NISTIR 7911

Assessment of Real-Time Factory Performance Through the Application of Multi-Relationship Evaluation Design

Brian A. Weiss
John Horst
Fred Proctor

http://dx.doi.org/10.6028/NIST.IR.7911

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 7911

Assessment of Real-Time Factory Performance through the Application of Multi-Relationship Evaluation Design

Brian A. Weiss
John Horst
Fred Proctor
Intelligent Systems Division
Engineering Laboratory

http://dx.doi.org/10.6028/NIST.IR.7911

April 2013

U.S. Department of Commerce
Rebecca Blank, Acting Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Under Secretary of Commerce for Standards and Technology and Director

Assessment of Real-Time Factory Performance through the Application of Multi-Relationship Evaluation Design

Brian A. Weiss, John Horst, Fred Proctor

National Institute of Standards and Technology

100 Bureau Drive, MS 8230

Gaithersburg, Maryland 20899 USA

1. INTRODUCTION

Successful operations within manufacturing environments require both accurate and precise information flow from one operation to the next. Incorrect, too little, or too much information can slow the manufacturing process and/or result in poor quality output. The National Institute of Standards and Technology (NIST) is developing and testing new integrated information models for use in manufacturing and quality measurement equipment [2]. Numerous information types are required and distributed during the steps of the manufacturing and quality measurement process [3]. Experts from the manufacturing quality community have been drafted to design and refine the Quality Information Framework (QIF) to ensure quality is appropriately injected into the manufacturing process. QIF is defined as "a suite of integrated XML Schema based standards enabling the seamless flow of information within computer-aided quality measurement systems" [7]. QIF presents a range of information including quality measurement plans, measurement results, measurement rules, measurement resources, and analysis of results.

The predominant information that QIF analyzes are Key Performance Indicators (KPIs). "KPIs are defined as quantifiable and strategic measurements that reflect an organization's critical success factors" [3]. Simply stated, KPIs must be recognized and understood in order to assess and improve manufacturing performance. KPIs exist to increase the understanding of lean manufacturing (minimizing waste) and to realize a company's vision of accomplishing their strategic objectives. As defined in the ISO/DIS 22400 standard, over 30 KPIs are detailed that are categorized as either informing on Production, Maintenance, Inventory, or Quality (see Table 1). Example KPIs include Worker Efficiency, Allocation Ratio, and Throughput Ratio, which are calculated from quantitative data collected within the manufacturing environment. Given that there are over 60 pieces of quantitative data (e.g., Actual Personnel Work Time, or Actual Personnel Attendance Time) used to calculate the various KPIs (e.g., Worker Efficiency = Actual Personnel Work Time / Actual Personnel Attendance Time), it is easy for an operator to experience "information overload." Determining which KPIs are more important than others and the relative importance of the functional areas within a manufacturing facility (e.g., an inventory area v. an assembly line) are significant challenges that must be overcome.

Table 1 – KPIs defined in Manufacturing Operation Management

KPI	Production	Maintenance	Inventory	Quality
Worker efficiency	X			
Allocation ratio	X			
Throughput rate	X			
Allocation efficiency	X			
Utilisation efficiency	X			
Overall equipment effectiveness index	X			
Net equipment effectiveness index	X			
Availability	X			
Effectiveness	X			
Quality ratio				X
Set up rate	X			
Technical efficiency	X			
Production process rate	X			
Actual to planned scrap ratio				X
First pass yield				X
Scrap ratio				X
Rework ratio				X
Fall off ratio				X
Machine capability index	X			
Critical machine capability index	X			
Process capability index	X			
Critical process capability index	X			
Comprehensive energy consumption	X			
Inventory turns			X	
Finished goods ratio	X			
Integrated goods ratio	X			
Production loss ratio	X			
Storage and transportation loss ratio			X	
Other loss ratio			X	
Equipment load rate	X			
Mean operation time between failures		X		
Mean time to failure		X		
Mean time to restoration		X		
Corrective maintenance ratio		X		

This paper begins to apply the Multi-Relationship Evaluation Design (MRED) methodology to this manufacturing problem to detail a process that devises test plan blueprints to assess the overall performance of a manufacturing operations facility along with its constituent functional areas and critical physical elements [14] . MRED also provides evaluators with a means of capturing the relative importance of KPIs within a manufacturing environment. These blueprints are invaluable in that they can focus test plan development to verify and validate performance regardless of whether the manufacturing operations facility is still in development or fully-developed.

2. Real-Time Factory Information (RTFI) Background

Manufacturing Execution Systems (MES) are defined as “information technology systems that manage manufacturing operations in factories” [16]. The essence of real time factory information is embedded within MES. Specifically, MES is defined to include the following activities [16]:

- Management of product definitions
- Management of resources

- Scheduling (production processes)
- Dispatching production orders
- Execution of production orders
- Collection of production data
- Production performance analysis
- Production track and trace

The MES scope is presented in a five level, functional hierarchy shown in Figure 1 [3]. This hierarchy grew from the ANSI/ISA-95 standard merger of the Manufacturing Enterprise Solutions Association's (MESA) initial MES structure and the Purdue Reference Model (PRM) [16] [17]. The core of MES lies in level 3, while process controls lie at levels 0, 1, and 2. The level 3 activities are often called Manufacturing Operations Management Systems (MOMS) and can be broken down into four primary operations: production, quality, logistics, and maintenance (some regard logistics as being inventory).

