Operational Sustainability Metrics Assessing Metric Effectiveness in the Context of Electronics-Recycling Systems

JENNIFER ATLEE AND RANDOLPH KIRCHAIN*
Engineering Systems Division and Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E40-421, Cambridge, Massachusetts 02139

In the past 15 years corporations and governments have developed a growing appreciation of the need for sustainability. However, there is still little clarity on how to move toward the goal of sustainability or measure improvements. Not only are there currently few operational metrics by which to practically assess progress toward sustainability, there is also little understanding of how to judge the effectiveness of such metrics. This paper presents a pragmatic approach to developing—and evaluating—system-specific performance metrics for sustainability. Electronics recycling is used as a case problem in developing and judging the effectiveness of such metrics. Despite growing concerns about the handling of end-of-life electronics, data availability is inconsistent, and there is still limited understanding of the electronics-recycling system as a whole. To begin to address the need for practical quantitative methods to assess system performance, several indicators were developed and applied to three U.S. electronics-recycling operations. These metrics were assessed based on the developed criteria that effective measures be useful, robust, and feasible. Results show that the current measure of “mass percent to landfill” is not sufficient to assess system performance. Relevance-weighted mass indicators with varying data requirements can provide additional insights on resource efficiency.

I. Introduction

Realizing global sustainability will likely require broad changes in both consumer and institutional behavior as well as in the technologies that enable the benefits of modernity. To accomplish such substantive change requires methods capable of identifying those decisions of individual stakeholders that move toward the overall goals. Recently, a number of individuals and organizations have examined the question of measuring sustainability performance (cf., the following section). To date, very limited work has been carried out to specifically evaluate these measures for their practical and effectual merit in the context of operational decisions. This paper uses the case of electronics recycling to assess the merits of such operational level metrics.

The issue of how best to deal with end-of-life electronics is a real, current, and messy problem based around a low value, high toxicity, complex waste stream about which there is little existing data or comprehensive understanding. The development of operational sustainability measures can contribute to improvements in the electronics-recycling system while the process of developing, assessing, and applying these measures for this application informs development of operational sustainability measures for other systems.

For the sake of brevity and because of its particular relevance to a material recovery focused industry, subsequent discussion will be limited to only one aspect of system performance: the sustainability of resource use. The authors acknowledge that such focus on resource sustainability omits important social and economic aspects of sustainability. Nonetheless, the principles discussed should be illustrative and serve as a basis for future efforts to include those factors.

To begin to address the question of developing operational sustainability measures for the electronics-recycling industry, this paper first develops a set of criteria for evaluating metric effectiveness. Existing and alternative measures for the industry are described and subsequently applied to three case facilities. The merit of the applied metrics is discussed using the developed criteria.

II. Measuring Sustainability: Status and Issues

Despite growing agreement on the conceptual definition of the goal of sustainability, questions remain on how to measure performance. Metrics are critical to accomplishing any goal insofar as they implicitly or explicitly define (1) system boundaries, (2) traits which are emphasized, and (3) the definition of improvement. Sustainability will not be successfully incorporated into firm actions until there are effective ways to measure progress toward it (1).

Efforts to develop sustainability metrics can generally be split into two categories. Some authors have cataloged exhaustive sets of indicators to evaluate a target industry (2, 3), while other researchers have developed composite indicators that try to address a broad set of issues (4–6). Despite recognition that such reporting would require extensive time and resources and, therefore, “further work is needed to develop a more simplified framework” (3), there has been little effort to match metric development to feasible data collection. To date, indicators of sustainability performance capable of informing the decisions of those effecting change in products, processes, and policy are little developed and have not been specifically evaluated for their practical and effectual merit. General efforts within the literature to define the dimensions of merit for sustainability metrics have resulted in the criteria catalogued in Table 1. Ultimately, these criteria can be summarized in a framework specifying that a successful metric must be (1) useful, (2) feasible, and (3) robust.

