Sustainability in Higher Education: Analysis and Selection of Assessment Systems

Antonios Maragakis1 & Andy van den Dobbelsteen1

1 Faculty of Architecture, Delft University of Technology, Delft, The Netherlands

Correspondence: Antonios Maragakis, Faculty of Architecture, Delft University of Technology, P.O. Box 5043, 2600 GA Delft, The Netherlands. E-mail: antonios.maragakis@gmail.com

Received: February 8, 2015   Accepted: March 4, 2015   Online Published: April 2, 2015
doi:10.5539/jsd.v8n3p1          URL: http://dx.doi.org/10.5539/jsd.v8n3p1

Abstract
There is a noticeable increase in interest with regards to sustainability in higher education. As institutions investigate, implement and market sustainability efforts, there is a myriad of sustainability assessment methodologies currently available. Although these assessment systems were created with the intention of helping sustainability in higher education institutions, they have ultimately led to an assortment of standards being used by institutions which do not help students and faculty assess the level of sustainability uniformly between institutions.

This paper combines relevant literature on sustainability assessment with empirical data to suggest an ideal assessment method to be used as the basis for a universal tool. It was found that the STARS system was the most suited system to be used as a basis for a future universal assessment tool.

Keywords: sustainable development, higher education, universities, sustainable education, sustainable assessment

1. Introduction
Over the years there has been an increased focus on sustainability in higher education. Policy makers (UNESCO, 2011) and students (Bone & Agombar, 2011) have placed a significant emphasis on sustainability, while institutions have responded by actively implementing sustainable initiatives. The term sustainability still has not been unequivocally defined; nonetheless, a plethora of universities are claiming to be sustainable in some way, shape or form. This raises the question of how to define and assess sustainability in higher education institutions.

Numerous publications (Ryan et al., 2010; Glasser 2009; Patrick et al., 2008; Perna et al., 2006) have investigated and analyzed the various assessment systems and with inventories of university initiatives currently available. However, none have gone so far as to suggest which assessment system would be best suited for standardized use. This is seen as a controversial step as the choice will have far-reaching implications in theory and practice (Shriberg, 2002).

In general, there has been resistance to standardizing assessments and/or rating institutions on sustainability. AASHE’s STARS, among other prominent sustainability tools, clearly makes the delineation that it is an assessment tool and in no way a rating or ranking system. It can be argued that this apprehension for standardizing sustainability within institutions neither benefits sustainable practices nor helps stakeholders (students, academics and administrators) identify the level of sustainability in an institution.

A standard sustainable assessment system would provide the basis for sustainability in an institution while also providing a standard for sustainability marketing. Selby et al. (2009) came to two very important conclusions about sustainability and marketing:

1) Sustainability messaging tends for the most part to treat ‘sustainability’ as synonymous with ‘environment’.
2) Rigorous institutional engagement with marketing sustainability credentials provides a beneficial feedback loop that deepens and embeds the commitment and adherence by administrators, academics and students.

These two conclusions make a clear case that a standard assessment would benefit by assuring that ‘sustainability’ is not misrepresented as a solely environmental issue while also assisting with the deepening of sustainability within the institutions culture.
Apprehension for standardizing assessment of institutions is directly opposes the needs of some higher education stakeholders. Maragakis & Dobbelsteen’s (2013) empirical study showed that 95% of potential or current students, staff and management in higher education agreed that there was a need for a uniform rating system. This demand would explain the rise of certain private initiatives, such as Princeton’s Guide to 311 Green Colleges (The Princeton Review 2011). By continuing to not act on creating a standardized system, scholars and practitioners may lose the ability to shape assessment and rating criteria for sustainability and could give rise to popular, yet potentially ineffective, methods of assessment that appeal to institutional stakeholders.

In an attempt to move this issue forward, this paper looks to review the existing literature on sustainability assessment methods and compare it to Maragakis & Dobbelsteeen’s (2013) empirical data in order to provide guidance as to what is the most suitable sustainability assessment system for higher education.

