Interdisciplinary measurement: A systematic review of the case of sustainability

Brink, M., Hengeveld, G. M., & Tobi, H.

Title: Interdisciplinary measurement: a systematic review of the case of sustainability

Matthijs Brink[1], Geerten M. Hengeveld[1,2], Hilde Tobi[1]

[1] Biometris, Wageningen University & Research
PO Box 16
6700 AA Wageningen
The Netherlands

[2] Forest and Nature Conservation Policy Group, Wageningen University & Research
P.O. Box 47
6700 AA, Wageningen,
The Netherlands
e-mail addresses:
geerten.hengeveld@wur.nl;
hilde.tobi@wur.nl

Corresponding author: Hilde Tobi

Abstract

Measurement in interdisciplinary research can be difficult due to incompatibility of measures used in the different disciplines. These difficulties arise when measurements from different disciplines need to be combined to quantify an attribute for an object of study. Sustainability is a fuzzy concept measured as an attribute of many different objects. Sustainability measurements are often a composition of environmental, economic and social sustainability measurements, which makes them interdisciplinary composite measurements. The present study aims to advance interdisciplinary sustainability research by a systematic review of the interdisciplinary measurement of the attribute sustainability using the Triple Bottom Line (TBL) framework and to investigate whether measurement issues differ between the environmental, economic and social dimension.

A systematic literature review was done based on a query, on Web of Science, that reflected the interdisciplinarity of sustainability science and the TBL framework. Abstracts were scanned for relevance and papers that passed the scan were downloaded and scrutinized for their measurement characteristics such as measurement level, aggregation of scores into composite measurements, and weighting methods. An iterative coding approach (top-down and bottom-up) was used until saturation was reached.

The search returned 467 papers, of which 77 contained sufficiently detailed information on the interdisciplinary measurement of sustainability, following the TBL framework. Not all papers in the final set contained sufficiently detailed information on the performed measurements, and if and how the distinct dimensions were combined into one interdisciplinary sustainability measurement. The 86 reported measurement models of sustainability were heterogeneous with respect to measurement levels used, the use of composite scores, weighting schemes and time synchronization between indicators. Two thirds of the measurement models presented composite scores of the attribute sustainability. The weights used in composite scores were in most cases determined using experts' opinions.

Albeit all included papers used the TBL as a framework, their measurement models do not form a coherent set. Fuzzy measurement practices occurred in all three dimensions of sustainability. On closer inspection, and wherever reports were sufficiently specific, interdisciplinary measurement models of sustainability turned out to be near-unique.

Keywords

Sustainability science; multidisciplinary; Triple Bottom Line; composite measurement; measurement models

1 Introduction

Sustainability science is a prime example of interdisciplinarity, because it addresses the interactions between nature and society (Clark & Dickson, 2003; Kates et al., 2001). Sustainability science is driven by real-world problems and the need for solutions, which transcend the division of traditional disciplines (Clark, 2007; Jerneck et al., 2010; Wiek et al., 2011). Bibliometric data confirmed that sustainability research is indeed more interdisciplinary than scientific research in general (Schoolman et al., 2012).

Multiple attempts to define sustainability resulted in the wide-spread vision of sustainability consisting of a natural, a social and an economic dimension (Kajikawa, 2008). The abundance of sustainability indices and assessment frameworks illustrates the multiplicity of meanings and interpretations of sustainability (for examples, see: Böhringer & Jochem, 2007; Gasparatos et al., 2008; Singh et al., 2012). This variety is caused by the semantic vagueness of the concept itself, and by the normativity in “what is to be sustained and for how long” (Parris & Kates, 2003, p. 8068).

Sustainability measurements are often composite measurements: they consist of a natural, a social, and, an economic dimension, which in turn consist of multiple indicators. By definition, a measurement model, needed to build the composite, is associated with multiple challenges (Nardo et al., 2005; Saisana & Tarantola, 2002). Given the interdisciplinarity of sustainability science, some challenges relate to differences between disciplines. The concepts studied in the social sciences as a soft system are less appropriate for measurement theory as defined in the natural sciences (Finkelstein, 2005). Besides the measurement differences between disciplines, collaborative natural-social science projects may also suffer from tensions and mutual mistrust because of differences in traditions or habits (Bergman, 2011; Fischer et al., 2011; Siedlok & Hibbert, 2014).