To evaluate candidate operational metrics, a short list of criteria was distilled from Table 1. The short list is generic and covers the basic elements needed in any indicator. Criteria that were redundant or not uniformly relevant were excluded. Together, these criteria should allow for reasonable and reproducible comparison or ranking of candidate indicators. These criteria are used as a qualitative guide to discuss the metrics applied to the electronics-recycling case.

A shortlist of indicators includes the following.

Useful: (1) addresses a clear goal; (2) simple-specific; (3) diagnostic; (4) comparable.

Robust: (5) subjective elements explicit; (6) reproducible; (7) nonperverse; (8) quality data available.

Feasible: (9) cost-effectively measurable.

* Corresponding author phone: (617)253-4258; fax: (617)258-7471; e-mail: kirchain@mit.edu.
While some of these criteria may appear obvious, many suggested indicators do not meet these standards. Those designing and selecting indicators would do well to keep a set of criteria on hand for reference. A brief explanation of each criterion follows.

**Useful.** 1. **Addresses a Clear Goal.** Stating a goal for an indicator ensures that there is a clear understanding of the desired change in the indicator (increase or decrease) (2, 14, 15). Measurement without a clear associated goal can result in a focus on the metric rather than the underlying improvement objective.

2. **Simple/Specific.** Although it is important for a measurement scheme to be comprehensive, there is a risk that complexity creates a barrier to adoption (16). Simple measures are more readily adopted and implemented.

3 and 4. **Diagnostic and Comparable.** The purpose of measurement is to enable comparison and problem solving. Good metrics facilitate both activities. To be diagnostic, a metric must facilitate the identification of patterns in metric results, development of hypotheses, and the determination of causation that underlies differences in metric values. To be comparable, a metric must be consistent and compatible with other relevant indicators. Consistency allows benchmarking, whereas comparability enables a broader level of analysis. Achieving comparability often involves some form of standardization.

**Robust.** 5. **Subjective Elements Explicit.** In a broad assessment like sustainability, multiple contributing factors must be evaluated simultaneously. Addressing the inherent tradeoffs necessitates the introduction of subjective or normative elements into the evaluation. Practitioners of life-cycle assessment (LCA) familiar with this issue have made some attempts to correct for this directly (13), while some LCA tools allow the user to choose multiple weighting systems to check sensitivity of the results (17).

**Feasible.** 6. **Reproducible.** Reproducibility requires clear boundary definitions, calculation, and data collection methods. As a more general test of robustness, the level of cross-metric consistency can be compared. Do metrics that are purported to measure the same phenomena give consistent results? Differences between measures’ results may be useful if the context for their divergence is understood.

7. **Nonperverse.** There are three primary ways in which a metric may generate perverse information: (1) A metric addresses an ambiguous goal where either directional trend could be considered beneficial. For example, if abatement expenditures are used as a measure of environmental performance, increasing expenditures on abatement could indicate either a worsening of environmental performance (requiring more effort to fix) or simply that there is newfound attention to an issue. (2) A metric encourages actions with unintended negative consequences or media shifting. For instance, if a primary metric is mass of toxic solid waste generated, improvement may be inferred through a shift to toxic air emissions. (3) In the rare case, a metric encourages actions that actually oppose the original goal.

8. **Quality Data Available (Accurate and Updated).** For a number of indicators suggested in the sustainability literature there is no data to conduct the measurement or the available data is inaccurate or rarely if ever updated. This could be because the topic was not previously a concern, because data collection was deemed not cost-effective, or because collection of the necessary data is simply not feasible. A measure should not be rejected simply because quality data is not currently available, if it can be shown that the required data is feasible and cost-effective to collect.

9. **Cost-Effectively Measurable.** While this seems the most self-evident, there are number of aspects of sustainability for which measurement is cost prohibitive or otherwise potentially infeasible, including resource availability, impact on biodiversity, and social welfare. In ad-
prehensive supply chain information on either material flows or environmental impacts.

IV. Existing Measures

Metrics affect the behavior of decision makers and the systems within which they work. Poorly selected metrics can cause performance to deviate from intended goals. It must therefore be asked whether existing recycling industry metrics (Table 2) are effective in moving toward sustainability and whether alternative metrics could better evaluate electronics-recycling efforts.

