Large-scale Internet benchmarking: Technology and application in warehousing operations

Andy Johnson [a,*], Wen-Chih Chen [b], Leon F. McGinnis [c]

[a] Department of Industrial and Systems Engineering, Texas A&M University, 237K Zachry Engineering Center, College Station, TX, USA
[b] Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan, ROC
[c] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

Effective benchmarking requires standards for measuring performance across a broad range of organizations. Industry-benchmarking measures the relative performance levels of similar operations, but gathering sufficient data to robustly characterize best performance is the primary hurdle to more widespread implementation. This paper describes the development of an innovative, large-scale benchmarking methodology which employs Internet technology to overcome the data collection problem. We discuss a particular instance of this approach, iDEAs-W—the result of an ongoing collaboration between academia and the warehousing industry. This paper's purpose is to demonstrate the methodology for the warehousing industry and share our insights to allow others to refine the tool for applications in other industries.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Performance benchmarking, the comparison of internal operations at one firm with the best practices at others, was popularized in the late 1980s when significant improvements in performance were realized by Hewlett-Packard and Xerox [1]. This interest continues today, with noteworthy projects such as the Open Standards Benchmarking Collaborative, a project of International Business Machines Corp., Procter & Gamble Co., Shell Oil Co., a unit of Royal Dutch/Shell Group, the U.S. Navy, and the World Bank [2]. Effective benchmarking requires standards for the measurement of performance across a broad range of organizations, and often the most relevant benchmarking information to improve operations arises from industry-level comparisons. Benchmarking studies can provide several benefits: (1) allowing firms to learn from others' experiences; (2) helping firms to analyze their own levels of performance relative to the competition; (3) identifying those firms with the highest (lowest) levels of performance which can then be studied to gain insights about the activities that correlate with high (low) performance.

Fig. 1 shows the three major components of a typical benchmarking study: data, methods and media. *Data* are the key performance indices (KPIs) or measures describing a set of comparable operations. *Methods* analyze and transform the collected data to useful information and/or managerial suggestions. *Media* are the channels by which the information is gathered and the results delivered. Although time- and labor-intensive, benchmark studies can be tightly focused, a favorable attribute in some data collection efforts.

In the typical approach, benchmarking teams use the three components in the following manner. Generally, data is collected from surveys mailed to users, phone interviews, and/or on-site face-to-face interviews. The teams analyze the information, commonly using the partial productivity method, where the level of a single output generated by a firm is compared to the level of a single input consumed. If, however, the benchmarking study uses multiple inputs and outputs, the partial productivity approach will produce several measures. The final step is disseminating the results in report and/or presentation formats.

Despite its appeal, there are drawbacks to this approach. At an operational level many firms lack the analytical tools and personnel to identify best/worst performance. Proprietary concerns prevent competing firms from sharing information. Data collection/limitation problems may arise when firms are unwilling to participate. Finally it can be difficult to assemble data for a peer group large enough to ensure confidence that the industry-level benchmarking has identified the best/worst performance.

Fortunately, a new, Internet-based benchmarking methodology is available to overcome many of these obstacles. While the project described in this paper has received considerable attention from

* Corresponding author.
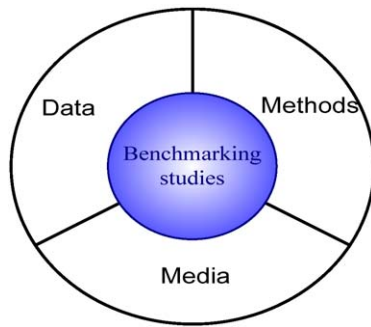 E-mail address: ajohnson@tamu.edu (A. Johnson).

Fig. 1. Components of benchmarking studies.

both industry (e.g. [3,4]) and academia (e.g. [5,6]), the current literature lacks discussion of the development of such a tool. Therefore, this paper describes the development of a general methodology using Internet technology to facilitate benchmarking, illustrating it with an example of an ongoing collaborative effort between academia and industry. The use of Internet technology reduces the difficulties of collecting data while still maintaining data security. Using the example of a warehousing industry-benchmarking tool developed by the partnership between industry and academia, we share our experiences to date, including evaluating method, implementation process, pitfalls, and extensions. We observe that the involvement of an academic institution may ease the proprietary data concerns of the participants. We suggest that the analyses and information that can be made available via online benchmarking as well as the relationship between academic institutions and warehousing industry practitioners can easily be replicated in other industries.

The paper is organized as follows: Section 2 introduces the evaluation methods. Section 3 describes the implementation of an online benchmarking Website, while Section 4 warns of potential pitfalls and discusses open research questions. Section 5 describes the significance of the iDEAs-W project and Section 6 gives some concluding remarks.