As the demand for research and solutions on sustainability becomes more pressing, so will problems associated with interdisciplinarity in such research (Stock & Burton, 2011). The present research is a systematic literature review aimed at answering the following questions on interdisciplinary measurement of sustainability:

(1) How has sustainability measurement been reported in interdisciplinary natural-social science research published in the period 2011- 2015?

(2) To what extent are difficulties in measurement of sustainability attributable to issues in interdisciplinary research, or to measurement issues in the natural and/or the social sciences?

2 Theoretical framework

Sustainability is an attribute ascribed to different objects, e.g. products, ecosystems, countries or policy interventions. Sustainability is fuzzy and intangible (Gasparatos et al., 2008), resulting in a large number of different tools and frameworks (Singh et al., 2012). There is no ‘gold standard’ in measuring sustainability, yet the tripartite model or Triple Bottom Line (TBL) is widely used. The TBL distinguishes an environmental, an economic and a social dimension and is sometimes extended with additional dimensions (e.g., institutional, Singh et al., 2012). The TBL reflects the maturation of sustainability science from an environmental field of study into an interdisciplinary field including the nature-society interaction (Kates et al., 2001; Martens, 2006; Ostrom et al., 2007).

Following Kates et al. (2001) and Ostrom et al. (2007), we focus on interdisciplinary measurement of sustainability. Interdisciplinary research is (Aboelela et al. 2007, p. 341):''any study or group of studies undertaken by scholars from two or more distinct scientific disciplines. The research is based upon a conceptual model that links or integrates theoretical frameworks from those disciplines, uses study design and methodology that is not limited to any one field, and requires the use of perspectives and skills of the involved disciplines throughout multiple phases of the research process''.

Whereas the maturation towards interdisciplinarity is considered a conceptual improvement (Ostrom et al., 2007), it may complicate empirical research (Kajikawa 2008). A first challenge is the operationalization of the attribute sustainability into something measureable (Tobi & Kampen, 2018). Different disciplines use different measurement theories and logic (Finkelstein, 2003). Disciplines differ in the concepts they study and, thus, measure. Generally speaking, concepts studied in the social sciences are considered less tangible than concepts in the natural sciences. In the first, the measure of a concept is usually dependent on the theory the researchers favour, and different theories mean different measures and measurements. Generally, formal measurement theory, as it exists in the natural sciences, is absent in the social sciences (Kampen & Tobi, 2011). According to Finkelstein (2005, p. 270), concepts from the social sciences are generally less 'countable' and measurement levels are usually lower. Consequently, the philosophical foundations and requirements of measurement in the social sciences are less strict than the norms applied by natural scientists (Maul et al., 2016). Therefore, methodological studies on interdisciplinary measurement and studies that make use of interdisciplinary measurement require a conceptual framework that incorporates these differences.

Portfolio representation of measurements may be used to understand and develop interdisciplinary composite measurement (Tobi, 2014). Portfolio representation explicitly and formally maps the way attributes are measured as (Sawyer et al., 2013):

$$\tau = \sum \tau(i) w(i) \tag{1}$$

Here τ, the measure of the interdisciplinary attribute is expressed as a weighted combination of mono-disciplinary characteristics $\tau(1), \ldots, \tau(N)$, with weights $w(1), \ldots, w(N)$, where the measurement of the interdisciplinary attribute requires four steps. These steps are: the operationalization of the attribute into components that together form the consitituting parts of the portfolio; the definition of relations between the components that form the measurement model; the design of data collection to ensure for example time, duration, observer and measurement equivalence, and after the actual data collection; data processing according to the measurement model.

In the present study, the term *operationalisation* is used for the process, or any step in the process, that "[...] iteratively decomposes the immeasurable attribute of interest [...] into components until measurable characteristics are reached in an explicit and auditable way" (Tobi, 2014, p. 229). In this first step in the measurement process the attribute is operationalized into components: dimensions, indices and indicators (see Figure 1) in a way that reflects the researchers' theoretical framework or theoretical model.

