Don't Trust a Management Metric,
Especially in Life Support

Harry W. Jones[1]
NASA Ames Research Center, Moffett Field, CA, 94035-0001

Goodhart's law states that metrics do not work. Metrics become distorted when used and they deflect effort away from more important goals. These well-known and unavoidable problems occurred when the closure and system mass metrics were used to manage life support research. The intent of life support research should be to develop flyable, operable, reliable systems, not merely to increase life support system closure or to reduce its total mass. It would be better to design life support systems to meet the anticipated mission requirements and user needs. Substituting the metrics of closure and total mass for these goals seems to have led life support research to solve the wrong problems.

Nomenclature

ESM	=	Equivalent System Mass
EVA	=	ExtraVehicular Activity
HEFT	=	Human Exploration Framework Team
LCC	=	Life Cycle Cost
Pr(LOC)	=	Probability of Loss of Crew
R&D	=	Research and Development
TRL	=	Technology Readiness Level

I. Introduction

THIS paper considers the use of metrics in management and describes their past use in guiding space life support research. A metric is an easily measured system property that is used as a standard of comparison. Metrics are widely used in systems management but they cause notorious and inescapable problems. Goodhart's law states that any metric that is used to control a management process will become distorted and will also misguide the process. Since this frequently occurs, metrics fail to represent reality and should not be relied on to make decisions. The metric of Equivalent System Mass (ESM) has been used nearly exclusively for decades in planning life support research and development (R&D). Selecting research projects and allocating funding based on a metric such as ESM typically causes the metric to be distorted and sometimes deliberately manipulated. Much more important, over emphasis on reducing ESM may have greatly harmed life support R&D by causing the neglect of much more important goals, such as cutting mission cost or improving reliability. Increasing the percent closure justifies recycling any waste material, regardless of cost or practicality. The Technology Readiness Level (TRL) metric naturally increases as R&D progresses. Advancing TRL seems the sure path to R&D success but it can fail because TRL measures technology push rather than mission pull. When metrics are used to allocate R&D funding, the metrics are subverted and the research misguided. The better systems engineering approach would first identify the user requirements and then work top-down to identify alternate implementations and make practical trade-offs. Metrics cannot replace comprehensive systems engineering.

II. Goodhart's law – the Murphy's law of metrics

Metrics are widely used to manage but frequently criticized. Goodhart's law and similar statements describe the basic metric problem. Using a metric to control a process will disrupt the process and distort the metric. Management metrics do not represent reality and cannot be trusted to guide action.

Not all metrics should be distrusted. Quantitative measures must be used in accounting, management, and engineering. They are typically factual, verifiable, and useful, for instance in defining the performance requirements

---

[1] Systems Engineer, Bioengineering Branch, Mail Stop N239-8.

for a system to be built and tested. Management by a metric is an attempt to guide an organization by controlling the future value of some parameter. Management metrics tend to be oversimplified and distorted representations of an incompletely understood complex process.

A. Goodhart's law

A simple statement of Goodhart's law is, "If a metric is made a goal, it becomes a bad measure." Any metric that is used for monitoring and control of human organizations becomes distorted and ineffective.

1. Original development of Goodhart's law

Goodhart's law was first proposed in the 1970's by the central bank economist Charles Goodhart to explain why adjusting the money supply fails to control inflation. Professor Goodhart suggested that any metric used for management loses the real world significance that originally made it relevant. People change their behavior to adjust the metric and to help them meet its target value. This breaks the relationship between the metric and the objective to be controlled. The metric becomes a bad measure of the real objective as soon as it is adopted as a substitute target. (Kay) (Wikipedia, Goodhart's law)

Goodhart proclaimed his law as a half-joking explanation of the failure of central bank monetary policy to manage inflation. Goodhart's law has strongly influenced modern central bank actions; "monetary targets are out and inflation targets are in." Goodhart later disparaged his law as merely a "mixture of the Lucas critique and Murphy's law." (Chrystal and Mizen)

B. Murphy's law

Murphy's law is the well-known engineering statement, "Anything that can go wrong, will go wrong." It was formulated in the 1950's aerospace industry as a pretended scientific law, the "newly discovered fourth law of thermodynamics." Murphy's law is a useful reminder to design defensively, to identify and manage risks.