2. Evaluating method

Traditionally productivity is measured as a ratio of a single output to a single input, termed single ratio productivity measures [7,8]. However, as production processes have become more complex, multiple inputs are often used to produce more than one output. This leads to a set of single ratio productivity measures which can be confusing to evaluate—a typical multiple criteria evaluation problem. If some measures are good and some are poor, is the firm performing well or poorly?

While there are several candidate approaches, such as AHP, balanced scorecard, and TOPSIS [9,10], we employ Data Envelopment Analysis (DEA) [11] as the system-based performance measure because of its foundation in production theory [12]. DEA allows the efficiency of each observed operation to be estimated relative to a best practice frontier identified from peers in the data set. Firms on the best practice frontier can be identified and differences and similarities between a firm under evaluation and the best practice firms can be observed. The relationship between DEA and single ratio productivity measure has been investigated and described [13].

DEA has been used in a variety of industries to evaluate performance, as demonstrated by 2000-plus academic articles identified by Cooper et al. [14] and Tavares [15]. Recent management science literature discusses efficiency measurement using DEA in a variety of contexts, for example, banking [16], insurance [17], education [18], telecom [19], and software [20].

A key concept in DEA and production theory is the production possibility set, which is the set of all input and output pairs such that the input can be used to produce the output. The boundary of this set is often referred to as the efficient frontier and describes how the most efficient production units trade off inputs and outputs. Rather than assume a functional form for the efficient frontier, nonparametric methods are used to construct a piece-wise linear boundary of the production possibility set using two common assumptions: free disposability and convexity. Free disposability means that excess input can be disposed of without cost. Convexity of the production possibility set is a technical assumption: if two production units are observed, call them W_1 and W_2, then any weighted combination of W_1 and W_2, say $W' = \theta W_1 + (1 - \theta) W_2$ with θ a scalar, $0 < \theta < 1$, is also a feasible production process. Typically, nonparametric efficiency models assume free disposability and convexity. The observations that produce the maximum output attainable from a combination of inputs at the existing state of technical knowledge are found to be efficient. All observations are bounded by the efficient frontier and those not lying on the frontier are defined to be inefficient, with their distance from the frontier (appropriately defined) used to determine their level of efficiency.

3. Implementation

Our goal is to provide an online benchmarking analysis that addresses the need for the performance assessment tools mentioned above. With this innovative tool any firm with Internet access can participate and view the individual performance analysis results in real-time. The large number of participants facilitated by easy electronic access provides a significant data set that can be used for comparative purposes.

Several other online benchmarking tools are available. For example Radgen [5] discusses a tool for benchmarking compressed air systems based on various cost measures. Radgen's tool allows the user to benchmark against itself or external benchmarks that can be identified by having the same 2 or 4 digit industry codes. Bogetoft and Nielsen [21] also develop an online tool to allow 50 different industries to benchmark their performance using capital and labor to generate gross profits. The application is not industry-specific, but compares across industries for the same country. Alternatively, Ma et al. [22] describe an Internet-based mobile agent technology that allows a host enterprise to systematically gather information from distributors' databases to compare distributor performance.

The methodology described in this paper assesses production performance and provides an option for industry-specific comparisons for product types. It allows the retrieval of the exact benchmarking information that best assists users to develop improvement strategies. Additionally it provides an integrated technical efficiency measure that is independent of cost, since the latter can vary widely depending on a production process's operational region.

Before benchmarking at an industry level can occur, an appropriate high-level model of the inputs and outputs must be established. Specific processes must be designed for collecting, maintaining and analyzing data, and reporting the results.

3.1. Input–output model

The specification of an input–output model is challenging and requires extensive domain knowledge of the production process. In this sub-section we provide a few general principles and then discuss the construction of a warehouse production model as a specific example.

DEA is a nonparametric efficiency estimation method that suffers from the curse of dimensionality, Simar and Wilson [23].

Thus, the amount of data to estimate larger models, as described by the sum of the number of inputs plus the number of outputs, grows rapidly in the model size. A parsimonious specification of the model is critical, with only the most critical inputs and outputs to the production process included. These are not necessarily the most costly goods since there are often public goods that cost a firm nothing, but the production process cannot be performed without these inputs or bad outputs, Fare et al. [24]. The inputs and outputs must also be substitutable; having more of one input allows the firm to have less of another input yet still maintain the output production level. It also means that reducing production of one output allows for more production of a different output, while holding inputs constant. If two variables, either inputs or outputs, are perfect functions of each other, one should be removed. However, DEA can be formulated as a regression problem, Kuosmanen and Johnson [25], so positive or negative covariance of two inputs or outputs is of no particular concern. In fact monotonicity of the frontier suggests that inputs should be correlated. Because cost or revenue does not necessary drive the selection of inputs and outputs, reviewing the financial statements of a firm does not typically contribute to the specification of a good model. Rather, the best models are typically the result of thoroughly understanding the underlying production process.