In the second step of the measurement process, researchers build a measurement model using the indicators, indices and dimensions that resulted from the first step. When the dimensions relate to diferent scientific disciplines, this is called an interdisciplinary measurement model. This measurement model defines how the components are assumed to relate to the attribute of interest. It serves as estimator of the attribute and it includes the weights assigned to components and the way scores are, or are not, aggregated into composite scores. The measurement model is complete when the relationships among all indicators, indices and dimension of the attribute are unambigously defined. How indicators can be put together in one measurement model depends on their measurement level (i.e. nominal, ordinal, interval or ratio measurement). The combination of indicators measured with different measurement levels poses problems, e.g., the nominal assessment of the type of pesticide used (substance A or substance B) is hard to combine with the ordinal assessment of habitat quality (low, medium, high) into a single measurement of sustainability.

The third step in the process is the design of data collection and this is influenced by time, budget and other practical constraints. For some indicators, researchers will have to choose between different available instruments. The data collection can be done by the researchers, by other observers, or even without human observation. Aspects of data collection design such as differences in observers, duration and time of observation and instruments may lead to unequivalent measurements. Note that equivalence does not equate to the absence of measurement error; persons and machines may introduce error by unintentionally making mistakes or insufficient calibration.

Data processing makes up the fourth step in the process. Data processing includes auditing the work agreed on in the previous step, digitalizing or transferring data from one format into another and the (mathematically) processing of the data to obtaine an attribute measure in accordance with the measurement model from the second step.

In the context of sustainability, Figure 1 depicts how the attribute sustainability may be operationalized following the TBL framework. From left to right, three dimensions are distinguished in the operationalisation of the attribute sustainability. Each of these dimensions calls for indices, e.g. the environmental dimension may include air quality, the economic dimension income and the social dimension well-being. These indices are operationalized into indicators. Because indicators for concepts in the social dimension (e.g. indicators for well-being) may be less measurable than indicators for the environmental dimension (e.g. O_3 as an indicator for air quality), the first sub-question for the present study is

(i) *To what extent are the measurement levels of indicators different between the environmental, economic and social sustainability dimensions?*

Composite measurements of sustainability, may call upon indicators with different measurement level (e.g. working hours as ratio level indicator, and satisfaction with job as ordinal level indicator in Figure 1). Indicators with different measurement levels cannot be processed and analysed in the same way (Stevens, 1946). Different procedures may be used to reach commensurable measurement scales, but the procedure chosen has implications for the composite measurement (Gasparatos et al., 2008,

p. 303). For example, the normalization of indicators involves the choice of a reference value, which introduces arbitrariness into the measurement (Böhringer & Jochem, 2007, p. 6). This arbitrariness can be avoided by analysing indicators separately, with the backdrop that no single composite is obtained that would have allowed simple comparisons between objects. The second sub question is therefore:

(ii) *If so, how are indicators with different measurement levels combined into composite measurements of sustainability?*

To come to a *measurement model*, the relations between indicators and the attribute must be defined, from right to left in Figure 1. One way is to express each index as the sum of equally weighted subsequent indicators, (e.g. income as the sum of different income indicators), and to repeat this for indices within a dimension, and for dimensions within the attribute. Another option is a measurement model that defines sustainability as a specific collection of nominal level sustainability criteria that have to be met simultaneously.

Most ways of combining dimensions of sustainability imply that the different dimensions may compensate for one another, defined as weak sustainability (Neumayer, 2003). The weights assigned determine how strong components contribute to the sustainability of the object under study. This stance implies that unsustainable use environmental resources may be compensated by economic progress (Davidson, 2010). A strong sustainability measurement model emphasizes the collective of the sustainability indicators,through a set of sustainability criteria that have to be met simultaneously. The choice between weak and strong sustainability calls for transparency. Afterall, whether indicators can or cannot compensate for one another has implications for the meaning of the sustainability measurement (Gasparatos et al., 2008).The way sustainability dimensions are weighted in composite measurements, and the distribution of weights is addressed in the question:

(iii) *How are indicators and dimensions from different disciplines weighted within composite measurements of sustainability?*

