The Use of Multi-Attribute Utility Theory to Determine the Overall Best-in-Class Performer in a Benchmarking Study

Authors:

Terry R. Collins, Ph.D., P.E.
Department of Industrial Engineering
Texas Tech University
Mail Stop 3061
Lubbock, Texas 79409
PH: (806) 742-3543
E-mail: terry.collins@coe.ttu.edu

(Corresponding Author)

Manuel D. Rossetti, Ph.D., P.E.
Department of Industrial Engineering
University of Arkansas
4207 Bell Engineering Center
Fayetteville, Arkansas 72701
E-mail: rossetti@uark.edu

Heather L. Nachtmann, Ph.D.
Department of Industrial Engineering
University of Arkansas
4207 Bell Engineering Center
Fayetteville, Arkansas 72701
E-mail: hln@uark.edu

James R. Oldham
Whirlpool, Inc.
6400 Jenny Lind Ave.
Fort Smith, Arkansas 72901
E-mail: james_r_oldham@whirlpool.com

The Use of Multi-Attribute Utility Theory to Determine the Overall Best-in-Class Performer in a Benchmarking Study

Keywords *Benchmarking, Performance Assessment, Multi-attribute Utility Theory, Sensitivity Analysis, Benchmarking Gap Analysis, Best-in-Class Performer*

Abstract *This paper investigates the application of multi-attribute utility theory (MAUT) to aid in the decision-making process when performing a benchmarking gap analysis. Multi-attribute utility theory was selected to identify the overall best-in-class performer for performance metrics involving inventory record accuracy within a public sector warehouse. A traditional benchmarking analysis was conducted on 21 industry warehouse participants to determine industry best practices for the six critical warehouse metrics of picking and inventory accuracy, storage speed, inventory and picking tolerance, and order cycle time. A gap analysis was performed on the critical metrics and the absolute best-in-class was used to measure performance gaps for each metric. The gap analysis results were then compared to the MAUT utility values, and a sensitivity analysis was performed to compare the two methods. The results indicate that an approach based on MAUT was advantageous in its ability to consider all critical metrics and define a best overall performer for these data. An approach based on MAUT allowed the assignment of priorities and analyzed the subjectivity for these decisions. MAUT added robustness to the decision-making process and provided a framework to identify one performer as best across all critical metrics.*

INTRODUCTION

For decades, practitioners in the public and private sector have adopted the benchmarking approach as a useful tool for performance and quality assessments. Landmark benchmarking studies have been performed and the results widely publicized over the years (Camp, 1989; Kolarik, 1995; McNamee, 1994; Yasin, 2002)). Benchmarking has many benefits to the organization; however, the data analysis aspect of the process is an area in need of further refinement. For example, how can it be proven that the best practices realized are actually the best? How can the relevance of best practices be assessed by an organization? And finally, what is the best method for determining the best practices?

A recent study discovered difficulties in determining the best-in-class performer because of dissimilar reporting statistics and varying analysis techniques (Roider, 2000). Another study finds that the adopting best practices are related to resource constraints, size of organization, and the comparability of data (Hinton, et.al., 2000). Classic benchmarking analysis tools of flow charts, matrix analysis, spider charts, and Z-charts “have no structured means to evaluate the data, characterize and measure performance gaps, and project future performance levels” (Barr & Seiford, 1996). Therefore, a benchmarking group must identify a correct data analysis tool to use. This research utilizes and validates the decision-based analysis tool of multi-attribute utility theory (MAUT) for the benchmarking gap analysis process. The results of these efforts are presented in this paper.

A warehouse benchmarking study in the public sector is used as a case study to compare the mathematical decision method of MAUT against traditional benchmarking practices. The case study first presents the traditional process benchmarking approach, which focuses on sampling other warehouses across multiple industries to identify best-in-class performance metrics for warehouse management. Next, a gap analysis is performed to compare general industry results to the public sectors' current operating procedures to identify opportunities for adaptation. The results of the gap analysis identify the best-in-class performer. The gap analysis results are compared with the MAUT technique to evaluate the applicability and fit of MAUT as a more quantitative decision making approach for benchmarking. Utility values and relative weights are assigned using the benchmarking data. Finally, sensitivity analysis is performed to determine the outcome effects of varying the relative weight values.