Two things go wrong when a metric is used to manage a system. The metric itself is distorted and misreported. The system monitored is redirected to improve the metric and then neglects its original purpose.

C. Lucas critique

Goodhart's law and Murphy's law are pseudoscientific pronouncements that teach real lessons, but the Lucas critique is a serious technical warning against using metrics in economic models. The Lucas critique made in 1976, a Noble prize achievement, states that it is simply not possible to guide economic policy using the metric relationships observed in historical data, such as the one between money supply and inflation that failed for Goodhart. Economic models predict the actions of intelligent economic decision makers using relationships observed in their past behavior. However the expected actions of the decision makers predicted by the models strongly depend on the decision makers' expected government economic policy. A new policy invalidates the original economic model. Altering the money supply to control inflation does not produce the results predicted by the original model. (Wikipedia, Lucas critique)

The Lucas critique and Goodhart's law tell us that metrics do not work. Economic analysts and policy makers cannot predict or control the economy using metrics because the very attempt to do so changes the economy and invalidates the metrics. If metric-based adaptive control is counterproductive, it may be better to use simple policy rules.

1. Fixed and trusted rules

In response to the Lucas critique, Kydland and Prescott published the article "Rules rather than Discretion: The Inconsistency of Optimal Plans." Mathematical control theory shows how to maximize an objective function if the past decisions and the current system state are known and if they and the current action completely control the next state and its outcomes. But rational economic agents make decisions based on their expectations of the future. If future policy changes alter agents' expectations, they are not able to successfully maximize the original objective function. Fixed rules that are established and trusted, such as future inflation targets, can lead to better performance. (Wikipedia, Lucas critique) (Kydland and Prescott)

D. Campbell's law

Campbell's law is a social science version of Goodhart's law, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." Campbell's law has been used to explain cheating following high-stakes testing in U.S. classrooms. (Wikipedia, Campbell's law)

E. The McNamara fallacy

Robert McNamara was the United States Secretary of Defense during the Vietnam War. His knowledge of systems analysis led him to quantify success in the war using enemy body count. Daniel Yankelovich formulated the McNamara fallacy as follows: "The first step is to measure whatever can be easily measured. This is OK as far as it goes. The second step is to disregard that which can't be easily measured or to give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily really isn't important. This is blindness. The fourth step is to say that what can't be easily measured really doesn't exist. This is suicide." (Wikipedia, McNamara fallacy) Metrics can focus attention so narrowly that all sense of reality is lost.

F. Metrics in engineering

Clearly, metrics are unsatisfactory in controlling inflation, optimizing social systems, evaluating teachers, and fighting wars. Metrics are widely used in engineering. Being human, engineers managed by metrics manipulate the metrics and change their behavior in response to the metrics.

"If people even suspect that metrics will be used for evaluation of personnel, they will be reluctant to report findings that might be detrimental to their careers. People will also be tempted to inflate positive indicators. Such massaging of metric data will erase objectivity and validity." (Naveda and Seidman)

"Whether or not our models are correct, and regardless of how well or poorly we collect and compute software metrics, people's behaviors change in predictable ways to provide the answers management asks for when metrics are applied…. Some of these actions seriously hamper productivity and can effectively reduce quality." (Hoffman)

"The sad thing is the number of managers and engineers who don't realize that given any moderately complex equation with multiple variables, values can be selected to generate any particular result." (Hoffman) Engineering metrics do not represent reality and cannot be trusted to guide management.

III. Managing life support research using metrics

Even sensible metrics measuring a real parameter can be manipulated and become counterproductive. Improving a metric is usually not a good planning guide, as has been found for life support research metrics.

The three metrics that have been used in space life support are system closure, Equivalent System Mass (ESM), and Technology Readiness Level (TRL). These metrics have been used at different times to justify life support research projects or select technology. Increased closure, lower ESM, and higher TRL all suggest better life support, but improving one of them will usually not improve the others or lead to overall better life support. Increasing closure may increase ESM rather than reduce it. Decreasing ESM may incur excessive costs or performance penalties. Requiring higher TRL technology could lead to the use of older less capable systems and even decrease closure or increase ESM.

A. Closure

The system closure metric for life support is the percentage of all life support material - oxygen, water, food, and other supplies for the crew - that is provided by recycling rather than supplied from Earth. The early goal of life support R&D was to approach a totally closed human ecosystem, independent from Earth. Closure is an important concept in ecosystem theory and artificial ecosystem design.

