Ecological Economics 136 (2017) 77–85

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

The Choice of the Sustainability Assessment Tool Matters: Differences in Thematic Scope and Assessment Results

Evelien M. de Olde [a,b,*], Eddie A.M. Bokkers [b], Imke J.M. de Boer [b]

[a] Aarhus University, Department of Engineering, Inge Lehmanns Gade 10, 8000 Aarhus, Denmark
[b] Animal Production Systems Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 25 February 2016
Received in revised form 26 December 2016
Accepted 13 February 2017
Available online 20 February 2017

Keywords:
Sustainability indicators
Tool Development
Framework
Coverage Analysis

ABSTRACT

While the number of farm-level sustainability assessment tools is growing rapidly, concerns are raised on whether the assessment results of different tools present similar and valid conclusions about the sustainability performance of farms. In this paper we analysed the thematic scope of sustainability assessment tools, and compared assessment results from sustainability assessment tools. A coverage analysis of four tools (RISE, SAFA, PG and IDEA) demonstrated the diversity in approaches to assess sustainability at farm level. Tool developers select different (sub)themes and indicators, and apply different methods for measurement and aggregation of scores. This variability in approaches results not only in different tools, but can also result in different conclusions on the sustainability performance of farms. Decisions made in the development of a sustainability assessment tool need to be transparent to understand and explain the results of a tool and support farmers in a sustainable development of their farm. To improve the transparency in sustainability assessment tools we presented a framework describing decisions made in the development of a tool. An increased transparency in sustainability assessment tools can reduce the risk on distorted assessment results and actions, and contribute to the trust and relevance of future sustainability assessments.

1. Introduction

A growing number of sustainability assessment tools have been created to support farmers and policy makers in developing agriculture in a sustainable way (Binder et al., 2010). Sustainability assessments are increasingly seen as an important tool toward more sustainable production. Next to the sustainability assessment of the farm's performance, assessments tools can be used to discuss and learn, and to identify obstacles in the development toward farm sustainability (Marchand et al., 2014). Indicator-based sustainability assessment tools are generally structured using three or four hierarchical levels (Gasso et al., 2015). On the highest, most abstract level the dimensions of sustainability are presented, followed by more specific themes and subthemes, and finally, indicators. Indicators provide information on the status of a (sub)theme.

In developing sustainability assessment tools, decisions concerning what is relevant to assess, and how to assess, are based on value judgements of those involved (Gasparatos, 2010; Gasso et al., 2015; Lélé and Norgaard, 1996). Sustainability assessment tools and the indicators included (thematic scope) vary widely due to differences in value judgements, prioritizations in the selection of indicators (i.e. using selection criteria such as budget, time and data availability), spatial and temporal scales, system boundary, and target groups (Binder et al., 2010; Bockstaller et al., 2009; Schader et al., 2014). This diversity in tools results in concerns on whether the results of different tools present similar and valid conclusions about the sustainability performance of farms (Bockstaller et al., 2009). So far, comparative studies have predominantly focused on the validity of individual environmental indicators included in tools (Bockstaller et al., 2009; Galan et al., 2007; Thomassen and De Boer, 2005). Little attention has been paid to comparing the assessment results and conclusions derived from sustainability assessment tools or to similarities and differences in the thematic scope of sustainability assessment tools. If sustainability assessment tools with a similar purpose and scope (i.e. dimensions, geographical, sectoral) provide different conclusions on the sustainability performance of farms, the validity of the tools could be questioned (Bockstaller et al., 2009). This can result in a lack of trust and action to improve the sustainability at farm level (De Olde et al., 2016a). The objective of this paper is to analyse the thematic scope of sustainability assessment tools at farm level, and to compare assessment results and conclusions derived from such sustainability assessment tools. To this end, we selected sustainability assessment tools for comparison and analysed their characteristics and structure. A coverage analysis was used to compare the thematic scope of the tools and the coverage on dimension level. Next, a comparison of assessment procedures and sustainability assessment results was made. The methods used to study the different aspects of sustainability assessment tools are introduced at the start of each section, as

* Corresponding author at: Animal Production Systems Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.
E-mail address: eveliendeolde@gmail.com (E.M. de Olde).

http://dx.doi.org/10.1016/j.ecolecon.2017.02.015
0921-8009/

Table 1
General characteristics of the used tools adjusted from De Olde et al. (2016c).

Tool	Full name	Publication	Origin	Version	Year
RISE	Response Inducing Sustainability Evaluation	Häni et al. (2003)	Switzerland (Bern University of Applied Sciences	2.2	2011
SAFA	Sustainability Assessment of Food and Agriculture Systems	FAO (2013a)	Multiple countries and institutes	3.0	2013
PG	Public Goods tool	Gerrard et al. (2012)	United Kingdom (Organic Research Centre)	1.0	2011
IDEA	Indicateurs de Durabilité des Exploitations Agricoles	Zahm et al. (2008)	France (multiple institutes)	3.0	2008

they are a follow-up of the results of previous section. Based on the experiences and results gathered by using the different sustainability assessment tools, a framework is presented that demonstrates the impact of decisions made in the development of a tool on the final results and the importance of transparency of sustainability assessment tools in the future.