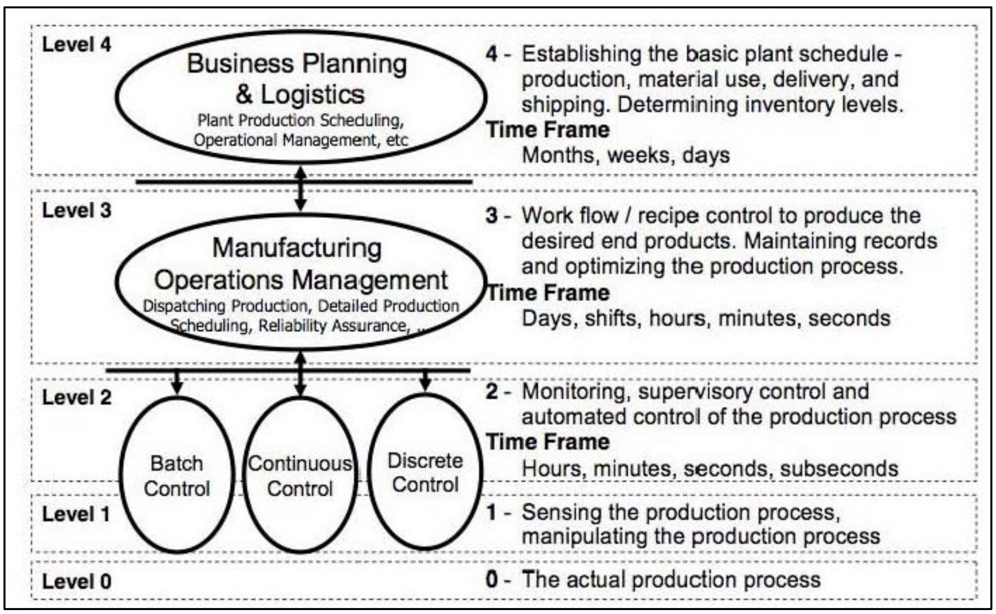

Figure 1: MES Functional Hierarchy

KPIs are typically produced at Level 3 and may be brought up to Level 4 for greater usage [3]. Information from levels 1 and 2 is often required to calculate the KPIs. KPIs are typically paired with company-defined thresholds to indicate when actions must be taken to improve performance, efficiency, and/or quality. When a KPI crosses a threshold, it may become necessary for a warning to be issued to a user/operator/manager or for an action to be performed. These KPIs are often shared among multiple technologies and processes that can contain multiple, proprietary formats that are not directly compatible. QIF is being developed to enable the seamless flow of quality information (e.g., KPIs) among multiple systems so information does not get "lost in translation." Given the magnitude of information to be shared and the understanding that a human operator consumes some of the information, it is critical to assess the priority of each KPI within an environment given its impact on the manufacturing process and the complexity of consumption by a human. In addition, the complexity of a manufacturing

environment, the magnitude of KPIs, and the technologies' maturity can make it challenging to discern which specific elements should be immediately assessed as compared to those whose assessment can wait. The MRED methodology presented in the following section presents a means to surmount these challenges.

3. MRED METHODOLOGY

MRED is an interactive algorithm that processes stakeholder-provided information from multiple input categories and generates one or more test plan blueprints that include their constituent test plan characteristics. MRED leverages the relationships among the inputs and the impacts the inputs have on the outputs to produce test plan blueprints. The overall model, including input and output, is shown in Figure 2. The model requires six different types of input in order for it to output one or more evaluation blueprints.

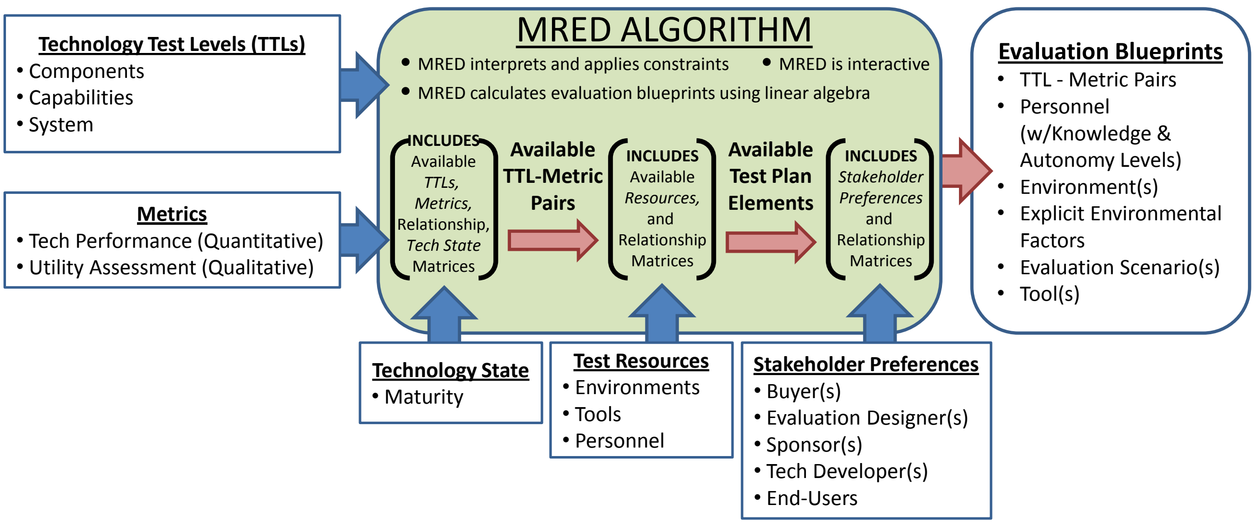

Figure 2: Overall MRED Model [14]

MRED includes 1) an interactive process to identify candidate test plan elements and eliminate those that are infeasible or unnecessary given relationships among these elements and 2) a method to capture and handle *Stakeholder Preferences* of evaluation elements while minimizing the load on the *Stakeholders* [14]. The following sub-sections highlight the MRED process, where extensive details can be found in [9] [10] [11] [12] [13] [14] [15].

3.1. Input Categories

MRED relies upon information, data, and preferences from the categories highlighted in Figure 2 (blue arrows). This section presents the critical inputs of MRED at a high level. Further details of these inputs can be found in [9] [10] [11] [12] [13] [14] [15].

3.1.1. Technology Test Levels (TTLs)

TTLs are defined as the technology's constituent *Components* and *Capabilities* along with the *System*, in its entirety [5]:

- *Component* – Essential part or feature of a *System* that contributes to the *System's* ability to accomplish a goal(s).
- *Capability* – A specific ability of a technology. A *Capability* is enabled by either a single *Component* or multiple *Components* working together.
- *System* – A group of cooperative or interdependent *Components* forming an integrated whole to accomplish a specific goal(s).

Figure 3 depicts a relationship diagram that highlights the technology dependence between *Components* and *Capabilities* where physical elements are necessary to produce functional elements [14]. Relationships among *Components* and *Capabilities* can be as simple as a one-to-one mapping (i.e., *Component 1* is solely responsible for producing *Capability 1* and no other *Capabilities*). Likewise, relationships can be as complex as multiple *Components* producing a single *Capability* (*Components 2, 3,* and *4* producing *Capability 2*) and/or as complex as a single *Component* contributing to multiple *Capabilities* (*Component 3* supporting the function of *Capabilities 2* and *3*).