1. The closure metric

At the time that the space station was being developed, the future life support system for the Moon and Mars was expected to include food production, waste recycling, and ultimately to be "totally closed except for losses due to leaks, EVA's, etc.," and to approach "complete closure of the food and solid waste loops." (Wieland) (Bilardo) "The goal for these (Moon and Mars) missions is a higher level of mass recovery, perhaps achieving 95% closure." (Wieland)

Figure 1 shows the number of person-days that can be supported by 1,000 kg of supplies from Earth for different percentages of closure achieved by the life support system.

International Conference on Environmental Systems

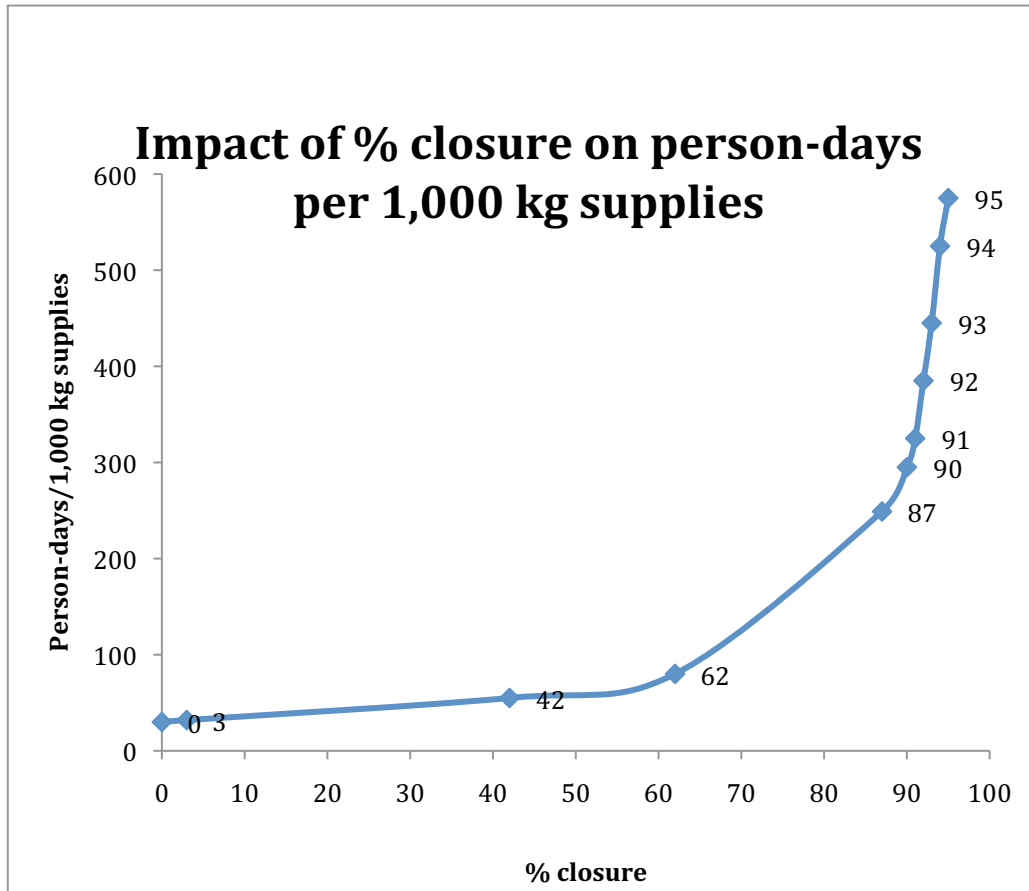# Impact of % closure on person-days per 1,000 kg supplies

Figure 1. Person-days for 1,000 kg supplies versus percent closure.

Figure 1 data is replotted from (Bilardo and Theis). The number of person-days that can be supported open loop, at zero percent closure, with 1,000 kilograms of air, water, food, carbon dioxide removing material, etc., is only about thirty. The life support technologies used to achieve the gradual increase in closure are shown in Table 1, with the resulting number of person-days supported per 1,000 kilograms.

Table 1. Life support technologies for increasing closure.