The usefulness of a measure depends on the needs of the stakeholders using the metric. While the ideal metric would be useful at all levels, could be aggregated, and be valuable for cross comparisons as well as real-time decision making, this is rarely achievable in practice. Original equipment manufacturers (OEMS), recycling operators, and regulators are each key stakeholders within the EOL electronics system, but each is concerned with different aspects of the system. For example, manufacturers have developed design guidelines for improving product recyclability and indicators that attempt to quantify this recyclability (24–31). In contrast, a recycler needs process (not product) based measures, such as mass processed per person-time, to be able to benchmark and evaluate aggregate operational performance and understand the impact of input, process, or demand changes. Whereas “recyclability” measures assess the effort involved in recycling (assuming a particular recovery process is followed), the mass recovery metric assesses the result of actual recycling efforts. Each provides key insights in different contexts, and ideally lessons from actual recycling operations would inform product design efforts (32).

Although current metrics are useful for operational decisions, they provide limited insights into two aspects of resource sustainability critical for this industry: recovery of economic and environmental value and reduction of emissions from end-of-life practices (36). Currently, to assess environmental performance of recycling activities, all stakeholders focus on the most easily measured—mass fraction of material going to landfill (37–39). While this goal appears to be uniquely congruent with resource sustainability, prior research indicates that this may not always be the case (36). Possible reasons include when any of the following apply:

- Mass is a weak indicator of environmental impact (e.g., impacts of toxicity).
- The material “recovered” by the first recycler is landfill or destroyed by a subsequent processor.
- The added cost of recovering additional material reduces collection such that net impact is increased.
- The effort (expense, energy, environmental impact) put into recycling the material exceeds the benefits reaped from its recovery.

It appears therefore that current system metrics do not necessarily identify trends toward sustainability improvement and may skew decision making away from underlying goals. This issue is indicative of the problem of using an overly simple measure. Exacerbating this shortcoming, recycler performance is typically assessed only on practices at the reporting facility, without taking into account key factors influencing recycling performance (e.g., inflow quality, product mix, or downstream material yields). Ideally this would be addressed by detailed system-wide analyses. Unfortunately, time and expense preclude this approach beyond the occasional case study. While one-off studies provide valuable insights, they do not fill the operational need for continuous feedback and regular benchmarking. The optimal approach is then to evaluate (and reassess) potential metrics in order to choose the simplest possible set

FIGURE 1. Schematic of EOL electronic material recovery pathways.
of metrics for the given need, while retaining an awareness of the gaps in vision that these metrics present.

The previous paragraphs describe the benefits and limitations with the current set of measures. Some measures are more useful for strategic decision making, while others are useful for tactical decisions. However, none address a broad set of resource concerns. Increased use of activity-based costing (ABC) could improve the data availability situation by facilitating better tracking of the true cost of processes and products (because costs are tied to production activities). However, ABC does not itself address the embedded value of EOL materials. The following sections describe the use of three relevance-weighted mass-based indicators. The efficacy of these metrics is compared in terms of the primary criteria: the usefulness, robustness, and feasibility of the metric.

IV.A. Evaluating Existing and Alternative Metrics. There are many possible weighting schemes that could be applied to augment a mass-based recovery measure. Because each represents key elements of a system’s resource sustainability, it is likely that a set of simple weighted metrics would be needed. The use of value, energy, and environmental impact as relevance-weighting measures is examined here.

1. Value-Based Recyclability. Economic value provides a signal of the utility of a good or service. For materials, value reflects (1) quality, (2) the cost of production or use (including energy consumption), and (3) scarcity rents for current use of that resource (40). As such, value-weighted assessment provides significant information about the effectiveness with which resources are reclaimed and returned to productive use, providing an indicator of both retained quality (41) and environmental impact (42).

The “recyclability” index developed in ref (41) uses this concept to examine resource recovery. The basic assumption of ref (41) is that “the recyclability of materials will be reflected by their monetary value”. This leads directly to the recyclability index \( Vp / Vm \), where \( Vp \) ($/kg) and \( Vm \) ($/kg) represent the market value of secondary and primary material, respectively (Figure 2).