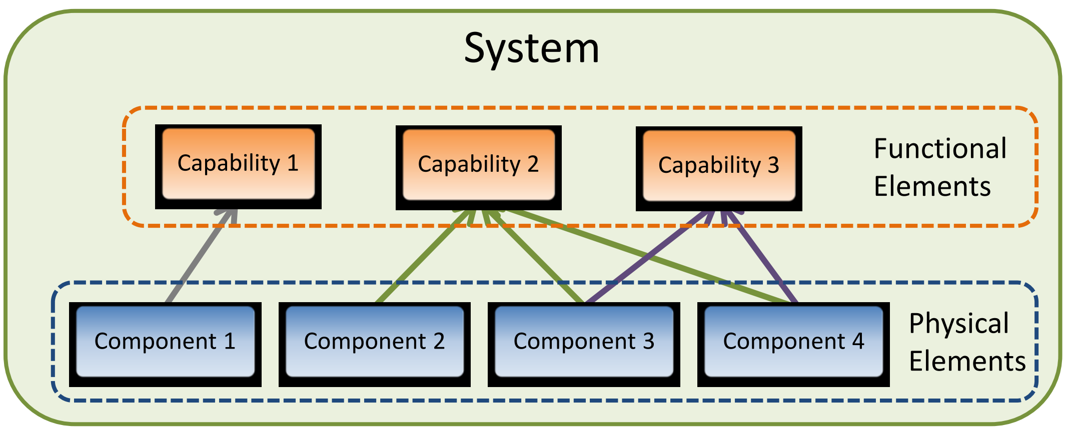

Figure 3: Relationships Between the Components and Capabilities of a Technology [14]

3.1.2. Metrics

Pertinent *Metrics* are also input according to the input TTLs. *Metrics* fall into one of two groups:

- *Technical Performance* – *Metrics* related to quantitative factors (e.g., accuracy, distance, or time)
- *Utility Assessments* – *Metrics* related to qualitative factors that express the condition or status of being useful and usable to the target user population. These metrics may be required by the program sponsor, to meet user expectations, inform the technology developers on their design, etc.

3.1.3. Technology State

"MRED defines *Technology State* as a technology's fitness for testing" [14]. *Technology State* is described by the factor of *Maturity* [12] [14]. *Maturity* is the fitness for operation of individual *Components*, *Capabilities*, and the *System*. A technology's *Maturity* has a direct impact on whether a specific TTL is ready for testing and what, if not all, *Metrics* may be reasonably captured given its current *Technology State*. As *Components* are integrated together to form the *System*, they enable specific *Capabilities*. Some *Capabilities* may be operational before the entire *System* is operational. Throughout the technology's development cycle, its *Maturity* is constantly updating. *Components* that have a *Maturity* value of fully-developed do not have any technological restrictions impeding the capture of all possible *Metrics*. Conversely, immature *Components* may be technologically-restricted from capturing all possible *Metrics*.

Maturity must be input into MRED for a *TTL* to be considered for testing. At any time during development, the *Maturity* of the *System*, its *Components*, and its *Capabilities* will fall into one of the following classes:

- *Immature* – The *TTL* has yet to be developed or is still in the process of being developed.
- *Fully-Developed* – The *TTL* is developed to the point of being operational and complete (no further development or revisions are expected).

The *Maturity* information for a technology's *Components* is typically gathered from the technology developers. These stakeholders are in the best position to provide this data since they are most familiar with the technology and have the most up-to-date information. The *Maturity* of *Capabilities* and the *System* is either provided by the technology developer or by MRED calculations [14]. It is important to note that Technology Readiness Level (TRL) definitions [4] [5] [6] are not relevant to MRED's concept of *Maturity*. While the TRLs are clearly defined for an entire technology (i.e., full system), the application of TRLs is not similarly consistent for a technology's constituent physical and functional elements. A *Component* cannot be tested at TRL-7 or above since these TRLs require a system prototype demonstration in the target environment. TRLs examine the *System* whereas MRED requires a means of assessing the *Maturity* of individual elements. Also, TRLs can only be reasonably assigned after a technology has either been judged by an individual or group as ready to move onto the next TRL (for the lower TRLs) or undergone a demonstration or evaluation in the corresponding conditions (for the higher TRLs). Otherwise, it is up to the *Stakeholders* to indicate the existing TRL of a technology (based upon prior demonstration) and make a judgment as to whether or not the *System* can be assessed at the next TRL. A significant benefit of MRED over TRLs is that MRED provides a concrete, documented path for proceeding from initial proposal through deployment, whereas the actual transition between TRLs is not defined (or even consistent).

3.1.4. Test Resources

This category of inputs signifies the availability of the viable *Environments*, *Tools*, and *Personnel*. They are defined as:

- *Environment* – The physical venue, supporting infrastructure, artifacts, and props that will support the test(s). The setting in which the assessment(s) take place can have a significant impact on the

data. The testing *Environment* can influence the behavior of the test personnel and limit which levels of a technology can be evaluated.

- *Tools* – The tools, equipment, and/or technology that will collect quantitative and/or qualitative data during the test [10] [14]. Tools also include the necessary hardware, software, and/or assessment personnel to produce the necessary *Metrics* from the captured data. *Tools* are split into those supporting the capture and/or generation of *Technical Performance* and *Utility Assessment Metrics*.
- *Personnel* – Individuals that will use the technology or indirectly interact with the technology. *Personnel* can be delineated as being either primary or secondary:
  - The primary *Personnel* directly interact with technologies, are classified as *Technology Users* (*Tech Users*) and are composed of three specific types of individuals: *End-Users, Trained Users,* and *Tech Developers*.
  - The secondary personnel are those that indirectly interact with the technology and fall into the following two categories: *Team Members* and *Participants*.

Test Resources have been greatly detailed in [9] [10] [14].