% closure	Life support	Person days/1,000 kg
0	Open, all supplies from Earth	30
3	Closed air, open water	32
42	Open air, 50% water recycle	55
62	Open air, 70% water recycle	80
87	Closed air, closed water	249
90	Dehydrated food	295
91	Minimum expendables	325
92	Increased recycling	385
93	Full human waste recycling	445
94	50% food grown	525
95	75% food grown	575

With minimum expendables, increased recycling, and growing most of the food, 1,000 kilograms of materials from Earth can support one person for 575 days.

International Conference on Environmental Systems

2. Problems with closure

Increasing the closure of space life support makes human space exploration more independent from Earth and reduces the expense of launching life support materials to sustain the crew. However, increasing closure is subject to diminishing returns and exponentially increasing costs. The most abundant and easiest to treat wastes, such as condensed atmosphere water or hygiene water, are treated first. Methane produced from exhaled carbon dioxide or human solid waste both are more difficult to recycle and are a much smaller part of the total ecosystem material circulation. Growing food is obviously needed for full closure but requires a large greenhouse, high power for lights, and massive supporting equipment.

Achieving high levels of material closure is theoretically appealing but is not cost effective. The closure metric was used to justify developing uneconomic technology for recovering difficult minor wastes and for growing food plants in space. Difficult and costly recycling efforts are still justified as needed to complete system closure. Percent closure was replaced as the major metric in life support R&D by Equivalent System Mass (ESM) in the late 1990's.

B. Equivalent System Mass (ESM)

Launch mass was used as a metric in the early days of rocketry. Equivalent mass is the total launch mass needed to provide and support a system.

1. The ESM metric

Traditionally equivalent mass included mass, volume, power, cooling, and materials and spares logistics, but in 1992 the crew time to operate and maintain the hardware was added. (Drysdale et al., 921241) (Levri et al., 2000-01-2395) This equivalent mass with crew time was renamed Equivalent System Mass (ESM) in 1998, when ESM was selected as the basis of the life support progress metric (Drysdale and Hanford, 39503) ESM was also then made the usual basis for life support technology selection, e.g., (Maxwell and Drysdale, 2001-01-2365).

ESM includes the mass of the flight hardware, the equivalent masses of the spacecraft power and the pressurized volume that the hardware uses, the mass of the cooling capacity required, the mass of resupply and spares, and an equivalent mass of the crew time. The pressurized volume is converted to mass using a mass equivalence factor, so many kilograms per cubic meter, based on the launch mass and volume of the structure. The power is also converted to mass at so many kilograms per watt, using the total mass and power capacity of the power system. The supplied power must be removed as heat. The allocated launch mass of the cooling system is similarly included in ESM. The launch mass of the resupplied materials and the spares is added directly to the ESM. The equivalent mass charged for crew time, so many kilograms per hour, is nonphysical and is difficult to determine but is needed to control crew time for operations and repairs. The addition of crew time shows that launch mass is not the only important factor in technology selection.

The life support systems for short missions are open loop. Life support R&D has aimed at reducing resupply mass for long missions by recycling and increased closure. But increasing closure can reduce the total cost only if it reduces the launch mass and cost. Reducing ESM does reduce the launch cost. Replacing closure by ESM increased the cost realism in life support R&D.

2. Problems with the ESM metric

There are four serious and fundamental problems with the ESM metric. First, ESM reflects only launch cost, giving a partial and biased view of life support cost. Second, ESM comparisons made for technology selection are unreliable and can be distorted, as often happens in metric reporting. Third, reducing ESM has guided life support R&D in a direction away from solving more urgent problems. Fourth, the use of the flawed ESM metric gives an unfortunate impression of unskillful management.

<u>ESM is partial and biased.</u> Equivalent system mass affects only launch cost. Mission and research planning should consider the entire cost, Life Cycle Cost (LCC). LCC includes development, launch, and operations costs. (Jones, 2003-01-2635) For recycling life support systems, the development and operations cost is usually much larger than the launch cost.