MULTI-ATTRIBUTE UTILITY THEORY

Multi-attribute utility theory (MAUT) provides a comprehensive set of quantitative and qualitative approaches to justify a decision between alternatives (Canada & Sullivan, 1989), such as identifying the best-in-class performer in a benchmarking study. A specific type of Multi-attribute decision theory MADT, called multi-attribute utility theory (MAUT), was evaluated for its applicability to benchmarking analysis. Utility theory takes into account a range of the consequences of a particular decision and the risks of this decision, just as probability theory does for uncertainty. MAUT was selected as a viable method for improving benchmarking analysis due to its relative ease of both formation and computation. The MAUT approach enables the decision maker to incorporate preference and value trade-offs for each metric and measure the relative importance of each (Keeney and Raiffa, 1993). While other MAUT studies have been performed, there still exists a need for documented applications of this type of analysis (Walls, 1995). Bordley (2001) describes the use of MAUT to perform gap analysis for service research. The resulting gap analysis discounts the gap between performance and expectations, providing more empirical inferences than conventional gap analysis (Bordley, 2001). It is expected that applying MAUT to benchmarking will yield the same benefits.

Multi-Attribute Utility Theory

The basic goal of multi-attribute utility theory (MAUT) is to substitute information with an arbitrary measure called utiles so that the information can be compared. The utile values range from a low of 0 to a high of 1, with intermediate values decided upon by the decision maker. The identified critical metrics are plotted on a graph from 0 (worst case) to 1 (best case). Then, a utility curve is plotted to model the subjective value of each outcome (Daellenbach, 1994).

The end result of MAUT is simply to maximize the combined utility value (Keeney & Raiffa, 1993). Each metric is assigned a utile value and is combined with other utile values to assess an aggregate utility value according to set mathematical procedures. These procedures are explained in detail in the next paragraph. MAUT allows the decision maker to develop reasonable preference criteria, determine which assumptions are most appropriate, and assess the resulting utility functions (Lindey, 1985).

Two types of multi-attribute utility theory are common in the current literature: additive and multiplicative utility theory. For *i* alternatives with *j* attributes, the additive utility model is expressed:

$$U(x_i) = \sum_{j=1}^{n} k_j * u_j(x_{ij}) \tag{1}$$

$$\sum_{j=1}^{n} k_j = 1.0 \tag{2}$$

where:

1. k_j is a relative weight factor of the *j*th attribute.
2. $u_j(x_{ij})$ is the utility of the outcome x_{ij} for the *j*th attribute.
3. All attributes are independent of each other.

For i alternatives with j attributes, the multiplicative utility model is expressed as:

$$U(x_{ij}) = \frac{\prod_{j=1}^{n}[K * k_j * u_j(x_{ij}) + 1] - 1}{\mathrm{K}} \tag{3}$$

$$\sum_{j=1}^{n} k_j \neq 1.0 \tag{4}$$

where:

1. k_j is a relative weight factor of the jth attribute.
2. $u_j(x_{ij})$ is the utility of the outcome x_{ij} for the jth attribute.
3. K is a scaling constant found by:

$$1 + K = \prod_{j=1}^{n}(1 + K * k_j) \tag{5}$$

and must be found iteratively.

4. All attributes are independent of one another; $-1 < K < 0$ implies utility independence.

For this research, additive utility theory (AUT) was chosen for the following two reasons: 1) AUT provides a more practical methodology due to easier computational analysis, and 2) AUT is easier to understand and explain to decision makers relative to multiplicative utility theory. AUT allows the benchmarking party to assign priorities to certain metrics and allows stratification of all critical metrics. Also, AUT can be applied using common spreadsheet software, which is readily available in most business settings. No components of the formulation require complex iterative solutions. This analysis method uses subjectivity in formulating the relative weight factors (k_i), which therefore requires sensitivity analysis to be conducted to ensure the robustness of the assessment.

Sensitivity Analysis and Additive Utility Theory

Additive utility theory requires personal subjectivity. Because of this, extensive and thorough sensitivity analysis is necessary for justifying the end objective scores. The purpose of this analysis is to determine how sensitive the outcome is to changes in the variable values. This step is crucial because small changes in assigned values could produce very different results. Sensitivity analysis identifies these small changes and allows the decision maker to decide if the values need to be adjusted. Also, sensitivity analysis can identify user bias and help the decision maker to re-evaluate the original criteria used (Daellenbach, 1994).