3.1.5. Stakeholder Preferences

This last category includes preferences from five specific stakeholder classifications [9] [14]:

- *Buyers* – Stakeholders purchasing the technology
- *Evaluation Designers* – Stakeholders creating the test plans including determining and identifying the MRED inputs
- *Sponsors* – Stakeholders paying for the technology development and/or evaluation
- *Technology Developers* – Stakeholders designing and constructing the technology
- *Users* – Stakeholders that will be or is already using the technology

Besides providing preferences on *Test Resources* (i.e., *Environment, Tools,* and *Personnel*), stakeholders also provide their preferences on the following [9] [10] [14] [15]:

- *TTL-Metric Pairs* - Defined as a specific TTL that is coupled with a *Metric* that can be generated from testing this specific TTL. The value of this output blueprint coupling is that many TTLs (if not all) can have more than one *Metric* captured during their assessment. Producing numerous *Metrics* from one TTL is advantageous since it likely reduces the testing cost per *Metric*.
- *Personnel – Knowledge Levels* – The *Personnel* involved in a test plan have varying levels of knowledge about the functionality and usage of the technology in addition to the testing environments. The scope of knowledge and their specific levels are defined by MRED for each test plan.
- *Personnel – Autonomy Levels* – *Personnel* are also assigned specific decision-making autonomy levels within a test plan.

- *Evaluation Scenarios* – The scenarios that oversee specifically what the technology will be exposed within the testing *Environment(s)*. The three types of evaluation scenarios defined within MRED are *Technology-based*, *Task/Activity-based*, and *Environment-based*.
- *Explicit Environmental Factors* – These factors are the characteristics present within the test environment that can affect the technology and influence the actions of the test *Personnel*. These factors relate to the inclusive environment, including *Participants*, structures, artifacts, etc. The two characteristics comprising *Explicit Environmental Factors* are *Feature Density* and *Feature Complexity*.

3.2. Output Test Plan Blueprints

Each set of blueprints defines one or more of the following:

- *TTL-Metric pairs*
- *Personnel* to act in primary and secondary (if necessary) roles
- *Knowledge* and *Autonomy Levels* for those *Personnel* who will directly and indirectly interact with the technology during the test
- *Environments* for testing
- *Evaluation Scenarios* describing the type of exercises in which the technology will be immersed
- *Explicit Environmental Factors* which indicate the levels of *Feature Complexity* and *Feature Density* within the *Environment*
- *Tools* to support the collection and analysis of data to generate the corresponding *Metrics.*

3.3. Key Relationships

Relationships are a core element to MRED and are defined among the various inputs and between inputs and outputs. Relationships defined between the inputs and outputs have been discussed extensively in previous work [9] [10] [11] [12]. The blueprint characteristics and their influences are shown below in Figure 4.

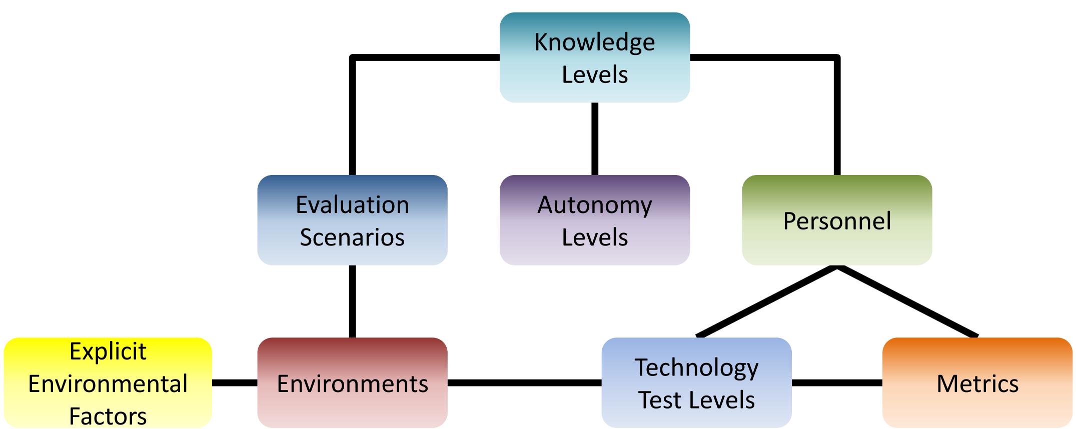

Figure 4: MRED Relationships among Test Plan Blueprint Characteristics [14]

Beside the relationships noted above, MRED also defines relationships within the *Technology Test Levels*. Specifically, relationships exist between *Components* and *Capabilities* that are highlighted in Section 3.1.1.

3.4. Filtering and Elimination

The MRED process, detailed in [14] [15], can be summarized in Figure 5 and Figure 6. The steps in Figure 5 are executed once the stakeholders have identified the *TTLs* for test consideration and the relevant *Metrics* along with the relationships between the *Components* and *Capabilities* and the relationships between the *TTLs* and the *Metrics*. The output of Figure 5 presents the feasible test plan characteristics and their corresponding relationships.