In practice, the selection of indicators is often constrained by data availability (Sala et al., 2015), or the ease of data collection and use (Liverman et al., 1988). Indicators may be obtained from old questionnaires, because conducting a new survey is too expensive or cumbersome. These practical delibarations may result in measurements at different points in time. Sustainability measurements are often performed to assess changes through time. The fuzziness of the measure increases with varying time conditions, e.g. when the measure in Figure 2 is made up out of environmental sustainability in year *t,* and economic and social sustainability in year *t*-1. The fourth and last sub question in this review is therefor:

(iv) *To what extent are interdisciplinary measurements of sustainability based on synchronized indicator measurements?*

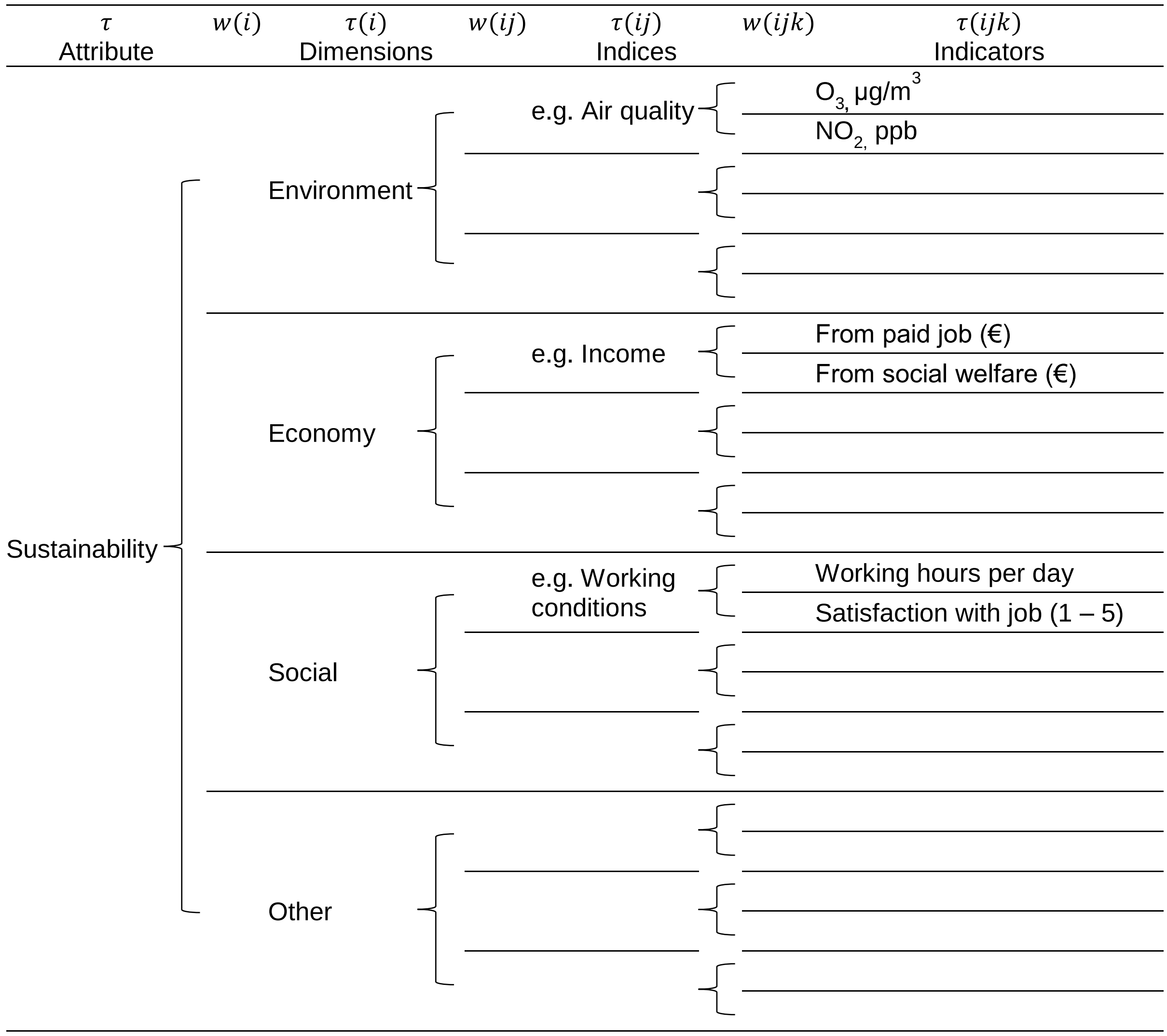

Figure 1. Portfolio measurement terminology illustrated with the attribute sustainability. The fuzzy attribute sustainability is decomposed into three or more dimensions. These dimensions are decomposed into multiple indices that are measured through one or multiple measurable indicators. To define a measure of the attribute sustainability, a measurement model is defined that creates a composite measure of the indices from the indicators, a composite measure of the dimensions from the underlying indices and a composite measure of the attribute from the dimensions (Tobi, 2014).

3 Methods

The selection of articles to be included in the systematic review is done in four steps (figure 2). First, all articles within the scope of the research questions are selected based on a query in Web Of Science. Second, these articles are screened for eligibility based on their abstract. Third, only articles from which the full text could be retrieved are included. Fourth, full article review was performed for eligibility to the search criteria.

To systematically review the measurement of sustainability in interdisciplinary research, the keywords in review question 1 were identified and operationalized in a search query (Table 1). The search was performed in the four core citation databases of ISI Web Of Science: Science Citation Index, Social Science Citation Index, Arts & Humanities Citation Index, and Emerging Sources Citation index. These were chosen because they reflect a wide collection of relevant disciplines and multidisciplinary and international research.