2. Methodology

This research focuses on reviewing the key elements from previous literature in order to provide a robust and complete framework for assessing the suitability of sustainability assessment systems. Specifically this research looks to extrapolate the key parameters used to rate sustainability assessment systems and combine them into a more comprehensive system in order to assess current systems in order to determine the most appropriate for use as a universal system.

Once a comprehensive list is created from assessment, a selection of sustainability assessment tools identified by the literature as being ideal will be selected and subjected to evaluation. The evaluation will focus on the framework of each sustainability assessment tool and will award marks of “Yes”, “No” and “Partially” in reference to fulfilling the evaluation criteria. To limit bias, each mark will be justified with reference to the sustainability assessments framework.

2.1 Research Questions

The primary question of this research is to seek the parameters and/or criteria that other authors have used or suggested to assess sustainability assessment tools in higher education.

The second research question is if the combining of these parameters can provide a meaningful comparison of assessment systems in order to determine an appropriate system for universal use.

2.2 Approach

A literature review will be conducted in order to identify the parameters and/or criteria in order to perform a review of existing sustainability assessments. A selection of current assessment tools will be selected based on the result from other literature review and from the survey results of Maragakis & Dobbelsteen (2013).

2.3 Literature Selection

The general topic of sustainability assessment has been exhaustively studied, perhaps better studied than sustainability itself (Kates et al., 2001). Sustainability assessments have been created for a wide range of international, national, professional and personal initiatives. Everything from sustainable farming to sustainable corporations has some methodology and guidance that is provided for from various sources. An example of this multi-tiered and growing market can be exemplified by the corporate sustainability assessment methods. Some organizations claim to assess the most sustainable corporations in the world, others assess the most sustainable corporations nationally (based on country), while others provide professional third-party sustainability assessment and finally other provide corporate sustainability assessment based on the niche in which that corporation is operating.

In addition to the existence of these sustainability assessments there also have been countless studies on the usefulness, comparison, categorization, etc. of these methods so as to provide discussion and improvement of these methods.

It is noted that the cores of all these assessments tend to be similar in nature. They all attempt to quantify sustainable initiatives using a variety of predefined or proprietary indicators. They all share a level of acceptance and criticism and they all aim to promote sustainability (although the term itself seems to vary greatly). With this in mind, all of these assessments, and the literature associated with them, would be potential sources for review. However, this would be a daunting task and would not necessarily assist in the purpose of this paper.

Thus, the scope of this paper is limited to publication related directly to sustainable assessments in relation to higher education. The literature specifically dealing within this scope is limited and provides key insight into the existing systems currently being used. The literature on this specific subject is assumed to have drawn from the existing knowledge on sustainability assessment allowing this paper to focus on determining the best possible
system to be used specifically for higher education.

For this assumption to be successful, significant literature was selected in such a way to provide for a specific review of comparable publication that represents the core of this paper. Of all publications studied from the last decade only two have dealt directly with the strengths and weaknesses of assessment systems for higher education. These are:


These two pieces of peer-reviewed work are assumed to provide guidance for the creation of a scholarly approach to comparing assessment methods. Their methodologies and results will be utilized in this paper, in conjunction with empirical data, to provide guidance for a standard assessment system.

3. Review of Literature

3.1 Institutional Assessment Tools for Sustainability in Higher Education: Strengths, Weaknesses, and Implications for Practice and Theory (Shriberg, 2002)

Prior to starting the review, it is noted that this publication is outdated in respect to the latest assessment tools and trends within the niche of sustainability in higher education. Although deficient in comparing new methods, the foundations of Shriberg are still relevant and useful for this paper.

Shriberg’s paper is arguably the basis for debate on the feasibility of a universal assessment system. The author touches on some of the key points that limit the implementation of a standardized system. Some of findings are:

- An effective tool needs to accurately portray the institutions current status but also integrate motivations, processes and outcomes in a comparable, understandable and calculable way.
- Tools capture baselines but do not provide mechanisms for comparisons.
- Tools converge on the parameters of:
  - Decreased throughput,
  - Incremental and systematic progress,
  - Sustainability education as a core function,
  - Cross-functional reach, and
  - Cross-institutional action.
- A universal tool debatably will overlook contextually important information.
- Sustainability ranking has been avoided due to resistance from administrators and others to ordering campuses on a subjective concept and goal.