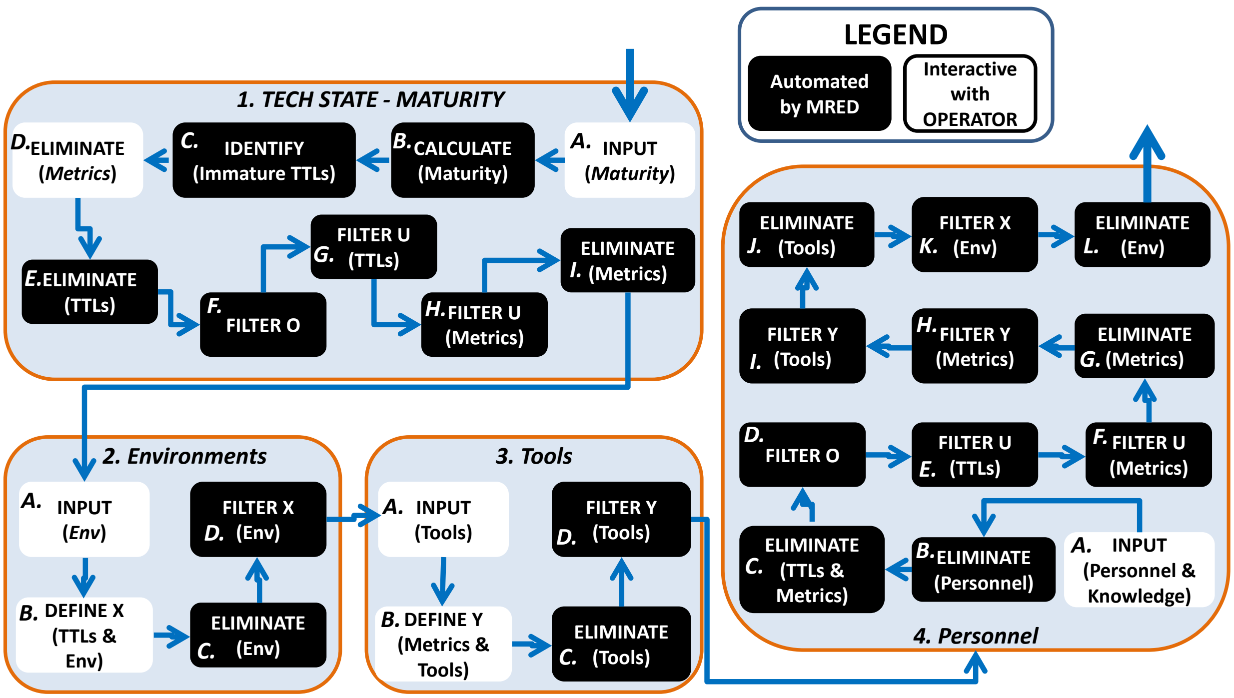

Figure 5: MRED Constraint Handling and Element Filtration Process [14]

3.5. Stakeholder Preference Capture and "Aggregation"

The execution of Figure 6 is based upon capturing stakeholder preferences of the indicated test plan characteristics on an 11-point scale using the method of Evaluative Voting [1] [13] [14]. Stakeholder preferences are captured on an ordinal, linguistic scale highlighted in Figure 7. This symmetric scale enables stakeholders to express their preferences linguistically as to their preference for or against the inclusion of an alternative or characteristic for test inclusion. Each linguistic preference corresponds to a value on a numerical scale enabling mean (as defined in Evaluative Voting) and standard deviation to be generated.

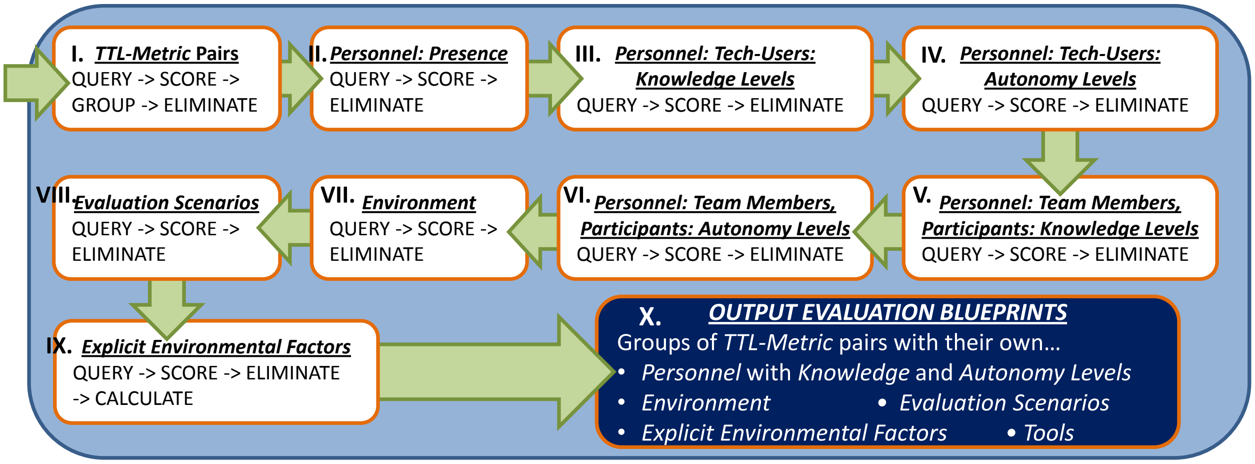

Figure 6: MRED Stakeholder Preference Capture and Handling [14]

Each box (I. through IX.) denoted in Figure 6 represents the capture (QUERY) of Stakeholder Preferences, the calculation of the means and standard deviations (SCORE), and the removal of the lowest scoring (ELIMINATE) characteristics. Further details of this process are presented in [14] [15].

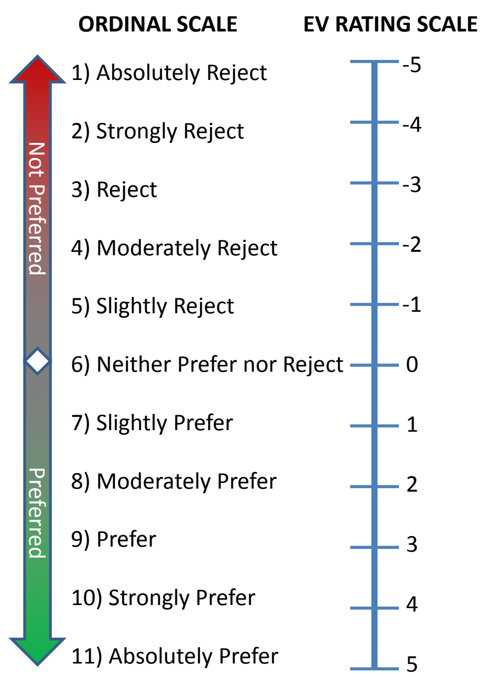

Figure 7: Example 11-point Scale to Capture Stakeholder Preferences for MRED [13]

3.6. MRED Benefits

MRED's blueprint generator enables the evaluation designer to formally define and document test plan blueprints including the documentation of stakeholder preferences. This offers significant advantages by allowing evaluation personnel the ability to:

- Formalize, and potentially standardize, the development of evaluation blueprints.
- Provide traceability of stakeholder preferences within a test plan, across multiple test plans and across multiple test events.
- Enable blueprints (i.e., one or more test characteristics) to be altered more rapidly while minimizing the burden on the stakeholders.

