number of articles per issue was multiplied by the number of issues that each journal publishes in a year. This number gave us the estimated number of total articles for a journal during the ten-year period we were studying. Finally, we calculated the percentage of BI articles by dividing the actual number of BI articles in a journal by the estimated total number of articles in a journal. As shown in Table 4, some top IS journals publish very little BI research while others devote a substantial amount of space to this research.

### Table 3. Number of BI Articles per Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of BI Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>5</td>
</tr>
<tr>
<td>1998</td>
<td>6</td>
</tr>
<tr>
<td>1999</td>
<td>10</td>
</tr>
<tr>
<td>2000</td>
<td>24</td>
</tr>
<tr>
<td>2001</td>
<td>17</td>
</tr>
<tr>
<td>2002</td>
<td>9</td>
</tr>
<tr>
<td>2003</td>
<td>26</td>
</tr>
<tr>
<td>2004</td>
<td>18</td>
</tr>
<tr>
<td>2005</td>
<td>25</td>
</tr>
<tr>
<td>2006</td>
<td>27</td>
</tr>
</tbody>
</table>

### Table 4. BI Articles as a Percentage of Total Articles

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>BI Articles/Year</th>
<th>Years</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Journal of Information Systems</td>
<td>18</td>
<td>20</td>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>MIS Quarterly</td>
<td>17</td>
<td>20</td>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>Information &amp; Management Communications</td>
<td>51</td>
<td>84</td>
<td>10</td>
<td>840</td>
</tr>
<tr>
<td>of the ACM Research</td>
<td>10</td>
<td>24</td>
<td>10</td>
<td>240</td>
</tr>
<tr>
<td>Journal of Management Information Systems</td>
<td>12</td>
<td>44</td>
<td>10</td>
<td>440</td>
</tr>
<tr>
<td>Journal of the ACM</td>
<td>5</td>
<td>30</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>Management Science</td>
<td>19</td>
<td>120</td>
<td>10</td>
<td>1200</td>
</tr>
<tr>
<td>Harvard Business Review</td>
<td>10</td>
<td>90</td>
<td>10</td>
<td>900</td>
</tr>
<tr>
<td>IEEE Transactions on Software Engineering</td>
<td>4</td>
<td>48</td>
<td>10</td>
<td>480</td>
</tr>
</tbody>
</table>

Analysis of Research Strategies in BI Research

The categorization of the 167 articles according to the nine research strategies produced the following results. Ninety-four articles (56%) were classified as Formal Theory/Literature review making it the most prevalent research strategy. Other categories, in decreasing order, are Field Study—Primary Data (21 articles), Field Study—Secondary Data (19 articles), Sample Survey (13 articles), Computer Simulation (10 articles), Lab Experiment (7 articles), and Field Experiment (3 articles). No articles were classified as either Experimental Simulation or Judgment Task.

Analysis of the research strategies over the ten year period from 1997 to 2006 (Table 5) illustrates that Formal Theory/Literature Review is the most prevalent research strategy. Other categories, in decreasing order, are Field Study—Primary Data, Field Study—Secondary Data, and Sample Survey are represented in almost every year of the time frame. These four strategies are exploratory in nature and indicate the beginnings of a body of research (Scandura & Williams, 2000).

BI Research Categories

As we analyzed the articles, five relatively distinct categories emerged. The Artificial Intelligence (AI) category consists of algorithms and applications of AI. The applications of the AI category addressed classification, prediction, web mining and machine learning. The Benefits category details how organizations have used data warehousing, data mining, and/or an enterprise-wide BI systems to achieve some measurable financial benefit. The Decision category contains articles related to improving overall decision making and includes such subjects as data modeling, decision-making, and decision modeling. The Implementation category covers project management issues.
in a variety of BI contexts including data warehousing, data mining, customer relationship management (CRM), enterprise resource planning (ERP), knowledge management systems (KMS), and eBusiness projects.

The final and most diverse category is Strategies. This category focuses on how to apply BI tools and technologies in the modern business environment. The category covers such diverse topics as improving internal performance (i.e., enterprise agility, marketing, and integrating business functions), working with external partners to improve collaboration in the supply chain, and providing the customer a better experience through customization/personalization and customer relationship management (CRM).

BI research covers diverse subjects ranging from practical applications of neural networks (Baesens, Rudy, Mues, & Vanthienen, 2003), to end-user satisfaction (Chen, Soliman, Mao, & Frolick, 2000), to the use of clustering as a business strategy to gain a competitive advantage (Porter, 1998). The researchers conducting BI research have backgrounds as varied as marketing (Cui, Wong, & Lui, 2006), management information systems (Watson, Goodhue, & Wixom, 2002), and computer science (Menzies, Chen, Hihn, & Lum, 2006). To integrate BI research stemming from such disparate backgrounds, we placed the BI literature into research categories. In Table 6, the BI categories are displayed along with some of the topics relevant to each category.