But cost is only one factor in technology selection and it is far from the most important factor. The addition of crew time to ESM concedes that mass is not all that matters. Other factors that should be considered relate to performance, availability, and safety. Performance includes requirements for the quantity and quality of the output product and the "-ilities" such as maintainability and reliability. The readiness of hardware for mission use is measured by its Technology Readiness Level (TRL). Safety can be measured by the number and criticality of the potential failures or by the life support contribution to the mission Probability of Loss of Crew, Pr(LOC). (Jones, 1999-01-2079)

ESM is biased in favor of recycling systems over use of stored and supplied materials. Supplying all materials has very high launch mass and ESM but very low development cost for the materials and their containers. Recycling systems have much lower launch mass and ESM, but they have very high development cost for the recycling

systems themselves. The LCC of life support recycling systems is usually much higher than the LCC of storage systems. (Jones, 2013-3407) (Jones, 2012-3418)

ESM is unreliable and can be distorted. When any metric is used to allocate funding, it will probably be manipulated. Two technology trade-off studies using ESM are notable. In the 1990's NASA was interested in applying biology in life support recycling and in growing plants for food in space. This seemed reasonable when closure was the accepted life support metric but ESM analysis had to be favorably adjusted to justify these important on-going projects. (Flynn et al., 981538) found, contradicting previous analysis, that a management-favored bioreactor water processor had about seven times higher ESM than competing non-biological systems. This led management to reorganize and tighten control of ESM analysis so that in the future there would be "no surprises." Jones (2006-01-2082) found, contradicting previous analysis, that the large mass and power of plant growth chambers makes growing food much more expensive than supplying food even for decade-long missions.

ESM diverted R&D from important problems. Life support analysis has spent decades, dozens of work-years, and millions of dollars supporting and applying ESM, including developing an extensive database useful only in computing ESM and an interactive system sizing computer tool that considers only ESM. The use of ESM (and closure before it) has completely misguided life support research. More millions have been spent developing the wrong technologies for the wrong purpose. Recycling saves launch mass over storage, so it was claimed that lower mass recycling systems were needed to further reduce mass. But diminishing returns set in. Recycling water and oxygen saves much mass, but a lighter recycling system saves relatively little. The obvious but neglected need is to develop more reliable, operable, maintainable systems. The NASA Human Exploration Framework Team (HEFT) found a need for high reliability life support for deep space missions, but no requirement for more material closure or for lower ESM recycling systems. (HEFT, 2010) (Jones, 2011-01-5037) The correct priorities were made clear in the mission planning literature but not appreciated in life support.

Use of ESM is inept. The program management results show using ESM was an error. But three things make ESM an attractive mistake. ESM favors recycling over storage by including only launch cost. ESM is a simple sum of factors, so it is very easily adjusted to provide technical support for projects selected for other reasons. And ESM allows management decisions to ignore technical details other than launch mass while dealing with complicated organizational and political factors. Using tools that are known not to work is unacceptable.

C. Technology Readiness Level (TRL)

Government technology development is usually explained as a process of advancing technology from research through prototype and test to application. Technology Readiness Level (TRL) is used by NASA, the US Departments of Defense and Energy, the European Space Agency, and others.

1. TRL metric

Each successive stage in technology development is indicated by a TRL from 1 to 9. The TRL levels are defined as follows: (Jones, 1999-01-2079)

TRL 1 - Basic principles observed and reported
TRL 2 - Technology concept formulated
TRL 3 - Critical function proof-of-concept
TRL 4 - Component or breadboard validated in laboratory
TRL 5 - Components validated in a relevant environment
TRL 6 - Prototype demonstrated in a relevant environment
TRL 7 - Prototype demonstrated in a space environment
TRL 8 - Design flight qualified
TRL 9 - System flight proven in mission operations

The space station technology selection process first eliminated all but the two highest TRL technologies for each function and then developed and tested prototypes of both. (Carrasquillo, Carter, et al.) Considering only the highest TRL technologies is very appropriate for a flight mission. R&D funding is limited, so only the better concepts are developed to a breadboard and only the best performing breadboards are developed into prototypes. A good candidate technology will have its TRL increase as continuing R&D investments are made. And if R&D investment decisions consider performance, cost, and operability, technologies with higher TRL will be better in those factors. Most important, the higher a technology's TRL, the lower the remaining cost, schedule, and risk to fly it.

TRL is an extremely useful metric for mission technology selection. Most missions have an operational objective. Missions are rarely developed to demonstrate new technology. Mission planners prefer flight proven technology over new systems. A mission that depends on new undeveloped technology is not likely to be approved.