4. MRED APPLICATION TO RTFI

The following section explores the feasibility of applying MRED to real time factory information (RTFI). The objective of this approach is to invoke MRED's structured and scientific process to:

1. Model an abstract manufacturing environment such that test plan blueprints can be generated to detail standards implementation verification and/or validation tests
2. Capture stakeholder preferences indicating the most favored and/or critical test plan characteristics including KPIs
3. Chronicle stakeholder preferences at multiple points in time to track the evolution of favored KPIs
4. Predict the critical KPIs for operations inclusion based upon stakeholder preferences
5. Provide stakeholders with a structured manner in which to identify pertinent test characteristics to assess a manufacturing environment
6. Leverage stakeholder preferences regarding KPIs to highlight points of emphasis for the QIF

4.1. Technology Test Levels

In order to meaningfully apply MRED to RTFI, it is critical to verify the alignment of MRED's hierarchical structure (Components -> Capabilities -> System) with a corresponding structure in a manufacturing environments. Fortunately, the IEC 62264 standard for Manufacturing Operations Management defines a hierarchical structure depicted in Figure 8 [3]. Note the similarities of Figure 3 to Figure 8. It is also important to note that the MRED terms *System*, *Functional Elements*, and *Physical Elements* are included to further highlight the correlation.

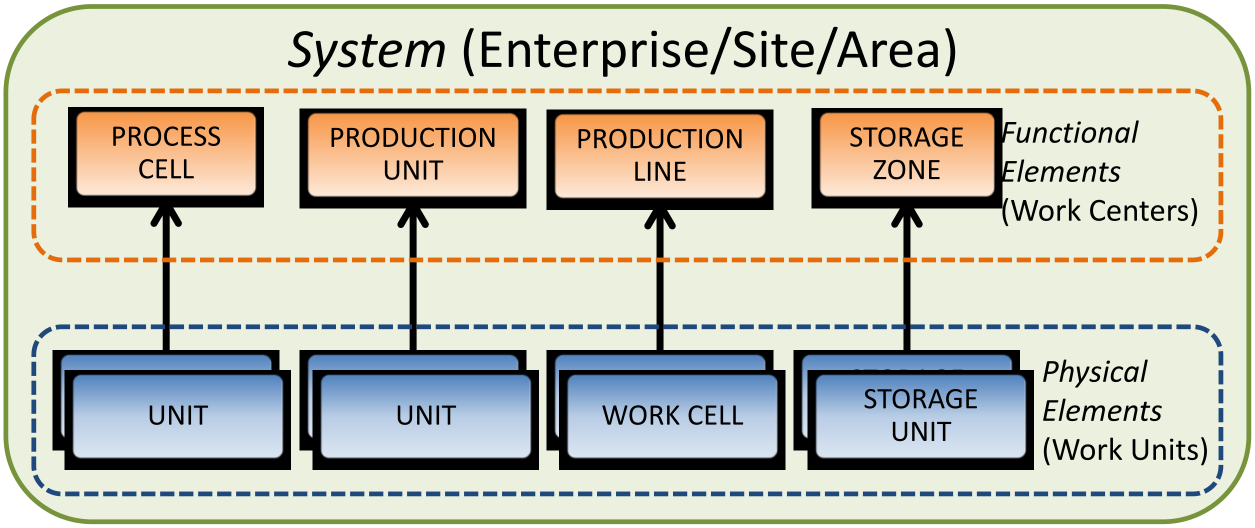

Figure 8: MRED's TTL Decomposition applied to a Facility's Operation Cluster Hierarchy

Given the Manufacturing Operations Management functional hierarchy presented in Figure 1 and noting that most KPIs are captured and/or generated at Level 3, further alignment of the manufacturing environment can be made with MRED:

- Work Units are synonymous with MRED's definition of *Components*. Work Units are physical elements where one or more work interact together to produce a *Capability*.
- Work Centers are synonymous with MRED's definition of *Capabilities*. Work Centers have functional objectives of fabricating a part, assembling a product, storing inventory, etc.
- The Enterprise/Site/Area is synonymous with MRED's definition of the *System*. This describes the whole of the manufacturing facility and is inclusive of its Work Centers and Work Units.

4.2. Metrics

KPIs are equated to being the *Technical Performance Metrics* of interest in the scope of RTFI as they are defined as quantifiable measurements that reflect an organization's success. CD22400 breaks down Manufacturing Operation Management into four categories as highlighted in Figure 9.

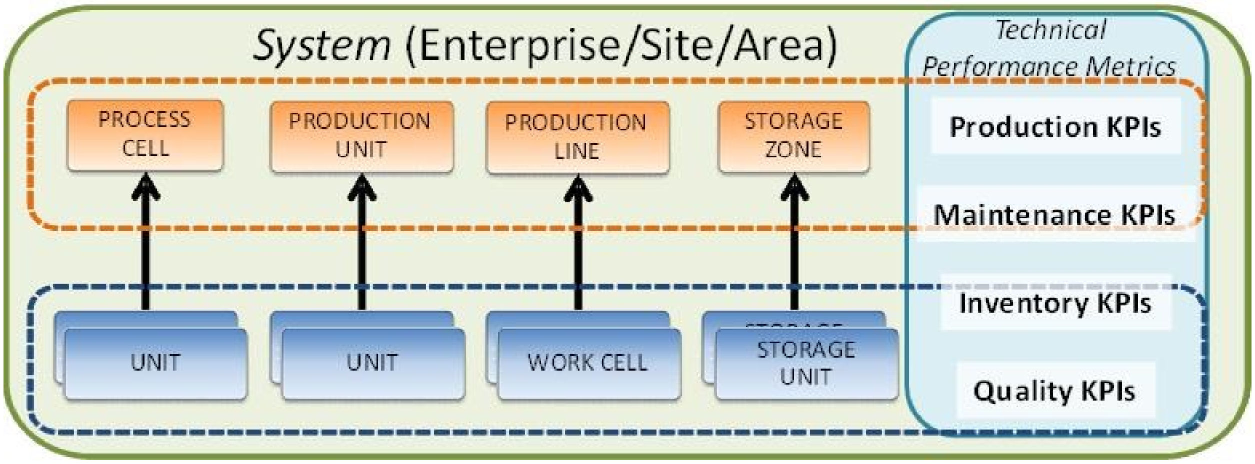

Figure 9: Inclusion of Metrics with MRED's TTL Decomposition applied to a Facility's Operation Cluster Hierarchy

Table 1 presents the specific KPIs defined in CD22400 within the four information categories defined for Manufacturing Operation Management. This list contains 34 different KPIs where successful operations equate to maximizing some KPIs (e.g., Worker efficiency, or Allocation throughput) and minimizing others (e.g., Set up rate, or Actual to planned scrap ratio). Given the complexity of many manufacturing facilities and the sheer volume of relevant KPIs, it is important to indicate which KPIs may be captured from which of the noted *TTLs*. Likewise, priorities derived from stakeholder preferences must be established to indicate which *TTLs* must be evaluated against the most preferred KPIs. The relationships (among KPIs and *TTLs*) and the derivation of stakeholder preferences are further discussed in Section 4.7.