In the warehouse industry, system performance assessment is yet to be widely implemented. Currently, the most common measures of performance are ratios, e.g. *lines shipped per labor hour*. This type of measure alone is not very informative because of the possibility for input substitution. For instance, with a given output level, warehouses with more capital may be able to use less labor since the capital substitutes for labor. Therefore, it is desirable to use a measure that considers all the significant inputs and outputs simultaneously.

The performance model used in our analysis of warehouses is based on Hackman et al. [26], and it was developed further through a series of round-table discussions with industry practitioners and the academics studying warehouse operations. The model identifies the most critical inputs and outputs that define the production process for the warehousing industry: labor, space, and equipment (inputs), and broken case lines shipped, full case lines shipped, pallet lines shipped, accumulation and storage (outputs).

Labor is measured as annual labor hours including both direct and indirect labor to perform receiving, moving, storing, retrieving, order picking and shipping. Indirect labor, such as management, planning, and equipment maintenance is included. However, indirect supporting personnel, e.g., security, cleaning staff, office assistants, accounting, human resources, customer service, and any labor assigned to value-adding activities, are not counted. Space is measured as the area (in square feet) dedicated to receiving, put away, storing, retrieving, order picking, packing, and shipping. Areas for supporting activities, e.g., office, bathrooms, cafeteria, break rooms, and space for value-adding activities, are not included. For multistory buildings or mezzanines, total usable square footage is reported, not building footprint. Rather than attempting to capture actual cost or book value to represent the value of the equipment, we developed an equipment inventory, and applied a standard cost to determine an equipment investment that is normalized across all warehouses.

A warehouse typically provides three valuable services: storing goods, demand processing and order consolidation. Hackman et al. [26] propose a storage function measure calculated by the formula:

$$p\sqrt{\text{broken case SKUS}} + (1 - p)(5\sqrt{\text{pallet locations}}$$
$$+ \sqrt{\text{floor storage sq. footage}}) \qquad (1)$$

where p denotes the proportion of lines picked as individual items or "broken case".

To clarify the services, we use some of the warehousing vocabulary as follows. A *customer order* is an individual customer's request to be fulfilled by the warehouse. It generally includes product types and the quantity for each. *Number of lines* is the number of different product types in a customer order. A *pallet line* is one that is picked and shipped by pallets. Similarly, *full case and broken case lines* are different product types picked and shipped in full case and individual items (sometimes referred to as "eaches"). The work in picking each of these different types of lines varies significantly; therefore all three are identified as separate outputs. To quantify *order consolidation*, we sum the lines shipped in a year and subtract the number of orders in that year and call this the accumulation index. Note that when all orders are "single-line orders" the accumulation index is zero. The data required to characterize resources and products are readily available to the warehouse manager, and the definitions are unlikely to vary from one warehouse to the next.

3.2. Benchmarking system architecture

Obviously Internet-based performance measurement is interactive; users can log in, enter data, select a peer group, and run a performance analysis. If the firm is found to be inefficient, the user can easily adjust various resource or output data values, or vary the comparison group. Interaction facilitates learning by helping the user to quickly process very large data sets and understand the relationships between the adjusted variables or the peer group and efficiency.

The mathematical computations required for DEA can be performed using any of a number of commercially or publicly available optimization engines. The following software is integrated in the current version of our benchmarking tool: PHP scripting language, MySQL^{TM} database (from MySQLAD), AMPL^{TM} optimization package (from ILOG), and Perl and Apache (from the Apache Software Foundation) to communicate between the various software packages.

We use a three-tier application (shown in Fig. 2) consisting of a Web-application platform, database, and decision support system. The user first sees the set of Webpages generated by the server. These pages are used to gather production process characteristics data from the user and to return the results. The data created by the user is maintained in the database, and can be edited in subsequent interactive sessions. Periodically the database is reviewed and analyzed to define "qualified" warehouses. These are observations that pass an outlier detection process and are complete, including contact information. When an observation is qualified, it is marked as an element of the reference set of production process observations used to calculate efficiency estimates for future users.

3.3. Functions

The two types of information provided by an Internet benchmarking tool are individual firm evaluations and industry-level trends. From our experience, we observe that participants tend to find their individual evaluations the most helpful, and often the evaluations are the initial motivators for participating in a benchmarking study. The following sections elaborate on the information and analysis provided both online and offline by our tool.

The current implementation of the online benchmarking tool can provide efficiency estimates, gap analysis (piecharts describing the connection between partial productivity analysis and the efficiency estimates as shown in Fig. 3), and practice and attribute information for the efficient production processes identified as benchmarks.