Selecting the highest TRL technologies is not appropriate in managing R&D. Low TRL alternates may be rejected prematurely. Changed missions and requirements may need new approaches. Some programs are specifically intended to do low TRL research, to find new game-changing ideas.

Essentially, higher TRL is the product of more R&D funding. Using higher TRL to allocate R&D funding and so to increase TRL creates a self-fulfilling feed forward loop. As seen in consumer markets, the technology winner can be the first to market rather than the technically best.

The government technology development process modeled by advancing TRL has a fundamental problem. Development should satisfy the customer, but government R&D splits the customer in two, the actual current funder and the hypothetical future user. Government research programs develop technology to be used later by potential future missions. The R&D effort develops early prototypes but does not flight qualify systems. Yet missions do not want to use technology that is not flight qualified. This causes the technology transfer problem, the mid-TRL gap. Because the user is not involved, the user-needed technology is not developed. The R&D program goal is to fund current projects, not to provide systems for future missions.

The objective of space life support R&D should be the development and demonstration of high TRL flight qualified systems that are superior to the current space station systems. It seems that bad metrics, a distorted selection processes, and unsuitable research projects are the result, the symptom and not the cause, of a technology development process that has gotten off track because of lack of user input.

D. Life support metrics overview
The three life support metrics are compared in Table 1.

Table 1. Closure, ESM, and TRL comparison matrix.

	Closure	ESM	TRL
Fundamental, important?	Yes	Yes	Yes
Where used?	Ecosystems	Life support	Government R&D
Purpose	Increase recycling	Cut launch cost	Track R&D progress
Effect	Recycle all waste	Recycle if saves mass	Justify R&D
OK as one elimination factor?	No	Yes	Yes
OK as sole selection factor?	No	No	Yes
Actually used alone?	No	Yes	No
Other factors suppressed?	No	Yes	No
Adjusted to affect funding?	Yes	Yes	Yes
Justified poor projects?	Yes	Yes	No

All three metrics measure important, interesting life support system parameters. Most metrics seem reasonable and useful when proposed, but they always have unexpected, counter-productive consequences. They become distorted and misguide the underlying process. These life support metrics are partial and narrow and therefore distracting and damaging.

Closure, by its very nature is specific to material recycling ecosystems such as space life support. Closure is a high level system metric, far separated from implementation details. It provides a general justification for any recycling regardless of cost but is not very useful in discriminating between technology projects.

ESM could be used for any hardware launched by rocket. Launch mass metrics were widely used early in the space age. Now ESM is used only in life support. ESM justifies any recycling system that saves mass. Higher ESM technologies are rightly rejected. ESM emphasizes the launch mass savings of recycling over stored materials. ESM was used as the sole life support selection factor and hardware cost, reliability, and other factors were ignored or suppressed. The ESM metric was sometimes managed to produce acceptable results, as often happens when funding is involved.

TRL was used in space station technology selection, which did not consider closure or ESM. But the mass, power, volume, and cooling elements of ESM were assessed, along with reliability and many other factors. Selecting for high TRL may eliminate some candidates too soon, but is unlikely to select an especially bad technology.

IV.   Resisting and replacing management metrics

The life support metrics of closure, ESM, and TRL are all currently used in life support analysis, but they are no longer used exclusively or as the major deciding factors. Other key performance parameters can be considered,

including cost, safety, reliability, and operability. But no formal project selection process is now implemented. It is still possible that people will again propose using unreliable management metrics. Management by metrics should be directly resisted and must be replaced by better methods.

A. Resisting metrics

The limitations of metrics are obvious, well known, and unavoidable. Metrics should be resisted by explaining that they do not work, pointing out that metrics are bad management tools, and emphasizing the greater importance of the real goals.

1. Metrics do not work

Goodhart's law is undeniable and inescapable. If a metric is used to control a human process, people will change their behavior in ways that distort the metric and disrupt the process. Metrics fail to represent reality and so cannot be trusted to guide management.

2. A management metric is a bad tool

Metrics are not used in effective management. Management metrics appear superficial, inept, and even insincere. The use of metrics has been discredited.

3. The important real goals are lost by using a management metric

A management metric is usually used instead of a more complicated and partly conflicting set of real goals. The metric is a substitute for the real goals. Life support must be made less costly, easier to operate, and more reliable, not necessarily to have lower ESM.