A least squares method of sensitivity analysis for additive utility theory was developed by Barron and Schmidt (1988). For known single attribute value functions $u_j(x_{ij})$, this method computes, for two independent alternatives, new k_j (noted as w_j in the equation) values required to make the total utility value of alternative x_i exceed the total utility value of alternative x_b by an amount Δ, whose value is decided upon by the researcher. The least squares method is expressed as:

$$\text{minimize} \sum_{j=1}^{n} (w_j - k_j)^2 \tag{6}$$

subject to:

1. $$\sum_{j=1}^{n} w_j a_j = \Delta \tag{7}$$

2. $$\sum_{j=1}^{n} w_j = 1$$

3. $$w_j \geq 0 \tag{8}$$

4. $$a_j = u_j(x_{ij}) - u_j(x_{bj}) \tag{9}$$

5. x_b is the best-in-class performer discovered for initial relative weights b_j; $i \neq j$. (Barron and Schmidt, 1988)

After the w_j values are found, sensitivity analysis can be performed. For example, if attribute *A* is deemed twice as important as attribute *B*, all w_j values violating this can be ruled out. Also, varying levels of Δ in the above formulation allows sensitivity analysis for the implied relationship between alternatives *i* and *b*, and the corresponding effect on all other alternatives (Barron & Schmidt, 1988).

In essence, a formal MAUT analysis forces the benchmarking party to clearly define its priorities and measure the attractiveness of a discovered best practice. This is especially crucial in benchmarking studies, as the effects of the study are far-reaching throughout the organization (Forger, 1998). As benchmarking studies continue to become more complex, traditional benchmarking tools do not apply to new research (Ammons, 1999), and the need for more powerful benchmarking techniques increases.

RESEARCH METHODOLOGY AND PROCEDURES

The research methodology is broken down into a two-stage process. First, existing case study data will be presented from the warehouse benchmarking study. Next, a MAUT comparison will be conducted using the gap analysis results of the benchmarking study. In each stage of the research process there are several steps or procedures. Steps 1-4 are related to the case study, while the remaining steps are procedures for the MAUT comparison. These procedures are:

1) Identify the goals and objectives for the benchmarking study.
2) Select critical benchmarking metrics using multiple criteria.
3) Perform a gap analysis on the benchmarking metrics.
4) Define gap analysis recommendations from the benchmarking study
5) Assign utility values to the critical metrics
6) Calculate and assign the relative weights for the selected benchmarking metrics.
7) Perform a sensitivity analysis on the utility values.
8) Analyze the results of the comparison.

Case Study Application

To evaluate the use of MAUT as a decision-making tool for benchmarking, a case study is presented using data from a warehouse study. The warehouse study used benchmarking to investigate warehouse best practices in the public and private sectors. The case study provides information on the goals of the benchmarking study and how the critical benchmarking metrics are selected. The traditional gap analysis approach and recommendations are presented to identify the benchmarking best practices.

- **Benchmarking Goals**

The benchmarking study compared a warehouse's inventory integrity procedure to that of their competition. The purpose of the study was to provide recommendations for improving record accuracy, identifying policies for physical inventories, and methods to sustain inventory integrity. The areas of particular interest for the sponsoring warehouse were as follows:

a) How does industry set tolerance levels for inventory accuracy reporting? That is, how much of an error in reported inventory levels is acceptable for recording inventory performance?

b) How does industry handle errors during a picking operation? That is, when an order is being filled, what actions are taken if an error is discovered during this process?

c) Does industry perform cycle counting or 100% wall-to-wall inventories? That is, does the warehouse ever conduct regular interval counts on particular items or check every single item in the entire warehouse? How often does this occur, and what methods are employed to achieve this?

d) What is the highest inventory and picking accuracy rates that can be expected? That is, assuming best practices are implemented, what are the predicted accuracy targets?