Overall efficiency estimates are calculated from both the input and output orientations based on a variable returns to scale DEA model [27]. The input efficiency estimates range from 0 to 1, where
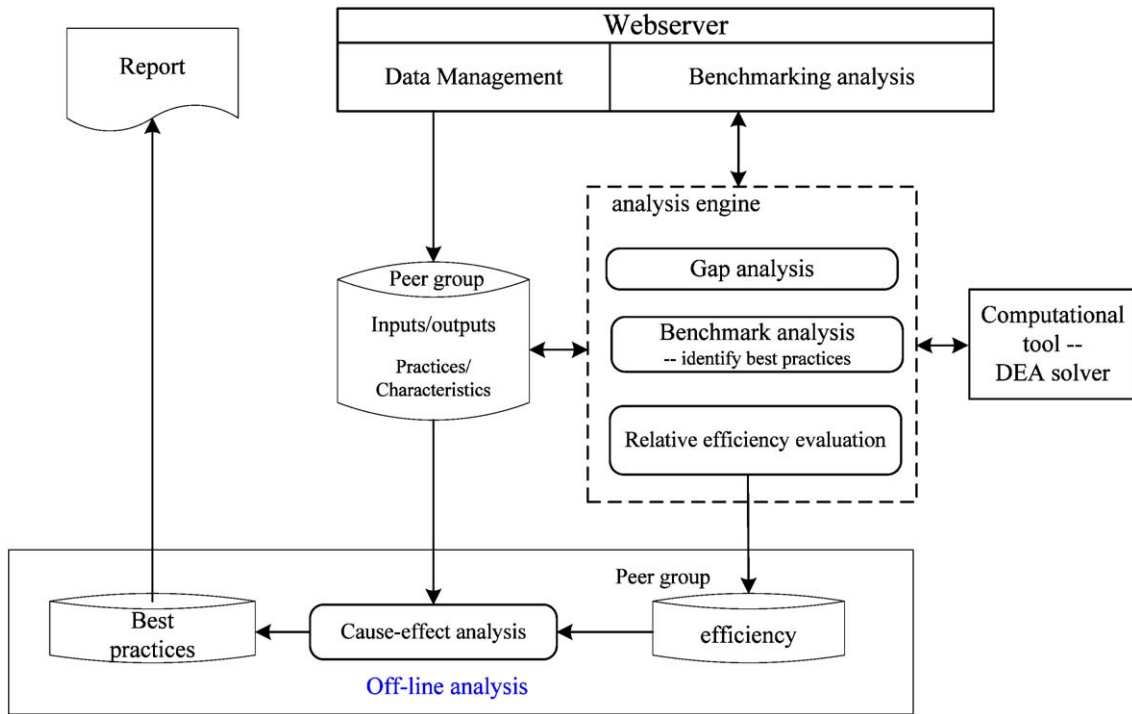
Fig. 2. System architecture.

1 indicates the operation is on the efficient frontier; thus efficient and any value less than 1 can be multiplied by the current (observed) input levels to estimate input levels that would be appropriate for an efficient operation producing the same output levels. Output efficiency estimates range from 1 to infinity, where 1 corresponds to an operation on the efficient frontier, or an efficient operation, and a value greater than 1 represents the amount by which all outputs must increase to achieve efficient operations for the given input levels.

Suppose an observed production process has an input efficiency level of 0.86 and an output efficiency level of 1.3. These numbers indicate the observation should be able use 86% of its current input levels and still produce its current output levels, and it should be able to produce 30% more output without changing its input levels.

In the variable returns to scale model it is not necessary for the input efficiency level to be the reciprocal of the output efficiency level. The efficiency levels give an indication of the magnitude of improvement possible for the given production process. Webpage-initiated calculations allow the user to view DEA efficiency estimates in seconds; the speed of calculation makes it easy to change the current values for various data elements describing the production process and to observe how the changes impact the efficiency estimates.

The Website also provides the more traditional partial productivity measures. To illustrate this in a warehousing context consider lines shipped per labor hour. For example, this single factor or partial productivity measure for warehouse i is turned into a corresponding single factor or partial efficiency measure by
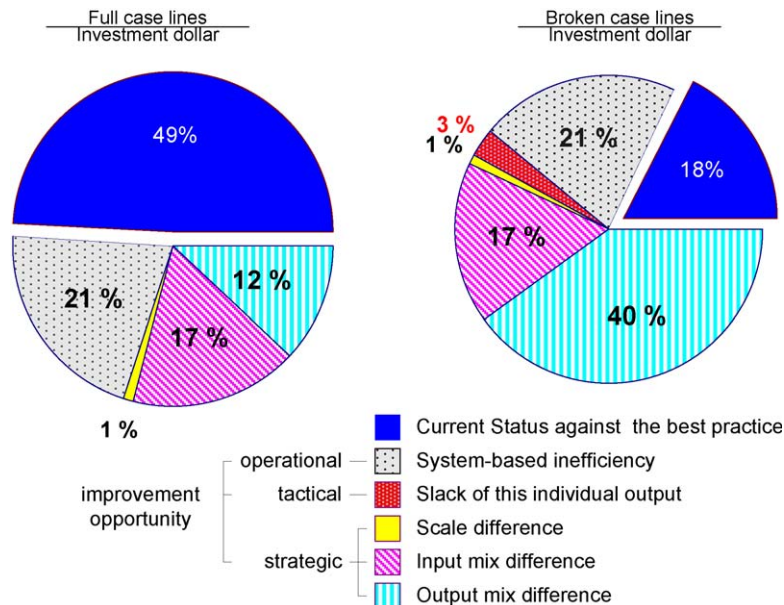