B. Replacing metrics

In the world of theory, ideal methods would easily replace management by metrics. In the real world, metrics are hard to replace.

1. The ideal methods of science and technology, of systems analysis and engineering

Science seeks to understand and technology to create the new. Both use critical methods to establish a firm base in reality, science uses experiments and technology relies on tests and demonstrations. Systems analysis uses data and logic to determine what systems to build. A system usually has several important but conflicting goals, such as low interest rates and low inflation, high multiple choice scores and writing skills, or destroying the enemy and building peace. The management process developed to balance and implement conflicting goals is systems engineering. It defines the user requirements, considers alternate implementations, and performs optimal trade-offs. Using a metric replaces the actual requirements and obscures the system goals.

2. Real organizations, real people

The actual work process of science, technology, and engineering organizations is selecting projects and funding them. Since this political process determines the fate of projects and people, it is an intense game played without rules. Understanding, creativity, and making optimal trade-offs are not prominent concerns. It is not surprising that nearly everyone could accept management by a metric such as ESM without objection. It is not surprising that all project advocates would make reducing ESM their major justification. It is not surprising, given the known history of metrics, that unfavorable ESM calculations could be adjusted and made better. It is not surprising that mentioning cost, reliability, or safety would be unpopular. The organizational reality differs much from the technical ideal.

3. If something can't go on, it won't

Fifty years ago, in the mid-1960's, I bought a new Ford Mustang, the hot car of the era. The salesman did not mention, and I did not ask about, value, fuel economy, reliability, or safety. They were not good. This year I bought a new Toyota Camry, a current favorite. Again the salesman did not mention, and I did not ask about, value, fuel economy, reliability, or safety. But they are quite good, fantastically improved over 50 years.

What caused this total change? Many things contributed, customer demand, foreign competition, activism by Ralph Nader, news reports, and government standards, tests, and regulations. Fundamentally, the automobile design situation was unacceptable and had to change. Much good engineering was done. Manufacturers and drivers now expect good value, economy, reliability, and safety. Social reality is now much closer to the technical ideal.

4. Improving the real by incorporating the ideal

Humans are social animals. We naturally, instinctively, unquestioningly adopt our group's values. We persuade and even force other group members to conform. If our group values are narrow, short term, group or self-serving, the group may well survive and have some limited, temporary success. Time and experience show that an enduring and respected organization must combine survival skills with broad, long term, socially approved values.

Life support has not done well by managing using the ESM metric and deliberately ignoring cost, economy, reliability, and safety. Life support needs justifiable goals, better management and engineering, and demonstrated results.

V. Conclusion

Metrics have two mutually reinforcing problems that are almost unavoidable in practice. Making project decisions and allocating funding using a metric nearly always causes the metric to be manipulated, distorted, and even falsified. But much worse, the actual work process being measured and managed is diverted and changed to improve the metric. Important considerations not measured by the metric are neglected.

Experience in life support confirms that metrics are corrupted and work is misdirected. Closure and ESM have been manipulated and, because they do not measure the true system objectives, have misguided research. TRL has usefully helped guide space station technology selection, but conducting research to advance TRL can neglect user input and commitment.

Why are metrics often used? A metric provides a formal, predictable, and controllable decision process. A metric is simpler, requires less data, and is easier to compute than a full trade-off study. The objective of analysis is often to support a known decision, and in this case a metric is sufficient.

The fact that life support R&D has had the wrong goals apparently explains why it has not emphasized developing the systems actually needed for future missions. But this obvious logic is wrong, backwards. In the theoretical systems engineering approach, the goals are set and then projects selected. In practice the projects come first. Researchers have technologies they want to develop and managers a need to expand R&D. The metrics are adopted to help advocate the projects. The misuse of metrics is a symptom rather than the cause of misguided R&D. The life support metrics support internally inspired technology driven projects. The objective should be to develop systems to meet external, higher level, long-term mission objectives.

The use of narrow metrics has caused life support research to advocate the wrong goals, increasing closure or reducing ESM, rather than providing the high reliability more operable hardware that mission designers would prefer. Using metrics to manage is a mistake. Metrics should be directly opposed. The best way to manage research is to use the systems engineering process to develop systems that meet user needs.

References

Bilardo, V. J., Jr., "The Physical/Chemical Closed-Loop Life Support Research Project," AIAA-90-3729, AIAA Space Programs and Technologies Conference, September 25-28, 1990, Huntsville, AL.