2. Selection and Structure of Sustainability Assessment Tools

To select tools to analyse the diversity in indicator-based sustainability assessment tools the following criteria were used: the assessment tool is focused at farm level, published by the tool developers in a peer-reviewed scientific journal or report, covers economic, environmental and social indicators, is suitable for livestock and arable agriculture in North-West Europe, is applied in multiple countries to enable contextualization, and the tool should be available in English and/or Danish to allow application on Danish farms (see for details De Olde et al., 2016c). Based on these criteria four tools remained from the initial list of 48 farm level sustainability assessment tools recently published in De Olde et al. (2016c). These four tools were: RISE (Häni et al., 2003), SAFA (FAO, 2013a), Public Goods (PG) (Gerrard et al., 2012), and IDEA (Zahm et al., 2008) (Table 1).

The following documentation was used to analyse the terminology, structure and selection of indicators, subthemes and themes in the tools: RISE (Grenz et al., 2012; Schoch, 2014), SAFA (FAO, 2013a,b), PG (Gerrard et al., 2012) and IDEA (Vilain, 2008; Zahm et al., 2008). When comparing the differences in terminology applied in the four tools, the need for aligning terminology in sustainability assessment tools becomes apparent (Table 2). The RISE manual mentioned the terms indicators and parameters for the theme and subtheme level. In an update of the tool in 2014, new terminology, given in Table 2, was introduced to align the tool more with terminology of SAFA (Schoch, 2014). What RISE defines as an indicator, we considered a subtheme, since it includes the assessment of various indicators. RISE and PG do not distinguish dimensions in their tool. In this study, the generally accepted terminology of SAFA was adopted.

Table 3 shows the number of dimensions, themes, subthemes and indicators in each tool. These elements present the structure of the tools. Overall, a high number of indicators are included in the tools. As a subtheme can be addressed by multiple indicators, a high number of indicators does not automatically mean that a tool has a higher attention or coverage of sustainability (sub)themes. For SAFA the number of indicators was derived from the tool manual. SAFA adds a fourth dimension of governance to the existing environmental, economic and social dimension of sustainability (FAO, 2013a). The RISE manual does not describe the number of indicators used in the aggregations to the subtheme level. RISE includes a complex aggregation of data and indicators. We studied the aggregations and identified 156 individual indicators to which a score is allocated. The number of indicators in PG is based on the number of questions, as for each question a score is defined. Finally, for IDEA the number of indicators is based on what IDEA identifies as criteria and covers all aspects to which a score is given.

3. Thematic Scope of Sustainability Assessment Tools

We used a coverage analysis to identify similarities and differences between the thematic scope of selected tools. A coverage analysis has been used in recent papers to evaluate the coverage of sustainability themes in sustainability assessment tools. Gasso et al. (2015), for example, analysed the effectiveness of themes and subthemes from existing sustainability assessment tools to cover sustainability issues in a specific case study. Whereas Schader et al. (2014) used a coverage analysis to evaluate the thematic coverage of six sustainability assessment tools, including RISE, as one of the indicators for the precision of tools. Schader et al. (2014) used one tool, the test version 1 of SAFA, as a reference for the coverage analysis. The paper, however, lacks transparency regarding the cut-off point in the coverage analysis. The tools selected by Schader et al. (2014) have been developed for different purposes (e.g. research, farm advice, policy advice). In our analysis, the selected four tools have similar characteristics (purpose, level of assessment and scope (i.e. covering environmental, economic and social dimensions) and compared all subthemes included in the four tools with each other. Similar to Schader et al. (2014) the coverage analysis focused on the subtheme level. This approach also enables a comparison of assessment results, since all tools include an aggregation of indicator scores to the subtheme level.

The coverage analysis evaluated whether the content of the subthemes of one tool is also addressed in subthemes of the other tools. First, the content of a subtheme was determined based on the description of the subtheme in the tool manuals and the indicators included. For example, the subtheme 'soil management' in RISE includes two aspects: the loss of agricultural area in the past 10 years, and knowledge collected about soil fertility (through soil analysis, humus balance and nutrient balances). Second, the subthemes of SAFA, PG and IDEA are studied to see if these aspects are also covered in one or more subthemes in these tools. The judgement whether something is covered is qualitative. After having explored coverages of several subthemes, we decided to have three levels of coverages. A high level of coverage (++) indicated that equal to and more than 75% of the content of a subtheme was found in one or more subthemes in the other tool. A coverage below 75% of the content was considered as intermediate coverage (+). No coverage (0) indicated that 0% of the content of a subtheme was found in any of the other tools (0). For example, when both aspects of the RISE theme soil management are covered in subthemes of SAFA, the coverage is considered to be high (++). When only one aspect is covered it is considered to be an intermediate coverage (+). As a result,

Table 2
Differences in terminology in tools.

RISE	SAFA	PG	IDEA
–	Dimension	–	Scale
Topic	Theme	Spur	Component
Indicator	Subtheme	Activity	Indicator
Indicator	Indicator	Question	Criteria

Table 3
Number of elements in each tool.

	RISE	SAFA	PG	IDEA
Dimension	-	4	-	3
Theme	10	21	11	10
Subtheme	50	58	57	42
Indicator	156	116	185	126