Fig. 3. Quantifying the inefficiency components: two examples of partial efficiency measures.

forming the ratio of warehouse i's partial productivity and the largest partial productivity value achieved by any warehouse in the database. Define BC_i as the number of broken case lines shipped in a year for warehouse i and L_i as the number of labor hours used in the same year for warehouse i, and I as the set of all warehouse observations in the comparison set. The broken case labor efficiency for warehouse i is calculated as

$$\frac{BC_i/L_i}{\max_I(BC_i/L_i)} \tag{2}$$

If there are n inputs and m outputs, there are $m \times n$ partial efficiency measures. Each of these measures is less than or equal to the warehouse's overall DEA efficiency measure. A gap analysis of the partial efficiency measures can identify the sources of inefficiency for each partial efficiency measure as shown in Fig. 3 (described in detail in Ref. [13]). Six sources of inefficiency are possible: operating inefficiency, scale inefficiency, labor hour slack, other resource substituting for labor hours, broken case lines shipped slack, and other output substitution for broken case lines shipped.

Table 1 gives a summary description of each component of inefficiency and the decision scope for improvement. By decomposing the inefficiency into its components, a manager can identify the particular factors limiting performance and the timeframe needed to implement the improvement activities.

Our Internet-based tool also shows information on efficient benchmark production processes that are similar to the inefficient processes in their mix of inputs and outputs. The manager of an inefficient production process can study how a similar process implements different practices to achieve more efficient production. In a warehousing context some examples of practices identified include: the use of velocity-based slotting, radio frequency (RF) dispatch, bar coding, automated sortation methods, and use of temporary labor.

Users can glean valuable information from analyzing the data set as a whole. Often managers are interested in specific practices or characteristics (attributes) that impact efficiency. Some examples of practices are given above, and examples of attributes include demand seasonality, demand variability, industry served, and target response time. These are characteristics of the environment or the conditions surrounding the production process which the manager in most situations cannot change or which would be very difficult to change. This analysis needs to be performed on the set of gathered data and thus is performed offline periodically. A two-stage, DEA and ordinary least squares approach as described by Banker and Natarajan [28] is used to identify correlations between computed efficiency levels and practices or attributes. These results can then be reported to the users via the Website, another favorable aspect of the Internet-based tool.

4. Pitfalls

This section discusses some pitfalls and issues associated with developing online benchmarking. They are important to resolve, and they also point to avenues for future research.

A critical step is to identify the benchmarking peer set. There are four considerations to address:

(1) Peers should use the same types of resources and produce the same types of outputs (although not all peers need to produce the same set of outputs);
(2) Peers should have access to the same technology;
(3) Identifying specific practices that may differ from one system to the next, but are controllable (e.g., use of particular technologies or operational methods is at the manager's discretion);
(4) Identifying system attributes which generally are not controllable, but which may affect performance (e.g., in the warehouse context, seasonality, demand volatility, or sku churn).

Security is a major concern, especially for proprietary information about a firm's financials and its operations. We therefore limit database access to only four individuals on the research team at Texas A&M University and Georgia Tech. When reference is made to benchmark production processes, it omits distinguishing characteristics such as firm name, location, and size. This allows the users to receive information about benchmark facilities without compromising the confidentiality of the submitted data.

While the Internet allows for the collection of data quickly and securely, data may be entered in error, or may not represent an actual facility. This increases the importance of designing effective data-filtering techniques which easily allow outliers to be identified [29,30]. Because DEA is a deterministic method, *any* error in data entry will affect the value of the identified best performance, even to the extent of misidentifying it.