The other analysis of the actual strengths and weaknesses of the eleven institutional tools available at the time seem to be subjective and provides more of a narrative opinion piece which is loosely connected to criteria proposed by Orr (2000) and the authors parameters, which are presented in the table below.

<table>
<thead>
<tr>
<th>Core issues of ecologically, socially and fiscally sustaining a society and campus</th>
<th>Ideal sustainability assessments</th>
<th>cross-institutional action</th>
</tr>
</thead>
<tbody>
<tr>
<td>What quantity of material goods does the college/university consume on a per capita basis?</td>
<td>Identify important issues</td>
<td></td>
</tr>
<tr>
<td>What are the university/college management policies for materials, waste, recycling, purchasing, landscaping, energy use and building?</td>
<td>Are calculable and comparable</td>
<td></td>
</tr>
<tr>
<td>Does the curriculum engender ecological literacy?</td>
<td>Move beyond eco-efficiency</td>
<td></td>
</tr>
<tr>
<td>Do university/college finances help build sustainable regional economies?</td>
<td>Measure processes and motivations</td>
<td></td>
</tr>
<tr>
<td>What do graduates do in the world?</td>
<td>Stress comprehensibility</td>
<td></td>
</tr>
</tbody>
</table>
The author’s review of the assessment methods, based on Table 1, can provide guidance for this paper. The time lapse since the printing of the article discussed has however seen the revision of the existing systems as well as the introduction of new systems ultimately making the Shriberg’s review outdated for the purposes of this paper.

3.2 Identifying Strengths and Weakness of Sustainable Higher Educational Assessment Approaches (Saadatian et al., 2011)

This publication, in contrast to Shriberg’s (2002), is directly relevant to this paper as it is relatively recent and deals with the predominant assessment methods currently available. Due to the recent nature of this research, it is assumed that the data and conclusions are still relevant and can assist in the development of this paper.

The authors took a different approach to measuring the strengths and weaknesses of the assessment systems. They utilized two theories and three criteria that were used as the basis of their evaluation. The theories were that of triple bottom line (Elkington, 1997) and that of avoiding subjective judgment (Connolly et al, 2000), which provided for the criteria of comprehensiveness, novelty and popularity.

The aforementioned theories and criteria formed their parameters of judgment. By conducting a literature review, archival review, interviews and research on internet popularity, the authors concluded that STARS and CSAF were the top scoring in terms of satisfying each of the theories and the three criterions.

Saadatian’s work should be applied cautiously, however, as several lapses were identified in the methodology and rigor of the tests. An example is the research conducted on the amount of Google search hits. No exact framework and keywords where provided, effectively eliminating the ability for other researchers to reproduce the results independently. Other critical lapses in presentation and academic rigor, for instance poor referencing and serious grammatical errors, were also noted and necessitate the need for the cautious use of the results.

3.3 Conclusions from the Literature Review

Both pieces of literature are a testament to the difficulties and subjectivity involved with the methodological analysis of the various assessment methods. Due to the vagueness of the term sustainability, along with the limited consensus on quantifiable indicators, there seems to be a certain amount of bias in both publications.

For example, Shriberg (2002) looks to assess the effectiveness of the actual metrics of the assessment methods beyond just the triple bottom line while Saadatian et al., (2011) assumes that the triple bottom line is an effective metric for sustainability and focuses on other criteria to judge the effectiveness of the assessment methods.

In both cases, results can be drawn as to useful methodologies and approaches. Shriberg (2002) offers literature on methods of actually assessing the usefulness of metrics used in the sustainability assessments in higher education. Saadatian et al. (2011) on the other hand explore other dimensions on the effectiveness of assessment methods beyond just the metrics that encompasses popularity and acceptance (preferences) of individuals involved in sustainability with higher education.