- **Identification of Critical Metrics**

The identification of the critical basic metrics is essential to successful benchmarking. If the improper metrics are chosen, the end result may be useless. In this study, warehouse square footage, number of employees, dollar value of material handling equipment, and types of items handled were reported. In addition, specific warehouse accuracy indicators such as inspection and order accuracy were sampled. The sponsoring warehouses identified the following critical metrics for the case study (Frazelle & Hackman, 1994):

a) *Picking accuracy* is defined as the number of correct picks performed divided by the total number of pick performed. That is, what percentage of the time is the picker able to select both the correct stock keeping unit (SKU) and the correct quantity?

b) *Inventory accuracy* is defined as the number of items found in its correct location and quantity when conducting an inventory. That is, what percentage of the time is the location on the shelf identical to the inventory record?

c) *Storage time* is defined as the time required placing new stock to a specific location in the warehouse. That is, how long does it take to move material from the loading dock to its stocking location?

d) *Order cycle time* is defined as the time required to complete an order once picking begins. That is, how long does it take to ship an order once the picking process began?

- **Gap Analysis and Recommendations**

A narrative style of gap analysis is used as described by Keehley et. al. (1997). This method consists of three parts: statement of the question, identification of the gap in the procedures, and recommendations for closing the gap. In the gap analysis section, the rational behind each question and the critical metric used for the analysis is described. The methodology was successful in developing recommendations for the specific areas of interest.

Based on the gap analysis, answers are provided to the original four key questions posed at the initiation of the study were obtained as follows:

a) How does industry set tolerance levels for inventory accuracy reporting?

The best-in-class performer for inventory accuracy doesn't use tolerance levels for reporting purposes.

b) How does industry handle errors during a picking operation?

The best-in-class performer for picking accuracy checked nearby locations and triggered an inventory to be taken. Also, a second party is sent to re-check the error.

c) Does industry perform cycle counting or 100% wall-to-wall inventories?

The best-in-class performer for inventory accuracy used control group and activity-based cycle counting. The use of 100% wall-to-wall inventories and its effect on inventory integrity is inconclusive.

d) What is the highest inventory and picking accuracy rates that can be expected?

The absolute best-in-class performers achieved a reported 99.999% inventory accuracy and 99.9% picking accuracy. This level is achieved by using cycle counting, radio frequency identification (RFID) for picking and storing operations, and a computer system that monitored all warehouse policies.

MAUT Comparison

- **Assigning Utility Values**

To assign utility values, a judgment must be made on what performance level is assigned its associated utility. The research team worked closely with experts associated with the study to assign the utility values. It should be noted that most benchmarking studies would deal with several groups by varying utility values for each metric. This research is not directly concerned with combining group utility values, but primarily examines the difference utility assignment has on the identification of the best-in-class for this data set. For case studies regarding group utility values, see Eyrich (1991), and Korpela and Tuominen (1996).

Tables I and II show the utility assignments used for picking and inventory accuracy utility assignments. To allow all pertinent data to be evaluated, the lower bounds of the utility curve had to be set. The lowest possible performance for both picking and inventory accuracy, 0.0%, is assigned a 0 utility value. This allows all responses for a particular question to receive a utility value for each metric regardless of other responses. However, the same method does not work for storage and order cycle time lower bound assignments. After multiple discussions, it was determined that any times equal to 120 hours or more should be assigned a 0 utility value. This upper bound is chosen for its application in the sponsoring organization's own warehouse procedures. Cycle times above 120 hours are unsuitable for this operation. For inventory and picking tolerances, it was determined that any tolerance level of 5% or more should be given a utility value of 0. This limit is chosen due to the fact that a decrease in 5% of actual accuracy severely affected best-in-class identification and the validity of the best practices identified.

[Take in Table I]

[Take in Table II]

After calculating the lower bounds, the upper bounds are calculated. The theoretical best-in-class is designated as a 100% level for picking and inventory accuracies and 0 time units for storage and order cycle times. Although these values are virtually unattainable, they provide an upper bound for these metrics. Also, inventory and picking tolerances set at 0% are considered best-in-class. Therefore, inventory and picking accuracies are not assumed under any tolerance and are interpreted as absolutes. To assign intermediate utility values, judgment on the relative difficulty of increasing accuracy or decreasing time must be determined. For example, an increase from 90% to 95% may be easier to achieve than an increase from 99% to 99.5%. A close breakdown of each basic metric will identify which levels of performance mark an increase in capacity and thus a higher utility value. To create the utility curves, the performance criteria for picking and inventory accuracies were analyzed first. It was decided that accuracy values increase rapidly from 95% to 100%, with increases in this range doubling every ½%. The graph becomes asymptotic (relative to 1 utile value) as the accuracy approaches 100%. It should be noted that the configurations around the extreme values for these curves is debatable. The utility values as they approach the highest level could suggest less utility gain. For example, if someone received a donation of $100, then received another $100, their utility value would be favorable. However, if someone received a donation of $100,000, then $100 more, their utility value would not be as favorable as in the first condition. This scenario existed in the utility curves for accuracy levels and cycle times. The expected utility gain in these curves continues to be high, even as inventory accuracy increases from 99.99% to 100.00%. However, the formulated curve represented the preliminary reasoning that examined how difficult change above a certain level became.