These five categories provided a subject-area classification for all of the 167 articles in our research sample. Fifty-nine articles were classified as Strategies, making it the most prevalent BI category. This category was followed by Artificial Intelligence (37 articles), Implementation (35 articles), and Decisions (26 articles). These four research strategies accounted for 94% of the articles in the sample. The category Benefits represented the remaining six percent (10 articles). These numbers illustrate the categories of BI research that have received the most and least attention in the IS journals.

### BI Category vs. Research Strategy

An examination of BI category versus research strategy (Table 7) identifies the research strategies used in articles on the various BI topics. Articles on all five BI topics used the FT/LR research strategy most often. Approximately one-half the articles in the Artificial Intelligence, Benefits, Decisions, and Implementation categories utilized the FT/LR strategy. However, almost three-fourths of the articles in the Strategies topic area used the FT/LR strategy.

Other than FT/LR, the technology-focused category of Artificial Intelligence used field-secondary and computer simulations the most frequently. The technology-focused category of Decisions used lab experiments most frequently. On the other hand, the less technical categories of Benefits, Implementation, and Strategies relied more heavily on sample surveys and field studies using both primary and secondary data.

The rationale for these findings is as follows. First, the FT/LR strategy is the most appropriate strategy for the early stages of research in any area. In these exploratory years of BI research, formal theory/literature reviews are appropriate for theory formulation and development. Second, researchers in business schools may be more skilled in administering sample surveys and performing field studies, and may not typically employ strategies such as laboratory experiment,

### Table 6. BI Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Topics</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Intelligence</td>
<td>Algorithms, Classification, Machine Learning, Prediction, Web Mining</td>
<td>37</td>
</tr>
<tr>
<td>Benefits</td>
<td>Data Mining, Enterprise-wide IS</td>
<td>10</td>
</tr>
<tr>
<td>Decisions</td>
<td>Data Modeling, Decision Making, Decision Modeling</td>
<td>26</td>
</tr>
<tr>
<td>Implementation</td>
<td>CRM, DM, DSS, DW, eBusiness, ERP, KMS, Project Management</td>
<td>35</td>
</tr>
<tr>
<td>Strategies</td>
<td>Collaboration, Competition, Customization, Integration, etc.</td>
<td>59</td>
</tr>
</tbody>
</table>
experimental simulation, judgment task, and computer simulation. For example, the majority of articles found in our search that used the computer simulation research strategy were written by computer scientists and computer engineers, while the majority of articles using sample surveys were written by business researchers. Finally, organizations are often less likely to commit to certain strategies (i.e., primary and secondary field studies and field experiments) because these strategies are more expensive for the organizations. These types of research strategies are very labor intensive to the organization being studied because records will need to be examined, personnel will need to be interviewed, and senior managers will be required to devote large amounts of their time to help facilitate the research project.

Areas for Future Research in Business Intelligence

It is interesting that the Theory Formulation/Literature Review research strategy remains the most prevalent strategy after so many years. The focus of BI researchers could shift to other research strategies. Also notable is that the Survey research strategy has been used so little. Perhaps additional survey research would provide value to this field. In addition, the Benefits topic deserves more attention. One of the reasons for so few articles published in this area is probably the difficulty in quantifying the benefits of improved decision making attributed to BI systems.

Limitations and Directions for Future Research

Our analysis of the BI literature is not without limitations. Future literature reviews could search a broader domain of research outlets. Further, future BI studies should consider the research gaps that we have identified in light of generalizability, precision of measure, and realism of context. Future efforts should also consider the five BI categories with respect to the research strategies.

In addition, much of the research in our sample addresses new technologies and issues in BI without attempting to explain the fundamental issues of information systems research as it relates to BI. This is to be expected in the exploratory stages of research in a subject area. However, our study indicates that enough research in FT/LR has been done to begin formulating guiding theories in BI research.

For researchers to continue to address important questions in BI, future studies need to employ a wider variety of research strategies. Scandura and Williams (2000) stated that looking at research strategies employed over time by triangulation in a given subject area can provide useful insights into how theories are developing. In addition to the lack of variety in research strategy, very little triangulation has occurred during the timeframe used to conduct this literature review. This absence of coordinated theory development causes the research in BI to appear haphazard and unfocused. However, the good news is that many of the categories and research strategies in BI research are open for future research efforts. We hope that this research analysis has laid the foundation for such efforts that will enhance the IS body of knowledge and theoretical progression relative to BI.

Author Bios

Zack Jourdan is an Assistant Professor of Management Information Systems at Georgia Southwestern University. He is a doctoral candidate at Auburn University. His current research interests include various aspects of information security and risk analysis.

R. Kelly Rainer, Jr. is George Phillips Privett Professor of Management Information Systems at Auburn University, Auburn, Alabama. He received his Ph.D from the University of Georgia. His current research interests include information security and risk analysis, and healthcare information systems. His most recent book (co-authored with Efraim Turban) is *Introduction to Information Systems*, 2007, John Wiley and Sons.

Thomas E. Marshall is an Associate Professor of Management Information Systems at Auburn University.
Auburn, Alabama. He received his Ph.D from the University of North Texas. His current research interests include information security, various aspects of database management, and geographical information systems.

References


Appendix A. Complete List of BI Articles