4.3. Component/Capability Relationships

Figure 10 presents a simplified MRED application to the manufacturing hierarchy presented in Figure 9. The area is composed of three work cells (A, B, C) where Work Cell A supports Production Line 1 while Work Cell B and Work Cell C yield Production Line 2. Both of these production lines are within Area X. This example can easily be expanded to represent a more complex manufacturing facility. Figure 10 also diagrams the relationships among these *TTLs* and notes that *Technical Performance Metrics*, in the form of KPIs, may be captured at each level.

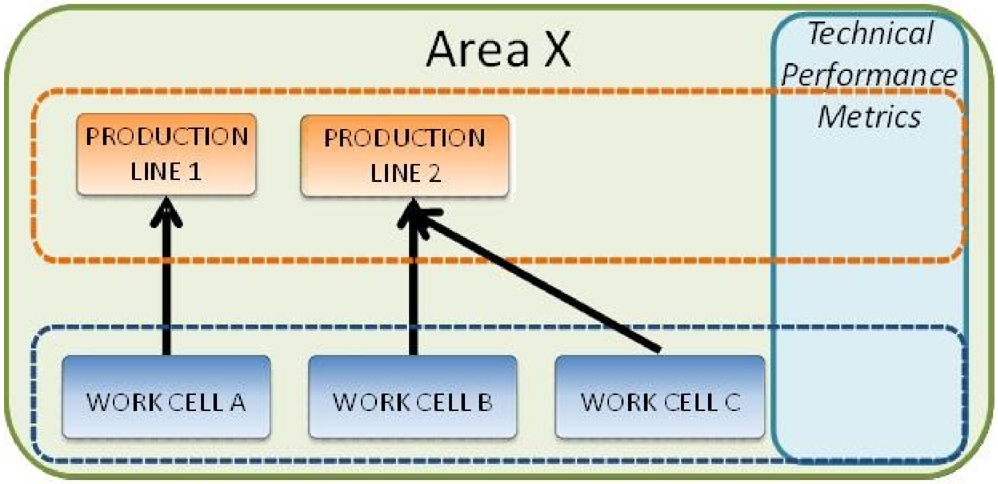

Figure 10: Component/Capability Relationships for an Example Manufacturing Area X

4.4. TTL/Metric Relationships

Table 2 presents a corresponding example TTL/Metric Relationship matrix indicating which KPIs would likely correspond to the example manufacturing TTLs. Note that some KPIs are relevant across all TTLs; other KPIs are only relevant for a few TTLs. Table 2 presents a very simple manufacturing example whereas most manufacturing facilities are significantly more complex, with more TTLs and Metrics (i.e., KPIs).

Table 2 – TTL/Metric Relationships for an Example Manufacturing Area X and its constituent TTLs

KPI	Work Cell A	Work Cell B	Work Cell C	Production Line 1	Production Line 2	Area
Worker efficiency				X		X
Allocation ratio				X	X	X
Throughput rate	X	X	X	X	X	X
Allocation efficiency	X	X	X			X
Utilisation efficiency	X	X	X	X	X	X
Overall equipment effectiveness index	X	X	X	X	X	X
Net equipment effectiveness index		X	X		X	X
Availability	X	X	X	X	X	X
Effectiveness	X	X	X	X	X	X
Quality ratio	X	X	X	X	X	X
Set up rate		X	X		X	X

4.5. Technology State

The next step in advancing through the MRED process is to define the technology state (as noted in Figure 5). If the entirety of a manufacturing facility is fully-developed, then all elements being considered for assessment can be evaluated against the relevant KPIs. Otherwise, the Technology State must be defined for each TTL, beginning at the *Component* level. The appropriate stakeholder (e.g., technology developer, or integrator) would indicate whether a specific Work Cell (i.e., *Component*) is either fully mature or immature. Given this maturity information and the relationships among the Work Cells and Production Lines (i.e., *Components* and *Capabilities*), Production Line maturity would be

calculated using the formula specified in [14]. The maturity of the entire Area (i.e., *System*) can now be calculated using the Production Line maturities and a similar formula, also presented in [14].

4.6. Filtering and Elimination

The filtering and elimination process begins in Figure 5 following the maturity calculations. First, KPIs are eliminated from immature *TTLs* (in the *TTL/KPI* relationship matrix) by stakeholders indicating which KPIs cannot be captured. This *TTL/KPI* relationship matrix is then automatically reviewed where a *TTL* is eliminated if there a no corresponding KPIs that can be captured (in its state of immaturity). The MRED process continues further by automatically updating relationship matrices and further eliminating KPIs that are no longer called (after *TTLs* were eliminated).

MRED continues this process with the identification, elimination, and filtering of *Environments, Tools,* and *Personnel* according to Figure 5. The identification of these elements and corresponding relationship matrices is done by the appropriate stakeholder(s) while the elimination and filtering steps are automatically performed by MRED. The filtering and elimination process ends with all of the available *TTLs, KPIs, Environments, Tools,* and *Personnel* being specified along with their updated, corresponding relationship matrices.

4.7. Stakeholder Preference Capture and "Aggregation"

The final significant process within MRED, captured in Figure 6, is capturing and aggregating the stakeholder preferences of the evaluation characteristics that yield test plan blueprints. It is critical that each KPI and each technological element be prioritized for assessment given the complexity of a manufacturing facility and the potentially large quantity of KPIs. As each manufacturing facility is unique, the stakeholders of one facility may place greater emphasis on specific KPIs as compared to the stakeholders of another facility.