Although originally applied to examine aggregate material flows, this metric can also serve as a weighting factor in the analysis of the recovery effectiveness of a recycling operation or industry leading to the value-retention weighted mass recovery index:

\[
\text{ValueRetentionIndex} = \frac{\sum_j Vp_j m_j}{\sum_k Vm_k m_{ki}}
\]

where subscripts \( i \) and \( j \) represent values for inflows and outflows, respectively, and \( k \) represents the \( k \)th embedded material in a given flow. \( m_j \) represents mass of a given flow.

This measure indicates those industries which are able to reclaim not only mass but also significant EOL value (\( Vp \)) relative to the value of materials which were originally consumed (\( Vm \)). The value-retention measure gives insight into aggregate system recovery effectiveness and provides a quantitative accounting of down-cycling. However, as the only indication of incoming material quality is primary material value (\( Vm \)), this metric provides only weak indication of the effectiveness of individual operators. Nevertheless, the value-retention recyclability measure should provide useful insight of aggregate system performance to electronic manufacturers and interested regulatory agencies.

To assess the performance of individual recyclers, the effect of varying incoming material quality must be accounted for. In a value context, this can be accomplished by integrating the residual value (\( Vr \)) of incoming EOL materials (i.e., the price paid by the recycler). The impact of recycling processes is thus characterized by the quantity (\( Vp - Vr \)) which is the value added by recycler activities. To compare various material streams, this quantity needs to be normalized. Conceptually, a useful normalization option is the quantity \( Vm - Vr \), which is the maximum possible value-added that a facility could achieve. Combining this with mass recovery information yields the value-added weighted mass recovery index:

\[
\text{ValueAddedIndex} = \frac{\sum_j Vp_j m_j - \sum_i Vr_i m_i}{\sum_k Vm_k m_{ki} - \sum_i Vr_i m_i}
\]

This measure works well except for cases with significant device/component reuse. In practice, this can be accommodated through segregation of data for product reuse and material recovery, with \( Vm \) as a measure of embedded use value or material value, respectively.

TABLE 2. Electronics-Recycling Measures: Information Is from References Cited and Primary Interviews with Stakeholders

<table>
<thead>
<tr>
<th>actor (right)</th>
<th>economic [quantitative]</th>
<th>mass based [quantitative]</th>
<th>additional criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>performance metric (below)</td>
<td>$/mass recycled or collected</td>
<td>% material collected, % recycled/product (33), material recycling rate (34)</td>
<td>requirement checklists</td>
</tr>
<tr>
<td>government [product or system]</td>
<td>$/mass recycling fees, system costs, disassembly time/product</td>
<td>total mass recycled, % recycled material/product, % “recyclable” material/product</td>
<td>LCA, recyclability (35)</td>
</tr>
<tr>
<td>OEM [product, process, or system]</td>
<td>mass/month, cost/mass processed, mass processed/time, recovered $/mass processed</td>
<td>% recycled or % diverted from landfill</td>
<td>facility based: energy use, EH&amp;S audit, ISO14000</td>
</tr>
<tr>
<td>recycler [process, or system]</td>
<td></td>
<td></td>
<td>“no export”, “no prison labor”</td>
</tr>
</tbody>
</table>

**FIGURE 2.** Value-based recyclability (\( Vr \) is used only in the value-added measure).
2. Other Impact-Weighted Metrics. While value-based weighting schemes give indirect insight into environmental issues (e.g., energy use, resource degradation), some externalities are not embedded in market derived values (e.g., emissions). To ensure unequivocal progress toward resource sustainability, measures are needed to address these discrepancies. A variety of life-cycle assessments (LCAs) have addressed EOL impacts for products and different recycling scenarios (43, 44). Conducting full LCAs on EOL material flows is not practical for making ongoing operational decisions; however, life-cycle approaches may form the basis for weighting schemes that augment mass-based assessments. Along these lines QWERTY (quotes for environmentally weighted recyclability) suggests the use of LCA weighting for product-specific assessments (26); the sustainability target method (STM) presents an “absolute” sustainability measure for material and product recycling (45); input–output LCA uses input–output analysis to augment material flow information (42). Exergy has also been suggested as a way to evaluate resource recovery (46–48); and energy analyses have been done comparing end-of-life strategies (49, 50). Each of these methods has particular strengths and weaknesses. However, the common problem is one of achieving an acceptable level of resolution with realistic data intensity. In addition, there is no standard accepted method to account for the impact from use of nonrenewable resources although some have specifically discussed the issue (51).