The search described in Table 1 performed on 29 March 2016 yielded N = 467. These references and their abstracts were subsequently screened for eligibility (Figure 2). For this screening inclusion and exclusion criteria were formulated related to the variables presented in Table 1. A sample (n = 20) of the abstracts was screened independently by two researchers to test the criteria. It was found that the raters often did not agree on the degree of integration or collaboration or interdisciplinarity expressed in the abstracts. After discussion, this criterion was dropped from the screening. Full screening was then carried out based on two criteria: sustainability and measurement. A paper was deemed eligible for further analysis if its abstract indicated that the paper dealt with sustainability measurements that comply with the TBL or an extension thereof, and that it reports on measurements of sustainability as an attribute of some object. The inter-rater agreement on these criteria for a new sample of abstracts (n = 20) was considered good enough (Kappa = 0.70). Automatic coding of relevant words in the abstracts was used in Atlas.ti to assist the screening of the abstracts and to increase completeness and speed of processing. After screening all abstracts, 187 articles remained for which the full-text was sought. As 25 articles (13%) could not be retrieved, a set of 162 full-text articles (Appendix A) was further scrutinized.

Table 1. Operationalization of review question into variables to identify papers

Words in RQ	**Conceptual framework applies to research ...**	**Operationalisation query**
Sustainability	that is on sustainability;	sustainab*
Measured	in which sustainability is measured or assessed as an attribute of some object;	sustainab* NEAR/10 (evaluat* OR assess* OR measur*)
Interdisciplinary	that mentions integration of disciplines, or collaboration between disciplines;	(*disciplin* OR integrat* OR collabor* OR synthesi* OR (field OR domain) NEAR/2 (academ* OR scholar* OR research))
Natural-social science research	that is about the natural science (environmental) and the social science side and the economic side;	(*envir* OR *ecol* OR *bio*) AND (*econom* OR financ*) AND (soci* OR communit* OR communal*)
Published in the period 2011-2015	published in the period 2011-2015.	*Timespan=2011-2015*