Many industries have extensive terminology and it can often be challenging to communicate a common understanding of the data being requested and to describe the analysis being performed. We have learned that direct communication is the most effective way to clear up misunderstandings. In providing support for our tool, we have experimented with different support mechanisms, e.g., telephone and email. We find that telephoning requires more effort for the support organization, in our case the faculty members involved in the project. We have found that adequate service can be provided by email with the faculty member having the option to phone the user. Various chat interfaces are now in wide use and could also be used to provide support for this type of tool; however, they are support-intensive mechanisms. Our project contains a relatively large number of users compared to the support staff.
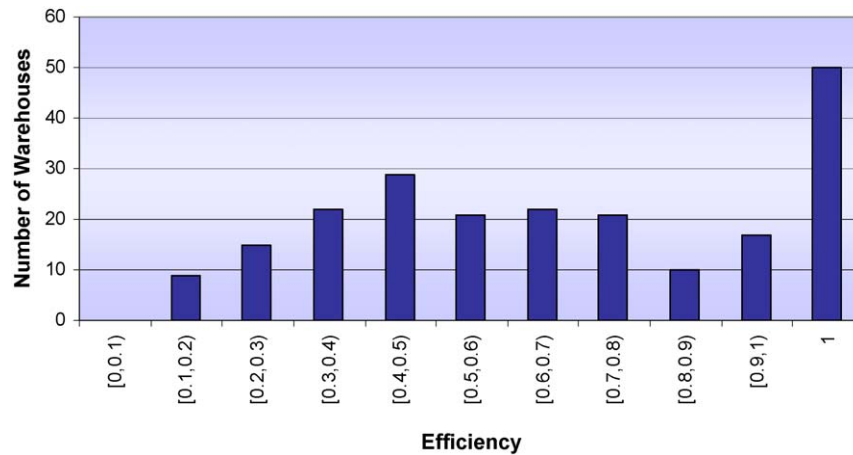
Table 1
A description and decision scope for each component of inefficiency for partial efficiency measures.

Factor	Description	Decision scope for improvement
Technical (operating) inefficiency	The gap to improve without changing current input–output structure and that can be achieved with better planning and execution	Short-term, operational
Resource slacks	Using the particular resource more than required.	Short-term, operational
Output slacks	Not fully utilizing inputs (capacity) to produce the specific output	Short-term, operational
Resource substitution	How much of the gap can be eliminated by reallocating resources (for this specific input)	Mid-term, tactical
Output substitution	The proportion of the gap due to the difference of output mix (for the specific output a firm selects as the productivity metric) against the best	Mid-term, tactical
Scale inefficiency	The performance gap due to the production level difference relative to the best	Long-term, strategic
Technology difference	The gap due to external environment (geography) or production technology	Long-term, strategic, or external environment

Fig. 4. Efficiency scores distribution.

Thus, it is crucial to design a user-friendly navigation and provide clear definitions of terminology. Moreover, DEA is developed on the basis of mathematics and economics and the analysis can be difficult to understand if the user is unfamiliar with these topics. The development of tutorials and explanations of the methodology are keys to wider adoption and use. The visualization and interpretation of analysis results are also desirable for interpreting the results.

5. Significance

Georgia Tech's Keck Virtual Factory Lab launched the Internet-based Data Envelopment Analysis System for Warehouses (iDEAs-W) project in 1999 to facilitate large-scale statistical benchmarking. The iDEAs-W Website is a component of the iDEAs project. The objectives are: (1) to demonstrate large-scale, Internet-based operational system performance assessment for a single industry, warehousing, and (2) to provide a cost efficient and useful benchmarking tool. This project is unique because it provides a fee-free, systems-based performance assessment via the Internet with a large data set used for comparison. It has been recognized by the warehousing industry as a useful tool as described in various trade publications [3,4,31,32]. This section summarizes the primary reasons for the success and describes other possible services.

Proper evaluation methods and information technology implementation do not automatically lead to the success of the online tool. One reason for the success of iDEAs-W is the availability of an initial data set. The data used in Ref. [26] provided an initial reference set, allowing the first users to receive real-time individual performance information. Without this dataset the first users would have only entered data and received no feedback, until a comparison set had been developed. In other words, the benchmarking application was immediately functional. Sometimes tools developed by academics fail to gain acceptance because user group is not clearly identified and the tool is not advertised. iDEAs-W has benefited from the support of the Logistics Execution Systems Association (LESA) of the Material Handling Industries of America (MHIA). MHIA has sponsored an information booth at five national material handling trade shows (ProMat 2001, NAMHS 2002, ProMat 2003, NAMHS 2004, and ProMat 2005). These trade shows have provided access and "face validity" for potential users.

As of April 2006, 390 warehouses have completed input and output data. After applying outlier detection methods, 216 warehouses were used in this study. For each warehouse we calculate an efficiency estimate based on the reported inputs and outputs. Fig. 4 shows the distribution of these efficiency estimates. About 21% of the warehouses have an efficiency estimate of less than 0.4, an indication that a large group of warehouses are relatively inefficient when compared to the others in the database. In addition 23% of the warehouses have efficiency estimates of 1.0. These are the warehouses that are used to identify best performance. The remaining warehouses, almost half, show considerable room for improvement.

Data is collected on a wide variety of attributes and practices. An attribute found to have a negative correlation with the efficiency estimates is seasonality. Warehouses with higher efficiency have, on average, a lower seasonality. Because resources are not completely flexible, high seasonality will cause some unavoidable negative effects on performance. Velocity-based slotting is a practice that positively correlated with high levels of performance, i.e. warehouses that slot according to the frequency of retrieval have, on average, a higher performance. For further information about the analysis of the entire database see Ref. [33].