An initial step in this process is to define the specific stakeholder population that will be providing their preferences for consideration. MRED defines five stakeholder groups as having an interest in technology assessments. The specificity of these groups can be tailored to the technology. Considering the manufacturing domain (in this example) and the emphasis on standards development, the stakeholder groups could reasonably be defined as 1) Owners, 2) Evaluation Designers, 3) Standards Organizations, 4) Technology Developers, 5) Technology Integrators, and 6) Users. MRED can accommodate preferences from these stakeholder groups where each would be weighted equally (future work in MRED is to explore weighting factors if the preferences of a stakeholder group are more/less important than others).

Stakeholder preferences for each *TTL-KPI* pair are now captured according to the first steps in Figure 6 and using the linguistic preference scale noted in Figure 7. The preferences would then be aggregated according to the aforementioned 11-point numerical scale where each *TTL-KPI* pair has an associated score from -5 to 5. The evaluation designer reviews these aggregate scores and determines which of the *TTL-KPI* pairs will be further considered for assessment and which will not at this time. The stakeholder preference capturing, scoring, and elimination process continues with respect to the personnel available

for testing. Stakeholder preferences must be captured for each grouping of *TTL-KPI* pairs given that some of the *KPIs* can be efficiently and realistically captured from the same *TTL* in a single test exercise whereas some of the *TTL-KPI* pairs can only be assessed with unique test considerations. This process repeats itself for the remaining test plan characteristics where the stakeholder preferences ultimately produce a test plan blueprint for each *TTL-KPI* pair and/or group.

MRED would provide the stakeholders with blueprints highlighting which *TTL-KPI* pairs should be tested according to the identified characteristics. MRED also captures each stakeholder's preference with respect to individual test plan characteristics so the information can be archived.

5. CONCLUSION and FUTURE WORK

This paper demonstrates the applicability of MRED to devising verification and validation tests. Initial steps are presented with an example at a high level. MRED applies measurement science to evaluation design in a manufacturing domain by implementing a structured process to rigorously filter and eliminate the unviable test plan characteristics while highlighting the stakeholders' preferred test plan characteristics. MRED's detailed application to this domain is planned for future work with a specific case study; development of the measurement science supporting automated optimal manufacturing. NIST researchers are planning to meet with industry partners to ascertain their preferences regarding specific KPIs and the technology test levels they are associated. This will include understanding the KPIs that are currently captured, the KPIs that were previously captured (and why they are no longer measured), along with the KPIs that could be captured in the future. NIST will use this data and work with industry partners to identify the most pertinent KPIs, determine how their effectiveness is measured, and isolate the characteristics that influence effectiveness. In this effort, MRED will initially be used as a tool to measure stakeholder preferences. After further consultation with the manufacturing community, MRED will be used to identify the most pertinent test plan blueprints. This data can be used to highlight the most critical verification and/or validation assessments of manufacturing systems and constituent elements.

6. REFERENCES

[1] Hillinger, C, 2004, "Voting and the Cardinal Aggregation of Judgments," *Munich Discussion Paper* 2004-9, http://epub.ub.uni-muenchen.de/353/.

[2] Horst, J., 2011, "Project Plans – EL Project Title and Number: Real-Time Factory Information, 7354004."

[3] International Organization for Standardization, 2011, "Automation Systems and Integration – Key Performance Indicators for Manufacturing Operations Management," *Draft International Standard ISO/DIS 22400-2*.

[4] Mankins, J., 1995, "Technology Readiness Levels," *Advanced Concepts Office, Office of Space Access and Technology, NASA,* White Paper.

[5] Mankins, J., 2009, "Technology Readiness Assessments: A retrospective," *Acta Astronautica,* **65**(9-10), pp. 1216-1223.

[6] NASA Systems Engineering Handbook, 2007, NASA/SP-2007-6105, Rev 1, http://ntrs.nasa.gov/.

[7] QIF – Quality Information Framework, 2012, http://qifstandards.org/, Last viewed: September 7, 2012.

[8] REQUIREMENTS FOR THE QUALITY INFORMATION FRAMEWORK (QIF) – An Industry Use Cases Research Report of the NIST Real-Time Factory Information (RTFI) Project, 2012.

[9] Weiss, B.A. and Schmidt, L.C., 2010, "The Multi-Relationship Evaluation Design Framework: Creating Evaluation Blueprints to Assess Advanced and Intelligent Technologies," *Proceedings of the 2010 Performance Metrics for Intelligent Systems (PerMIS) Workshop.*

[10] Weiss, B.A., Schmidt, L.C., Scott, H., and Schlenoff, C.I., 2010, "The Multi-Relationship Evaluation Design Framework: Designing Testing Plans to Comprehensively Assess Advanced and Intelligent Technologies," *Proceedings of the ASME 2010 International Design Engineering Technical Conferences (IDETC) – 22ND International Conference on Design Theory and Methodology (DTM)*.

[11] Weiss, B.A. and Schmidt, L.C., 2011, "Multi-Relationship Evaluation Design: Formalizing Test Plan Input and Output Elements for Evaluating Developing Intelligent Systems," *Proceedings of the ASME 2011 International Design Engineering Technical Conferences (IDETC) – 23RD International Conference on Design Theory and Methodology (DTM)*.

[12] Weiss, B.A. and Schmidt, L.C., 2011, "Multi-Relationship Evaluation Design: Formalizing Test Plan Input and Output Blueprint Elements for Testing Developing Intelligent Systems," *ITEA Journal*, **32**(4), pp.479-488.

[13] Weiss, B.A. and Schmidt, L.C., 2011, "The Multi-Relationship Evaluation Design Framework: Producing Evaluation Blueprints to Test Emerging, Advanced, and Intelligent Systems," *ITEA Journal* **32**(2), pp.191-200.

[14] Weiss, B.A., 2012, "Multi-Relationship Evaluation Design: An Interactive Test Plan Designer for Advancing and Emerging Technologies," Ph.D. dissertation, Mechanical Engineering, University of Maryland, College Park, MD.

[15] Weiss, B.A. and Schmidt, L.C., 2012, "Multi-Relationship Evaluation Design: Modeling an Automatic Test Plan Generator," *Proceedings of the 2012 Performance Metrics for Intelligent Systems (PerMIS) Workshop.*

[16] Wikipedia, 2012, "Manufacturing execution system," http://en.wikipedia.org/wiki/Manufacturing_execution_system.

[17] Williams, T., 1994, "The Purdue enterprise reference architecture," *Computers in Industry*, **24**(2-3), pp. 141-158.