Next, the sponsoring organization and research team analyzed the utility assignments for storage and order cycle times. The utility assignments used for storage and order cycle times are presented in Tables III and IV. It is decided that any times less than one complete day (24 hours) or less would determine the bounds from a utile value of 0.2. This is chosen to allow time ranging from 8 hours to 5 hours to have a marked improvement in utility assignment. This line would again break at 2 hours, allowing even more value to be assigned to shorter cycle times.

Finally, the utility assignments to use for the tolerance levels used in reporting accuracies were analyzed. Tables V and VI show the resulting utility assignments used for inventory and picking tolerances. As stated earlier, the lower bound for tolerance limits is assigned as 5%. It is decided to construct the utility curve as a linear relationship between the best possible (0%) and worst possible (5%). This decision is made to allow an equal penalty to be assigned as inventory tolerance increased.

The utility curve could now be plotted and all associated utility values for each response can be calculated. Because the methods for fitting curves to the utility assignments are most commonly done by hand, it was decided to use a linear relationship between each pair of data points to set intermediate utility values. Also, this linear relationship allows the easiest interpolation of intermediate metric values. Figures I-VI show graphical representations of the utility curves used. Each predetermined point is shown on the curve.

[Take in Figure I]

[Take in Figure II]

[Take in Figure III]

[Take in Figure IV]

[Take in Figure V]

[Take in Figure VI]

- **Calculation of Relative Weights**

To evaluate the total utility for each warehouse, all six metrics must be compared. However, one metric may be favorable over others due to the design of the survey and from the responses acquired. The following list outlines the reasoning for choosing a relative ranking scale: 1) Which metrics are more important: picking accuracy, inventory accuracy, storage time, or order cycle time? How does inventory and picking tolerances relate to these metrics?, and 2) Why is one metric more important and how much more important (in terms of k_j values)?

The selected relative weights can deal with these problems and help justify the identification of the best-in-class across all metrics, which adds robustness to the results. The relative weights are used in calculating the total utility for each participant. The sponsoring organization and the research team discussed the priorities of the current warehouse policies. It is deemed that the metrics should be arranged in the following order of importance: 1) Picking accuracy, 2)Inventory accuracy, 3) Storage time, 4) Order cycle time, and 5) Inventory and picking tolerances.

From this original examination, it is determined that picking and inventory accuracy rates identified best-in-class more than storage and order cycle times and tolerance values. Therefore, the following relative weight assignments are formulated. Table VII shows the initial relative weights assigned to this data.

[Take in Table VII]

- **Sensitivity Analysis**

A sensitivity analysis is then performed to justify the relative weights, which are ultimately affecting the identification of the best-in-class performers. With the initial utility value for each participant calculated, sensitivity analysis is conducted for each relative weight used in this calculation. Each weight factor is altered to identify the critical weights that will change the identity of the best-in-class performers. This step is crucial in the analysis of each weight factor to ensure a top performer is not eliminated or created by marginal changes in each relative weight.

To perform sensitivity, the least squares method described by Barron and Schmidt (1988) in Equations 6-9 is used. The best-in-class participant (x_b) is identified and used for the associated calculation. A pair-wise comparison is then made to each participant to calculate the associated k_j values necessary to have equivalent total utility values ($\Delta = 0$). Then, the cumulative utility values for all participants are calculated to discover if best-in-class identification has changed. Once tested for sensitivity and justified weight factors are found, the best overall performer is identified and recommendations are gathered from this respondent.

- **Data Analysis of MAUT Results**

After sensitivity analysis, the participant with the highest combined utility value is identified. Then, their associated responses are analyzed. The specific operating procedures for the entire warehouse accuracy process are evaluated and recorded. Gap analysis is performed from this data to the home processes to develop recommendations for improvements. Once this gap analysis is complete, the recommendations realized through MAUT are compared to the previous analysis.