Benchmarking services for a specific user group are possible, and may provide more precise and meaningful improvement guidelines. For example, iDEAs-W-BISG is an industry-specific version of iDEAs-W that was developed for the Book Industry Study Group (BISG). It is a multi-year collaboration between the Keck Virtual Factory Lab at Georgia Tech, BISG, and Texas A&M University. In the first iteration warehouse performance for fiscal year 2005 was analyzed, and in the second iteration 2006 data was analyzed. By tailoring the iDEAs-W tool to BISG more than twice as much data was collected for each warehouse. The book industry was interested in the effects of specific technologies, i.e. bar coding and radio frequency identification (RFID) on operational performance. A significant positive difference was observed between warehouses that used bar coding and warehouses that did not. The percentage of warehouses using RFID technology was too small to identify a statistically significant difference in performance. A detailed assessment of the current practices of the industry and an investigation of the correlations between efficiency and warehouse attributes and warehouse practices was performed in the offline analysis. The BISG found the information beneficial in identifying improvement strategies.

Analyses for specific groups give better visibility to the potential users and provide more precise suggestions. However, as addressed in the beginning of this section, the availability of a large peer group and an initial data set is critical. A more general-purpose tool and analyses are suggested if a connection to a specific peer group is not available.

6. Conclusions

This paper has demonstrated the feasibility and effectiveness of using Internet-based technology and advanced performance

assessment techniques to evaluate industry performance. Online benchmarking mitigates the difficulty in collecting data, reaches a wider audience, and maintains data security. We have discussed an example the warehousing industry, but similar performance measurement concerns arise elsewhere. Basic input/output models for other industries can be identified, with situation-specific variables describing the attributes and practice, and current methodology and the Internet implementation can be modified to fit these other applications.

Online analysis allows managers to identify improvement strategies based on the practices of the benchmark warehouses, and offline analysis helps them to identify the industry trends regarding attributes that can result in more efficient design of warehouses. As global competition increases, online bench-marking will become more important in identifying best practice behaviors and sharing information about improvement strategies.

Online services will also allow academics and practitioners to interact on a larger scale than otherwise possible. An electronic platform gives academics the opportunity to test algorithmic or methodological improvements and the calculation speeds make it simple to adjust the models. These advantages hold true as well for industry practitioners and ultimately the end-users.