Bilardo, V. J., Jr., and Theis, R. L. A., "Regenerative Life Support Technology Challenges for the Space Exploration Initiative," Proceedings of Third International Conference on Engineering, Construction, and Operations in Space III, Denver, CO, May 31 - June 4, 1992.

Carrasquillo, R. L., Carter, D. L., Holder, D. W., Jr., McGriff, C. F., and Ogle, K. Y., Space Station Freedom Environmental Control and Life Support System Regenerative Subsystem Selection, NASA TM- 4340, February, 1992.

Chrystal, K. A., and Mizen, P. D., "Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy," Bank of England, November 15-16, 2001. (Reference in http://en.wikipedia.org/wiki/Goodhart%27s_law)

Drysdale, A. E., and Hanford, A. J., Advanced Life Support Research and Technology Development Metric - Baseline, CTSD-AD-384, JSC 39503, November, 1999.

Drysdale, A., Thomas, M., Fresa, M., and Wheeler, R., "OCAM - A CELSS Modeling Tool: Description and Results," SAE 921241, Society of Automotive Engineers, Warrendale, PA, 22nd International Conference on Environmental Systems, 1992.

Flynn, M., Fisher, J., and Borchers, B., "An Evaluation of Potential Mars Transit Vehicle Water Treatment Systems," SAE 981538, 28th International Conference on Environmental Systems, 1998.

HEFT, Human Exploration Framework Team, Phase I Closeout, NASA Steering Council, September 2, 2010.

Hoffman, D., "The Darker Side of Metrics," Software Quality Methods, 2000.

Jones, H., "Comparison of Bioregenerative and Physical/Chemical Life Support Systems," SAE 2006-01-2082, 36th International Conference on Environmental Systems, 2006.

Jones, H., "Design and Analysis of a Flexible, Reliable Deep Space Life Support System," AIAA-2012-3418, 42nd International Conference on Environmental Systems, 2012.

Jones, H., "Equivalent Mass versus Life Cycle Cost for Life Support Technology Selection," SAE 2003-01-2635, 33rd International Conference on Environmental Systems, 2003.

Jones, H., "Mission Design Drivers for Life Support," AIAA-2011-5037, 41st International Conference on Environmental Systems, 2011.

Jones, H., "Multiple Metrics for Advanced Life Support," SAE 1999-01-2079, 29th International Conference on Environmental Systems, 1999.

Jones, H., "Storage or Recycling Life Support for Mars?" AIAA 2013-3407, 43rd International Conference on Environmental Systems, 2013.

Kay, J, "The law that explains the folly of bank regulation," Financial Times, Sept. 12, 2012.

Kydland, F. E., and Prescott, E. C., "Rules Rather than Discretion: The Inconsistency of Optimal Plans," Journal of Political Economy, 1977, vol. 85, no. 3, pp. 473-491.

Levri, J. A., Vaccari, D. A., and Drysdale, A. E., "Theory and Application of the Equivalent System Mass Metric," SAE. 2000-01-2395, 30th International Conference on Environmental Systems, 2000.

Maxwell, S., and Drysdale, A., "Assessment of Waste Processing Technologies for 3 Missions," SAE 2001-01-2365, 31st International Conference on Environmental Systems, 2001.

Naveda, J. F., and Seidman, S., B., IEEE Computer Society Real-World Software Engineering Problems: A Self-Study Guide for Today's Software Professional (Google eBook), John Wiley & Sons, Feb. 22, 2013, p.. 250.

Wieland, P., "ECLSS Development for Future Space Missions," AlAA 90-3728, AIAA Space Programs and Technologies Conference, September 25-28, 1990, Huntsville, AL.

Wikipedia, "Campbell's law," URL: http://en.wikipedia.org/wiki/Campbell%27s_law [cited 17 September 2012].

Wikipedia, "Goodhart's law," URL: http://en.wikipedia.org/wiki/Goodhart%27s_law [cited 17 September 2012].

Wikipedia, "Lucas critique," URL: http://en.wikipedia.org/wiki/Lucas_critique [cited 17 September 2012].

Wikipedia, "McNamara fallacy,' URL: http://en.wikipedia.org/wiki/McNamara_fallacy [cited 17 September 2012].