Energy and environmental impact (EI) weighted measures for recycling facilities are applied in (52) and described below. The energy and EI measures do not supplant the measures described above. Rather, they serve to illuminate common issues and clarify strengths and weaknesses of these approaches relative to the value measure, with specific attention given to the level of reappraisal of recovered material. The energy and EI-weighted mass indexes have the same form, thus the following discussion of methodology applies to both measures.

Like the value metrics above, which focus on the “value retention” or value saved from material recovery, the energy metric for a facility would focus on the energy “saved” due to material recovery. However, whereas value does this implicitly, the energy (and EI) measure must do so explicitly. In general terms, the “energy saved” is the difference between the energy used to create a marketable commodity out of (1) a recycled “ore” or (2) a mined ore (or other traditional source). Unfortunately, secondary materials (aside from metals) are rarely complete substitutes for primary materials.

The value measure assumes that substitutability determines market prices, with a lower value reflecting poorer material properties (41). In contrast, the energy and EI measures must explicitly compare the secondary commodity with the primary commodity that would otherwise be used for a specific application. For example, a recycler trying to determine the most energy-saving way to recycle leaded glass from cathode ray tubes (CRTs) would make the following comparison (Figure 3). The leaded glass can (1) be sorted, crushed, and cleaned, then sold to make new leaded glass or (2) be crushed and sold as-is to a lead smelter as flux (53). For option (1) the secondary material replaces new glass; the comparison is between the energy required to make either end-of-life material or virgin silica, lead, and other ingredients into new glass. For option (2) the material is replacing either sand or lead-containing silica wastes. In this case, the differences in energy use are transportation energy and any impact on smelter efficiency.

Following this logic, the energy-weighted mass recovery index for a facility would be as follows:

$$\sum \sum (E_{p_{ij}} - E_{r_{ij}}) m_{ij}$$

where $E_p$ = total energy use to make the primary material for use in application 1, $E_r$ = total energy use to prepare the secondary material for use in application 1, and $E_{max}$ = the total energy use to make the most energy-intensive primary material for which the secondary material could be substituted (for instance, in Figure 3, $E_{max}$ would be equal to $E_p$, rather than $E_r$).

The environmental impact, or EI-weighted mass index, parallels the energy measure. As with the energy measure, the EI measure is the EI-weighted sum of the material flow masses based on the net environmental impact of making secondary commodities available and displacing use of the corresponding primary commodities. The metric used is as follows:

$$\sum \sum (E_{p_{ij}} - E_{r_{ij}}) m_{ij}$$

where $E_p$ = total energy use to make the primary material for use in application 1, $E_r$ = total energy use to prepare the secondary material for use in application 1, and $E_{max}$ = the total energy use to make the most energy-intensive primary material for which the secondary material could be substituted (for instance, in Figure 3, $E_{max}$ would be equal to $E_p$, rather than $E_r$).

V. Assessing Metrics in Use

The preceding discussion highlights the conceptual advantages and disadvantages of three recycler performance metrics. To understand the practical value of these metrics for improving performance assessment, they were exercised against representative data from recycling facilities. Three facility provided data on commodity flows for 2003; the other five facilities interviewed did not consistently perform even this level of data collection.

The three case facilities each take in a full range of EOL electronic products. Nevertheless, each facility has a distinct customer focus; focusing on telecom, manufacturing (or prompt scrap), or product reuse. Table 3 provides an overview of the value characteristics of the material processed by the three study facilities.