Acknowledgements

References

[1] R.C. Camp, Benchmarking: The Search for Industrial Best Practices that Lead to Superior Performance, Quality Resources and ASQC Quality Press, New York, NY/Milwaukee, WI, 1989.
[2] W.M. Bulkeley, Business, Agencies to Standardize their Benchmarks, Wall Street Journal, New York, 2004, p. 2.
[3] Anon, On-line assessment tool for better warehousing, The Distributors and Wholesalers Advisor (2001).
[4] Anon, How good is warehouse performance? Modern Materials Handling (2002).
[5] P. Radgen, Comparing efficiency: Internet-based benchmarking for compressed air systems, Proceedings ACEEE Summer Study on Energy Efficiency in Industry 3 (2005) 119–129.
[6] E. Thanassoulis, M.C.S. Portela, O. Despic, Data envelopment analysis: the mathematical programming approach to efficiency analysis, in: H.O. Fried, C.A.K. Lovell, S.S. Schmidt (Eds.), The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press, 2008.
[7] J.W. Kendrick, Understanding Productivity: An Introduction to the Dynamics of Productivity Change, Johns Hopkins University Press, Baltimore, 1977.
[8] J.A. Tompkins, J.A. White, Y. Bozer, J.M.A. Tanchoco, Facilities Planning, John Wiley & Sons, Inc., New York, 2003.
[9] L. Berrah, V. Cliville, Towards an aggregation performance measurement system model in a supply chain context, Computers in Industry 58 (2007) 709–719.
[10] W.A. Tan, W. Shen, J. Zhao, A methodology for dynamic enterprise process performance evaluation, Computers in Industry 58 (2007) 474–485.
[11] A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, European Journal of Operational Research 2 (1978) 429–444.
[12] H.O. Fried, C.A.K. Lovell, S.S. Schmidt, The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press, 2008.
[13] W.-C. Chen, L.F. McGinnis, Reconciling ratio analysis and DEA as performance assessment tools, European Journal of Operational Research 178 (1) (2007) 277–291.
[14] W.W. Cooper, L.M. Seiford, K. Tone, Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Kluwer Academic Publishers, Boston, 2000.
[15] G. Tavares, A bibliography of data envelopment analysis (1978–2001), in: Rutcor Research Report, Rutgers Center for Operations Research, Rutgers University, Piscataway, NJ, 2002.
[16] A. Soteriou, S.A. Zenios, Operations, quality, and profitability in the provision of banking services, Management Science 45 (9) (1999) 1221–1238.
[17] J.D. Cummins, M.A. Weiss, H.M. Zi, Organizational form and efficiency: The coexistence of stock and mutual property-liability insurers, Management Science 45 (9) (1999) 1254–1269.
[18] S. Grosskopf, K.J. Hayes, L.L. Taylor, W.L. Weber, Anticipating the consequences of school reform: a new use of DEA, Management Science 45 (4) (1999) 608–620.
[19] T. Sueyoshi, Measuring efficiencies and returns to scale of Nippon Telegraph & Telephone in production and cost analyses, Management Science 43 (6) (1997) 779–796.
[20] R.D. Banker, S.A. Slaughter, A field study of scale economies in software maintenance, Management Science 43 (12) (1997) 1709–1725.
[21] P. Bogetoft, K. Nielsen, Internet based benchmarking, Group Decision and Negotiation 14 (3) (2005) 195–215.
[22] H. Ma, Z. Deng, W.D. Solvang, An online approach for distributor benchmarking, Benchmarking 11 (4) (2004) 385–402.
[23] L. Simar, P.W. Wilson, Statistical inference in nonparametric frontier models: recent developments and perspectives, in: H.O. Fried, C.A.K. Lovell, S.S. Schmidt (Eds.), The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press, 2008.
[24] R. Fare, S. Grosskopf, C.A.K. Lovell, C. Pasurka, Multilateral productivity comparisons when some outputs are undesirable—a nonparametric approach, Review of Economics and Statistics 71 (1) (1989) 90–98.
[25] T. Kuosmanen, A.L. Johnson, Data Envelopment Analysis as Nonparametric Least Squares Regression, Operations Research 58 (1) (2010) 10–21.
[26] S.T. Hackman, E.H. Frazelle, P.M. Griffin, S.O. Griffin, D.A. Vlasta, Benchmarking warehousing and distribution operations: an input–output approach, Journal of Productivity Analysis 16 (2001) 79–100.
[27] R.D. Banker, A. Charnes, W.W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management Science 30 (9) (1984) 1078–1092.
[28] R.D. Banker, R. Natarajan, Evaluating contextual variables affecting productivity using data envelopment analysis, Operations Research 56 (1) (2008) 48–58.
[29] W.-C. Chen, A.L. Johnson, A unified model for detecting efficient and inefficient outliers in data envelopment analysis, Computers & Operations Research 37 (2) (2010) 417–425.
[30] A.L. Johnson, L.F. McGinnis, Outlier detection in two-stage semiparametric DEA models, European Journal of Operational Research 187 (2) (2008) 629–635.
[31] Anon, New online tool benchmarks your warehouse vs. others, Inventory Reduction Report 5 (2002).
[32] Anon, Online resource helps DCs benchmark performance, creates industry database, Distribution Center Management 37 (2002).
[33] L.F. McGinnis, A.L. Johnson, M. Villarreal, Benchmarking Warehouse Performance Study: Summary of Results for Data Collected Through April 2006 for Internet-based Data Envelopment Analysis for Warehousing, W.M. Keck Virtual Factory Laboratory, School of Industrial and Systems Engineering, Georgia Tech, 2006.
[35] L.F. McGinnis, Best of Breed Warehouse Performance Assessment, Council on Logistics Management Annual Conference, Chicago, IL, 2003.

Andrew L Johnson is an Assistant Professor in the Department of Industrial and Systems Engineering at Texas A&M University. He obtained his B.S. in Industrial and Systems Engineering from Virginia Tech and his M.S. and Ph.D. from the H. Milton Stewart School of Industrial and Systems Engineering from Georgia Tech. His research interests include productivity and efficiency measurement, warehouse design and operations, material handling and mechanism design. He is a member of the INFORMS, National Eagle Scout Association, and German Club of Virginia Tech. His email address is ajohnson@tamu.edu.

Wen-Chih Chen is an Assistant Professor at Department of Industrial Engineering and Management in National Chiao Tung University, Taiwan. He received the B.S. degree in Industrial Engineering from the National Tsing Hua University, Taiwan in 1995. He received his Ph.D. from School of Industrial and Systems Engineering at Georgia Institute of Technology, USA, in 2003. His current research interests focus on performance evaluation and decision making analysis in logistics and semiconductor manufacturing.

Leon F. McGinnis holds the Gwaltney Chair for Manufacturing Systems in the College of Engineering at The Georgia Institute of Technology, with faculty appointments in Industrial and Systems Engineering and Mechanical Engineering. His teaching and research address modeling, analysis and design of discrete event logistics systems, and he leads a number of inter-disciplinary and industry sponsored research initiatives. Professor McGinnis is a member of IEEE and INFORMS, and a Fellow of IIE.