The research of Saadatian et al. (2011) needs to be used cautiously as there are some fundamental questions as the quality of the research. However, the results of STARS as one of the highest ranking assessment methods is also in line with other literature from GreenerU (2010), which also found that STARS is one of the most prominent external assessment system because of its comprehensive and holistic nature.

This literature review has provided some key metrics for further analysis. Utilizing Orr’s (2000) criteria, the triple bottom line in relation to higher education institutions can be explored in depth for each assessment method. Shriberg’s (2002) criteria provide for a more in-depth review of cross-institutional metrics beyond just the social, economic and environmental parameters. While most of the criteria of Saadatian et al. (2011) have been addressed with the previous two metrics, the metric of popularity has not, providing a significant factor for determining the effectiveness of a system.

4. Review of Empirical Data

In late 2012, Maragakis & Dobbelsteen (2013) conducted a broad survey of assessment systems within higher education that provided some useful empirical data. These results provide a first step in quantifying the needs of stakeholders (students, staff and management). One of the needs identified, and indeed motivation for this research, was that 95% of respondents agreed that institutions need to be uniformly rated.

The results of the 203 survey respondents showed that the STARS, Princeton Review Green Rating and College Sustainability Report card were the most popular assessment methods, with STARS being the most popular of the three.

Of all the assessment methods, STARS was the clear preference of stakeholders as the most appropriate metrics
for assessing sustainability within higher education.

5. Discussions from the Literature and Empirical Data Review

One of the most important conclusions from the literature and empirical data review is that each research focused on a different set of assessment systems. This does not affect the usefulness of Shriberg’s (2002) findings as his research primarily provides a comprehensive methodology for assessing assessments rather than explicit results. The different sets of assessment systems do however limit the ability of directly comparing the results of Saadatian et al. (2011) with Maragakis & Dobbelsteen (2013).

This inability to directly compare the two research publications also raises questions as to the validity and comprehensiveness of each of the publications. Maragakis & Dobbelsteen’s results have provided a section in their data collection for “Other” assessment systems which proved to be statistically insignificant, thus eliminating some of the uncertainty of not including other assessment systems, such as AISHE and CSAF. However, Saadatian et al. have not allowed for any potential assessment omissions and significant questions are raised as to the validity of the results. Even though the results are partially supported by GreenerU (2010), it should be noted that GreenerU is also an inflexible analysis based on a specific set of assessment methods and it could be argued that this raises more questions on the validity and comparison of the two results.

It should nonetheless be noted that STARS is consistently ranked as one of the top systems. Although there is no way to compare the various research results directly, it can be argued that STARS’s superiority has been proven both against various methods and through different research methodologies. While this is not a definitive result it does provide for the formation of a trend that STARS is currently the most popular system.

Since the literature and data cannot be directly compared, all the results will need to be considered in this analysis. Saadatian et al. (2011) concluded that STARS and CSAF were the highest ranked assessments based on the research conducted. Based on survey results Maragakis & Dobbelsteen (2013) concluded that STARS was the best assessment method.

It should be noted that GreenerU (2010), which was referenced but not assessed, concluded that STARS and the College Sustainability Report Card were the most popular. As the College Sustainability Report Card has since been suspended, it will not be considered in this research.

6. Comparing Assessment Methods

Based on the review, STARS and CSAF are the candidates for most appropriate sustainability assessment system to uniformly rate higher education institutions. A comparison of these two methods using the criteria of Orr (2000), Shriberg (2002) and Saadatian et al. (2011) was conducted using a simple ‘Yes’, ‘No’ or ‘Partially’ measurement. An explanation for each criterion ranking is provided for after Table 2.