- **Comparison of Results**

Using the previously obtained data, a pair-wise comparison is made to identify the different suggestions made for each question. A comparison is made between what suggestions the original study provided compared to the suggestions provided by MAUT. For each question, a pair-wise comparison is made to identify the different suggestions realized for the four critical questions posed during the previous research: 1) How does industry set tolerance levels for inventory accuracy reporting?, 2) How does industry handle errors during a picking operation?, 3) Does industry perform cycle counting or 100% wall-to-wall inventories?, and 4) What is the typical inventory and picking accuracy rate that can be expected? It is anticipated that the suggestions identified through MAUT are similar for some questions, while others may be very different. However, the careful consideration of applying MAUT is to add robustness to the decision criteria that identified these critical metrics. The purpose of this analysis is to examine the additional information MAUT provided for this data set.

DISCUSSION OF RESULTS

- **Case Study Data**

After completing the questionnaires, the twenty-one participants' responses to the basic warehouse metrics were recorded. Table VIII shows the raw data recorded for these twenty-one warehouses. With the data recorded, a quick comparison was warranted to ensure the previously formulated utility assignments were valid for this data set. Picking accuracy ranged from a low of 85.000% to a high of 99.999%, with three participants not recording this metric. Inventory accuracy ranged from a low of 80.000% to a high of 99.900%, with four participants not recording this metric. Storage speed ranged from just under one hour up to 48 hours, with three warehouses reported varying storage times. Order cycle time ranged from just 10 minutes up to 120 hours, with three warehouses reported varying cycle times. Inventory accuracy tolerances ranged from 0.0% to 5.0%; the majority of warehouses did not use inventory tolerances for calculating accuracy. Picking tolerances ranged from 0.0% to 5.0%; again, the majority of warehouses did not use picking tolerances for calculating accuracy.

[Take in Table VIII]

From the original assignments presented in the methodology section, the research team felt the original utility assignments would work for this data set. The original ranges allowed both extremes of each metric to be evaluated using MAUT. All warehouses that did not record a particular metric or reported varying cycle times were automatically assigned a "0" utility value for that particular metric. The decision for assigning a "0" utility to these metrics was to allow each warehouse to be compared across all metrics. Although this potentially eliminated best practices, it was a focus of the research to identify methods that sustain predictable and repeatable accuracy and cycle times across several metrics. When formulating the original critical metrics list, it was required that all criteria be evaluated regardless of the value of other metrics. This was important in both the identification of critical metrics and sustaining utility independence of the data. The next step was to convert each warehouse metric into its new utility value and evaluate the data using the relative weight factors.

- **Utility Values and Relative Weights**

To calculate each utility value, the raw data presented in Section A was entered into an Excel spreadsheet. The utility curves for each metric were then used to calculate the proper utility assignment for each metric. Taking each response presented in Table IX, the data was plotted on its associated utility curve. Then, the value was mapped to the curve to assign the utility value to each performance metric. Table IX shows the individual utility assignments made from the original data.

[Take in Table IX]

After the utility values were assigned, the relative weights were multiplied by each of the corresponding metrics to arrive at a total additive utility value. The combined utility values were derived from Equation 1. Table X shows the results of the initial utility value assignments for these data and the relative ranking of each warehouse's combined utility.

From this analysis, Warehouse 10 had the highest additive utility value of 0.807. Warehouse 21 had the lowest additive utility of 0.050. This was due to the lack of reporting only inventory and picking tolerances at 0.0%; all other metrics were not recorded or varied widely. The next step was to use sensitivity analysis to check if slight variations in relative weight factors would affect the identification of best-in-class status.

- **Sensitivity Analysis**

Sensitivity analysis is performed according to the method presented by Barron and Schmidt (1988). This method calculated the required change in relative weight factors to make a particular warehouse's additive utility value equal to the original best-in-class' additive utility value using the least squares principle. Equation 6 was solved using the Excel solver program. Then, the utility data was re-evaluated using the new relative weight factors to produce the corresponding additive utility for all participants. Table XI shows the results of the sensitivity analysis, the relative weights calculated, and the corresponding warehouse deemed best-in-class for this set of relative weights. The best-in-class (BIC) column for each warehouse ID denotes the particular warehouse that was identified as best-in-class under the calculated w_j values.