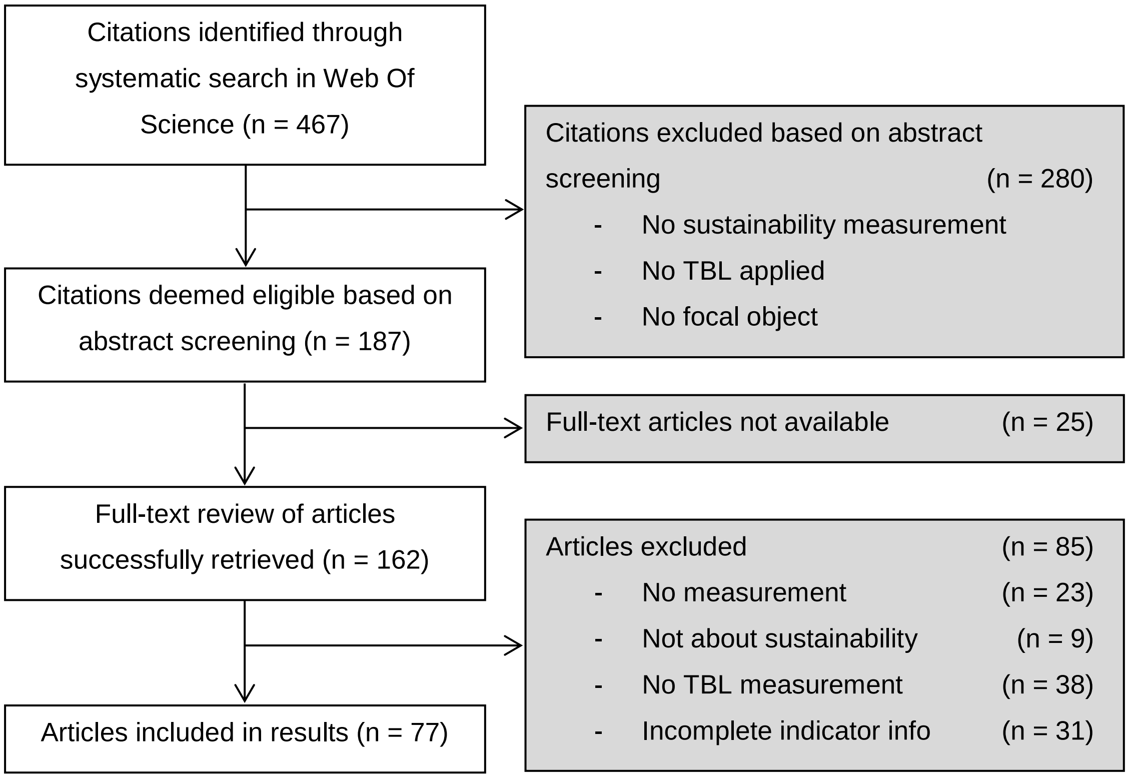

Figure 2*.* Flow chart of eligibility review process

Of the 162 retrieved full papers, 85 were excluded for one or more reasons as depicted in Figure 2. This resulted in a final set of 77 papers. Some key characteristics of the 162 retrieved papers and the papers in the final set, are listed in Table 2. The total number of different scientific journals was 85 and 43 respectively. When an article presented separate measurement models, models were considered independently. Hence the results include 86 measurement models obtained from 77 papers (Table 2).

Study objects in the final set ranged from the housing market in European countries to beef production in the US, from automotive components to structural concrete columns, and land use policies in Mongolia to solid waste management systems in Thailand.

Data was collected through coding of reports on the measurement models using Atlas.ti. For variables on measurement levels, TBL dimensions, and composite reporting, coding was top-down according to a predefined coding scheme (coding scheme available on request). Other variables like weighting schemes and synchronized measurement were coded iteratively using an open coding approach until saturation was reached. The resulting scheme was then applied top-down to ensure coding consistency over all papers and models. Lastly, component names and study object names were coded by extracting text fragments from the articles under study. Data was extracted on different levels: the measurement model (attribute) level, dimension level, lower intermediate levels and, finally, the indicator level.

Table 2. Description of included papers

	Final set (N=77)	Full-text review (N=162)
Publication year		
2011	12	24
2012	13	26
2013	13	31
2014	15	34
2015	24	47
Journal with at least 3 articles in final sets		
Ecological Indicators	10	14
Journal of Cleaner Production	7	13
Sustainability	4	6
Agricultural Systems	3	3
Ecological Economics	3	4
International Journal of Life Cycle Assessment	3	4
Most commonly mentioned keywords		
environmental sciences	38	77
indicators	23	30
sustainability	22	56
management	20	44
engineering, environmental	19	36
Number of measurement models in paper		
1	71	
2	5	
5	1	
Total number of models	86	

4 Results

The 86 sustainability models under review reported a total of 1.817 indicators, of which 670 environmental, 475 economic, 499 social and 173 belonging to other (e.g. institutional) or multiple TBL dimensions (e.g. socio-economic). Regardless of the TBL dimensions, most of the indicators (1.163, 64%) were reported as interval or ratio level indicators. Table 3 depicts measurement levels of the indicators for the different TBL dimensions.