Although there is a depth of knowledge that exists regarding criteria to judge sustainability assessments, this research has actively chosen to focus on significant work that has dealt solely on this subject. This approach was taken to use a peer-reviewed framework that would promote an unbiased, comprehensive and non-overlapping comparison. Weaknesses in the approach have been noted and it is expected that as new research continues to be published, these criteria may need to be revisited.
Table 2. Comparison of STARS and CSAF based on review criteria

<table>
<thead>
<tr>
<th>Core issues of ecologically, socially and fiscally sustaining a society and campus (Orr, 2000)</th>
<th>STARS</th>
<th>CSAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>What quantity of material goods does the college/university consume on a per capita basis?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>What are the university/college management policies for materials, waste, recycling, purchasing, landscaping, energy use and building?</td>
<td>Yes</td>
<td>Partially</td>
</tr>
<tr>
<td>Does the curriculum engender ecological literacy?</td>
<td>Yes</td>
<td>Partially</td>
</tr>
<tr>
<td>Do university/college finances help build sustainable regional economies?</td>
<td>Partially</td>
<td>Partially</td>
</tr>
<tr>
<td>What do graduates do in the world?</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ideal cross-institutional sustainability assessments (Shriberg 2002)</th>
<th>STARS</th>
<th>CSAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify important issues</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Are calculable and comparable</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Move beyond eco-efficiency</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Measure processes and motivations</td>
<td>Yes</td>
<td>Partially</td>
</tr>
<tr>
<td>Stress comprehensibility</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saadatian et al. (2011)</th>
<th>STARS</th>
<th>CSAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popularity</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Totals</th>
<th>Yes</th>
<th>9</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partially</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

For the first criterion, “What quantity of material goods does the college/university consume on a per capita basis,” a review of both STARS and CSAF offers multiple areas that touch on this field. However, STARS directly deals with this in Operational (OP) Credit 17: Waste Reduction and categorizes the waste on a per capita basis. CSAF offers multiple indicators that cover this topic; however it fails to provide a per capita figure.

For the second and third criteria, “What are the university/college management policies for materials, waste, recycling, purchasing, landscaping, energy use and building” and “Does the curriculum engender ecological literacy” respectively, both the STARS and the CSAF provide indicators dealing with these subjects, however there is a key difference with the measurements that sets STARS apart from CSAF. CSAF proves to be an excellent tool for measurement while STARS provides both an excellent tool for measurement while also providing guidance. For example, the policies section within the CSAF is based upon the percentage of sustainable policies as compared to the total number of policies within an institution. Although this may provide a more robust way of gaining credit for sustainable policies, STARS looks to actively promote specific verbiage within the various policies and awards credit on a “per section” basis than as an institution as a whole. The same is true with eco-literacy as the STARS method has it integrated in various facets of the educational process while the CSAF approaches it tends to be much more vague and robust.

For the fourth criterion, “Do university/college finances help build sustainable regional economies,” it is arguable that neither method fully embodies the regional aspect. STARS provides some verbiage in various sections that promote regional integration, however falls short of providing anything of actual value with regards to this criterion.

For the fifth criterion, “What do graduates do in the world,” it is unfortunate to note that neither assessment method has post-graduation metrics.

For the sixth criterion, “Identify important issues,” the term “important” is somewhat debatable. This being kept in mind, both methods identify important issues with regards to sustainability. STARS groups the requirements
in four overarching themes while the CSAF provides 169 indicators. In both cases, it is arguable as to how “important” the actual composition of each measurement is, however it is apparent that there is substantial effort and thought in identifying the “important” issues.

For both the seventh and the eighth criteria, “Are calculable and comparable” and “Move beyond eco-efficiency,” it is apparent that both assessment tools provide their own unique, but effective way for calculating and comparing a robust set of requirements that move well beyond just eco-efficiency. The STARS system offer a calculable and comparable system that is based both on quantitative and qualitative information. The CSAF offers hard metrics based on 169 indicators that provide an overall quantifiable measurement that takes into account both quantitative and qualitative information. Both tools move well beyond just eco-efficiency, but it is noted that a significant portion of both tools focus the bulk of their metrics, in all facets of the institution, on eco-efficiency.

The ninth criterion, “Measure processes and motivations,” provides a slight advantage for the STARS method. While both tools measure process and motivation, STARS provides a more comprehensive and supporting methodology that supports and measures qualitative progress as compared to CSAF’s more quantitative approach. This is primarily an issue when trying to deal with motivations as these are more qualitative factors that may be hard to quantify.