[Take in Table XI]

This sensitivity allowed the calculation of new relative weights and evaluation of these new values. All calculated relative weights produced zeros in one or more attribute weight, which signified eliminating one or more metrics from the data set. The research team did not desire eliminating any of the original critical metrics, so any sensitivity results containing a 0 in any column was deemed unacceptable. However, the sensitivity analysis proved that marginal changes in the original relative weights would not alter the identification of best-in-class. Also, one warehouse was unable to converge to a solution, as all metrics were below that of the best-in-class performer. That is, the least squares procedure will not converge with this series; the best-in-class would always have a higher utility for any and all relative weights greater than or equal to zero. Finally, five warehouses calculated a least squares solution of 0.50 and 0.50 for inventory and picking tolerances, respectively. The result of this sensitivity analysis produced multiple best-in-class rankings, as most warehouses did not use tolerances. However, it was noted that these five warehouses all had metric values that were less than that of the best-in-class performer. Therefore, convergence was dependent only on the tolerance levels used, eliminating all other metrics. Therefore, the sensitivity analysis provided adequate proof that marginal changes in relative weights would not change the identification of the best overall performer.

- **Comparison Between Methodologies**

Both analysis types revealed the best performers did not use tolerance levels. This implied accuracy levels were accurate as given and no allowances were used to calculate accuracy levels. For errors that occurred during picking, the process of flagging the error by the actual picker and filling a partial order and shipping remaining items later or per the customer's instructions was identified. The analysis did not identify whether or not nearby locations were checked for reconciliation of the error. Both analysis techniques revealed that a second party was sent to re-check for the error. For cycle counting, the analysis identified random sample cycle counting and not conducting 100% wall-to-wall inventories. In contrast, the previous analysis identified different types of cycle counting, and 100% wall-to-wall inventories were conducted. Table XII shows the gap analysis for the four key questions posed in the original study.

[Take in Table XII]

CONCLUSIONS AND IMPLICATIONS FOR USE IN THE BENCHMARKING PROCESS

This research followed a comprehensive application of MAUT and sensitivity analysis to a benchmarking study. This analysis differed from the classis benchmarking approach where recommendations are made based on traditional benchmarking tools. Comparing the suggestions of each method showed some similarities as well as some differences. However because of the various benchmarking methods existing today, a thorough pros-and-cons approach should be taken to determine if MAUT is useful for a particular study.

- **Pros of Using MAUT**

Multi-attribute utility theory proved effective in establishing priorities of several critical metrics and provided a method to compare these metrics across several participants. The most powerful advantage to using MAUT for this research was its ability to consider all critical metrics and define a best overall performer for these data. Also, MAUT allowed the comparison of different types of data to be directly compared. For example, accuracy percentages and cycle times were converted into identical units, making comparison easier. In addition, the use of MAUT allowed further investigation of the data gathered and provided a different look at the best practices discovered. The data used in classical benchmarking methodologies was easily re-used for the MAUT analysis, providing even more information. Finally, MAUT provided a framework to identify one performer as best across all critical metrics. Because the questions sampled were affected by multiple metrics, this fact became critical for best practice identification.

- **Cons of Using MAUT**

One of the drawbacks to using MAUT to benchmarking was the decision on weighting factors, both for modeling the utility curves and calculating relative weights (k_j). Although this method was chosen for its ability to handle individual preference, it introduced the possibility of user bias. That is, the results were heavily dependent on these original values. If an industry-wide benchmarking study was conducted, there may be several discrepancies to which metric received more or less importance and how much these opinions differ. Another drawback to applying MAUT to these data was the lack of robust sensitivity calculations for new relative weights (w_j). Although the sensitivity analysis presented in this research showed marginal changes in relative weight factors would not alter the best-in-class status, the resulting zeros in every analysis line did not allow the consideration of any new possibilities. However, this is suspected to be a case-specific problem unique to this data set. One consequence in using MAUT is that the method identifies an overall best performer. Although this was critical for the scope of this study, other benchmarking efforts may benefit from spreading their best practice identification to absolute top performers for a chosen question. However, the breakdown of each critical metric and the question it directly affects simply provides more detail to the research; absolute best-in-class evaluation can still be achieved with this method.