For personal computer manufacturing, significant environmental impact and cost comes from the ultrarefining of raw materials (49). Component reuse saves this processing, whereas material recovery, generally, does not. Ideally metrics should not only resolve [material recovery performance] but also identify the significant benefits from component refurbishment and reuse. Nonetheless, ultimately all electronic
material is destined for material recovery or disposal. The measures presented here were thus applied to evaluate the material recovery fraction only. However, the analysis of the use of these metrics for reuse was also explored in (52). The substantial value difference between reuse and recycling indicated by these measures was consistent with an analogous energy analysis (49).

Data collection and analysis are described further in refs (52, 54) and in the Supporting Information with results presented in Table 4. While all three facilities expressed similarly high levels of mass recovery (equal to, or greater than 98%), the other measures highlighted substantial differences between the ability of the three facilities to repurpose the material effectively.

On the basis of the value measures, none of the facilities repurpose material as effectively as indicated by the mass measure. However, both facilities 1 and 2 generate material recovery streams retaining some 40% more value than those from facility 3. When the condition of incoming materials is considered, (with the relative value-added measure) this discrepancy is even greater. Facilities 1 and 2 add significantly more value to their incoming flows than does 3. The energy and EI results imply more effective recovery than the value measures, but still less effective than indicated by the mass measure. Like the value metrics, the energy and EI metrics provide significant additional resolution. However, the relative ranking of facilities differs from the value measures. Facility 3, which showed the lowest value, has the highest energy and EI results, whereas the energy and EI measures differ in which facility comes in second. The two primary mechanisms that could explain these differences:

1. The relevance of particular materials differs between the perspectives of value, energy, and EI.
2. The EI and energy measures assume (a) different processing intensity than actually occurs, whereas the value measures implicitly considers processing costs; (b) a single recycling pathway for each material, whereas the value measure captures the actual level of application for each commodity; (c) full recovery of materials, whereas value accounts for the degree of recovery actually occurring; (d) full recovery of material properties, whereas value accounts for degree of material degradation.

If these differences are due to mechanism 1, using multiple measures is appropriate, as the measures provide complementary information. If instead these differences are due to mechanism 2, then the value measures are most likely to provide an accurate picture of facility recovery given existing data availability. The efficacy of these measures is assessed below.

VI. Assessing Measure Effectiveness

To determine the potential effectiveness of a metric one must explore the usefulness, robustness, and feasibility of the measure as applied. In this case, primary consideration is given to the balance of their informative value and cost-effectiveness.

Measure usefulness depends on whether they are simple, diagnostic, and address a clear goal. What is diagnostic depends on the goal, and goals may differ by stakeholder. In the case of electronics recycling, customers concerned only with liability reduction may find the mass recovery measure sufficient. However, if other goals are intended by the recycling activity (e.g., reduced environmental impact and resource conservation), then mass is inadequate. For these goals, the three metrics evaluated can provide more comprehensive insights.

The value-based metrics provide significant additional resolution concerning the material recovery practices of the three operators. Two facilities retain and add more value, preventing material down-cycling and, presumably, the attendant demand for additional primary material. The EI and energy measures also provide additional resolution, but in contrast to the value measure, these measures do not inherently differentiate between subsequent uses based on information available at the recycler. Subjective elements within EI measures are generally not explicit, reducing measure robustness. Without data on material pathways and degree of loss it would be difficult to ensure reproducibility of these measures.

Measure robustness requires the availability of quality data, and feasibility requires a metric be cost-effectively measurable. While the energy and EI measures have the potential to be highly informative, their data intensity poses
the default methods; Centre for Environmental Assessment of Products and Material Systems; Chalmers University of Technology, Technical Environmental Planning. 1999.


(15) WBCSD. Measuring Eco-Efficiency; 2000.


(18) Puckett, J.; et al. Exporting harm: the high tech trashling of Asia; Basel Action Network (BAN); Silicon Valley Toxics Coalition (SVTC): 2002.


(34) USGS. Metals-Recycling; 2002.


(44) LCA study of the product group personal computers in the EU Ecolabel scheme; Atlantic Consulting: London, 1998.


Received for review May 16, 2005. Revised manuscript received September 29, 2005. Accepted November 2, 2005.

ES050935L