For the tenth criterion, “Stress comprehensibility,” there are no doubts that both systems, in their own way, stress comprehensibility.

For the eleventh and final criterion, “Popularity,” it is clear that both tools are popular. However when trying to say which is more popular, a case can be made that STARS is the most popular of the two. Although both ranked high on Saadatian et al. (2011), in Maragakis & Dobbelsteen (2013) CSAF was indirectly proven to be popular. Although CSAF was not included directly in their survey set, the “Other” category, which could reference CSAF indirectly, was not statistically significant in the results.

7. Conclusions and Discussions

7.1 Interpretation of Results

The comparison found in this research is a first step in showing that STARS may be the most suitable basis for a uniform rating of sustainability in higher education institutions. Based on criteria set forth in previous research as well as empirical survey results, it is clear that STARS is a methodology that could be used as the cornerstone for a universal rating system.

Although both STARS and CSAF are useful tools for assessment, STARS is notable a better system. Neither system was perfect and both are comparable, however STARS exceeds in fulfilling nine of eleven criteria proposed in this research, in comparison to CSAF’s ability to fulfill five. Although the criteria were selected to promote an unbiased, comprehensive and

The research also showed that STARS offers a certain level of guidance as well as assessment. Although this was not a specific topic of research in this paper, it is important as institutions that are interested in applying sustainability will have a tool that will provide guidance and measurement.

Finally, based on the data collected by Maragakis & Dobbelsteen (2013), STARS is clearly preferred by stakeholders. It is also noted that, although not conclusive, various pieces of literature have also ranked STARS as one the better assessment tools, adding validity to this research and the data collected.

7.2 Discussion of Method Used for Comparison

The comparison is a first step to combine literature and empirical data to select a universal assessment system for higher education; however the limitations of this research need to be identified.

Firstly, as previously mentioned, this research is based on limited research material that is in many cases empirical, weak or incomparable. There are significant holes within this research resulting from the level of uncertainty in the literature used, especially of Saadatian et al. (2011), and the empirical nature of the survey conducted by Maragakis & Dobbelsteen (2013). These uncertainties could potentially be further researched in order to ascertain if indeed STARS and CSAF are the premier assessment methods to be used as a universal system.

Furthermore, the utilization of Orr (2000) and Shriberg (2002) as criteria is also a limiting factor of this research. Again, as previously discussed, sustainability assessment may be more thoroughly researched than the actual science of sustainability itself (Kates et al., 2001). The assumption that the literature used for this research is a culmination of specific efforts to research sustainability assessment methods in higher education could
unknowingly eliminate other useful criteria that could have affected the results of this research. As a last statement, the actual comparison itself is subject to the bias of the researchers. The ‘Yes’, ‘No’ and ‘Partially’ measurements used to compare the two methods is subjective and based on the interpretation of the researchers. Although most of the measurement results can also be supported by literature (both directly and indirectly referenced by this research), they are still subject to research bias and opinions. For example, are the three levels of measurement selected the most appropriate for this study, or should a scale have been created? What is the quantifiable level of “partially” for each of the eleven criteria? These are some examples of potential bias in the results. But, considering that this research is conducted as an empirical study and aims to provide a starting point for further research, these limitations should be noted and addressed in further research without discounting the relevance of this study.

7.3 Recommendations

Based on the results, it is recommended that further research be conducted on the applicability of STARS as a universal rating system. Although this research has shown it has potential to be the most suitable system for universal use, there are still some concerns and shortfalls of the STARS system that are noticed both in fulfilling the criteria set by this research and in other literature.

7.4 Outlook

Based on the results and recommendations, it is recommended that further research be conducted on the applicability of STARS as a universal system. An analysis of the system, focusing on the strengths and weaknesses, and integration of the data from Maragakis & Dobbelsteen (2013) can provide specific insight on the steps needed to make STARS a universally applicable, and acceptable, tool.

Acknowledgements

Special thanks and acknowledgement is given to Alison Erlenbach for her support on this paper.

References


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).