- **Contributions**

In this research, the application of MAUT was analyzed for its applicability to benchmarking analysis. This research was successful in proving the following points for this comparative study:

1. MAUT provided stratification of all critical metrics chosen and allowed for direct comparison between them.
2. MAUT allowed the research team to assign priorities and analyze the subjectivity of these decisions.
3. MAUT provided a mathematical method for comparing trade-offs and identifying best-in-class.
4. MAUT added robustness to the decision criteria and was suspected to increase robustness as the amount of critical data increased.

A formal MAUT application to benchmarking is recommended for any party who requires a method to compare several metrics simultaneously. The MAUT model works best for small benchmarking efforts where the research team can clearly define priorities within the study. Also, the tests for sensitivity allow the benchmarking party to evaluate its choices for weight factors and to adjust them if necessary. It should be noted that the MAUT method finds a "best overall" performer for all critical metrics, rather than a "best-in-class" performer for each critical metric. Therefore, the proper method to employ depends on the end result desired. This research used the same data set for both analysis methods. MAUT could be used in addition to the previous methodology to gain even more insight on the information. Through the initial investigation of MAUT on benchmarking, some suggestions for future research were formulated at the conclusion of the study.

REFERENCES

Ammons , D. (1999), “A proper mentality for benchmarking,” *Public Administration Review*, Vol. 59, No. 2, pp. 105-109.

Barr, R. S. and Seiford, L.M. (1996), “*Benchmarking and Performance Improvement Tools for Manufacturing and Service Processes*,” presented at the National Science Foundation Design and Manufacturing Grantees Conference, Albuquerque, New Mexico.

Barron, H. and Schmidt, C. (1988), “Sensitivity analysis of additive multi-attribute value models,” *Operations Research*, vol. 36, no. 1, pp. 122-127, 1988.

Bordley R. F. (2001), “Integrating gap analysis and utility theory in service research,” *Journal of Service Research*, vol. 3, no. 4, pp. 300-309, 2001.

Camp, R.C. (1989), “Benchmarking: The search for industry best practices that lead to superior performance.” *ASQC Quality Press*, Milwaukee, MN.

Canada, J.R., and Sullivan, W. G. (1989), “*Economic and Multi-attribute Evaluation of Advanced Manufacturing Systems*,” Prentice Hall, New York, NY.

Daellenbach, H.G. (1994), *Systems and Decision Making*,” Wiley, New York, NY.

Eyrich, H.G. (1991), “Benchmarking to become the best in breed,” *Manufacturing Systems*, vol 9, no. 4.

Forger, G. (1998), “Benchmark your warehouse for future success,” *Modern Materials Handling*, vol. 53, no. 12, pp 39-41.

Frazelle, E. H., and Hackman, S. (1994), “The warehouse performance index: A single-point metric for benchmarking warehouse performance*,” Material Handling Research Center Final Report* #93-14, Georgia Tech.

Hinton, M, Francis, G, and Holloway, J. (2000), “Best practice benchmarking in the UK,” *International Journal of Benchmarking,* vol. 7, no. 1, pp. 52-62.

Keehley, P and MacBride, S. (1997), “*Benchmarking for Best Practices in the Public Sector*,” Jossey-Bass Publications, San Francisco, CA.

Keeney, R. L. and Raiffa, H. (1993), “*Decisions with Multiple Objectives: Preferences and Value Tradeoffs*,” Wiley, New York, NY.

Kolarik, W.J., (1995), “*Creating Quality: Concepts, Systems, Strategies, and Tools*,” McGraw-Hill, New York, NY.

Korpela, J. and Tuominen, M, (1996), “Benchmarking logistics performance with an application of the analytical hierarchy process,” *IEEE Transactions on Engineering Management*, vol.43, no. 1, pp. 323-333.

Lindey, D.V. (1985), “*Making Decisions*,” Wiley, New York,NY.

McNamee, D. (1994), “*Reinventing the Audit: Frameworks for Change*,” Mc2 Management Consulting, California.

Rodier, M.M. (2000), “A quest for best practices,” *IIE Solutions*, vol. 32, no. 2, pp. 36-39.

Walls, M.R. (1995), "Integrating business strategy and capital allocation: An application of multi-objective decision making," The Engineering Economist, vol. 40, pp. 247-266.

Yasin, M.M. (2002), "The theory and practice of benchmarking: Then and now," *International Journal of Benchmarking*, vol. 9, no. 3, pp. 217-234.