Overall, there was an association between dimension and the indicator's level of measurement (chi-square with 6 dF, p<.001). Economic indicators were reported as interval or ratio level indicators more often (71%) than environmental (66%) and social indicators (55%). Social sustainability indicators were more often nominal/ordinal than indicators for environmental and economic sustainability (38% vs. 23% and 26% respectively). Indicators with indistinct reports on measurement levels were most prevalent in the group of other or multiple dimensions (16%).

The above confirmed to some extent the expectation that measurement levels of indicators would differ between the three TBL dimensions. Lower measurement levels are less prevalent among environmental and economic sustainability indicators than among social sustainability indicators, yet they are no exception in any dimension.

Table 3. Indicator measurement levels by sustainability dimension (n=1.817, within 86 models)

		Dimension				
		Environmental	Economic	Social	Other/multiple	Total
Nominal/ordinal	N of indicators	153	125	191	35	504
	% of indicators	23	26	38	20	28
Interval/ratio	N of indicators	444	336	272	111	1163
	% of indicators	66	71	55	64	64
Unknown	N of indicators	73	14	36	27	150
	% of indicators	11	3	7	16	8
Total	N of indicators	670	475	499	173	1817
	% of indicators	100	100	100	100	100

In 27 measurement models no composite measurement at the attribute level was presented (Table 4). Most of these models reported seperate indicators. Seven models did report dimension composites of which 3 were the result of a measurement model with interval/ratio measurement levels only (Table 4). For 59 out of the 86 measurement models (69%) composite measurements at the attribute level were presented, most frequently based on a mixture of (unknown) measurement levels (n=26) or on the nominal/ordinal level (n=20). The attribute level measurement was often complemented with measurements on dimension composites and seperate indicators (with n=37 each).

Table 3 also shows that 41 models contained a mixture of, occasionally unknown, measurement levels complicating the presentation of a attribute level composite. Most of these models (26/41) did however not present an attribute level composite.

Overall, the report of composites at the attribute level did not replace the report of seperate indicators; more than half of the models that showed an attribute level composite also gave seperate indicators (37/59). Nearly all (24/27) of the models that showed no composite attribute measurement did present seperate indicators, often of mixed and/or unknown measurement levels.

When looking at the number of indicators from each dimension on which the 59 composites were based, the average number of ecological, economic and social indicators was 7.8, 5.8 and 6.1 respectively, a non-significant difference (repeated measures ANOVA, Huyn-Feldt corrected F(1.450, 81.707)= 2.427, p= 0.110).

Table 4. Reports of composite scores by measurement level patterns (N= 86 models).

	All interval/ratio	All nominal/ordinal	Mixed and/or unknown	Total
No attribute level composite	9	3	15	27
Dimension composites	3	3	1	7
Lower level composites	3	0	2	5
Separate indicators	8	2	14	24
Attribute level composite	13	20	26	59
Dimension composites	8	14	15	37
Lower level composites	4	1	8	13
Separate indicators	8	15	14	37
Total	22	23	41	86

The weighting methods reported reflected very clearly the great variety in composite sustainability measurements (Table 5). The weighting methods reported varied from simple schemes involving equal weights, to multivariate schemes taking covariances into account, or expert opinions. Four models explicitly investigated sensitivity to uncertainty of weights.

Most frequently mentioned were schemes based on expert opinions. This implies that decisions on how composite scores were calculated were not made by the authors, but by people they deemed experts. This class was very diverse as it included different types of multiple criteria decision models, analythic hierarchy process, analytic network process, Delphi, and more.

Table 5. Reported weighting methods within composite measurements of sustainability (N=86 models)

	Model composite	Dimension composites	Lower level composites
Based on expert opinions	35	28	9
Equal weights approach	12	7	4
Multivariate approach	5	3	
Uncertainty addressed	4		
Other	3	6	5
Total	59	44	18

For about one in every three measurement models it was impossible to assess whether indicator measurements were time synchronized, regardless of whether or not an attribute composite was given (Table 6). Synchronized indicators were reported for about one in every three measurement models overall, somewhat more frequently in attribute composite measurement models and somewhat less frequently in composite measurement models at a lower level. This suggests that there is a broad awareness of the relevance of synchronized indicators to the the composite measurements provided. However, there is no consensus yet on the need to report the use of (non) synchronized indicator measurements.

Table 6. Reports on time of measurement within sustainability measurement models.

	No attribute composite	Attribute composite	Total
Match reported	7	23	30
Differences reported	11	17	28
No (clear) reports	9	19	28
Total	27	59	86

Discussion

The present study was, to our knowledge, the first to systematically study reports on the interdisciplinary measurement of sustainability relying on the TBL framework. There was substantial heterogeneity in the reported measurement models of sustainability; we observed the presence or absence of a composite; a range in number of steps between lowest level indicators to highest level composite; near-unique weighting methods combining all measurement levels or not, and the presence or absence of synchronized indicator measurements. This heterogeneity occurred both within and between dimensions, therefor we could not attribute difficulties in measurement to the natural or social sciences specifically, or to problems with interdisciplinarity.

All measurement levels were found in all dimensions, but indicators for the social dimension were relatively more frequently of the lower measurement levels (nominal or ordinal) than indicators of the environmental and economic dimension. In the papers analysed, the average number of environmental indicators per model was slightly larger than the average number of economic and social indicators, although not statistically significant. Whether the number of indicators was associated with the width and complexity of that dimension, the interest of the research team or the measurement levels or data availability was out of the scope of this research.

One in three models did not yield an attribute level composite at all, possibly because of problems in the construction of a composite caused by lower measurement levels (Singh et al., 2012). Composite sustainability measurement models require quantification (also when relying on lower measurement levels), normalization and aggregation. The exact weighting methods reported were near-unique and most frequently based on expert opinions. This suggests that researchers may have been reluctant to impose what they, as experts themselves, think of as sustainability. Instead, what is considered sustainability is mediated through their opinion of who is, and who is not, an expert. This drawing on different experts probably contributed to a lack of standardized interdisciplinary measurement of sustainability.

This systematic review was limited to the query as depicted in Table 1. The query and selection required a clear theoretical framework and produced reliable results. The query was informed by and limited to interdisciplinary collaboration within the TBL framework presented in papers as such. The query called for explicit referrals to interdisciplinarity and natural-social science research which means that papers written by authors who saw no need to refer to interdisciplinarity or multi-disciplinary collaboration were not included. This may have favoured papers reflecting research within which interdisciplinairty and multidisicplinarity was an issue. Determining the level to which this affects the research results is beyond the scope of this paper. The substantial reduction from 467 citations in the initial set to 77 in the final set, was the result of a systematic procedure that ensured the final set to consist of papers which reported actual measurement.

The coding that yielded the data for the analysis, was based on measurement theory but relied on the completeness and clarity of the papers. Because not all papers were clear, nominal and ordinal measurement levels were taken together, just like interval and ratio measurement levels. The presented categories are the result of the reported measurement in the final set of papers, and coding and classification decisions made by the authors. Generating uniform and unambiguous ways to describe measurement practices meant loss of detail of the individual models.

This study showed that ‘fuzzy’ measurement is not the sole domain of social sciences, but is common in all dimensions in sustainability measurement. Even within one framework, measurement models were near-unique. There were differences in approaches between studies and heterogeneity in measurement levels in both social sciences and natural sciences input into the measurement models. As a result no clear conclusions could be drawn on whether difficulties in measurement of the overarching attribute of sustainability is due to the interaction between the sciences, or has roots in either one of the sciences. Thus, we could not answer our second research question. This may explain why we found little attention in the papers for measurement validation, that is the assessment of the extent to which the measure reflects the attribute of interest (Borsboom et al., 2004). In measurement validation the match between the indicators and the interdisciplinary theoretical framework is explicitly judged (Tobi & Kampen, 2018).

There is a need for sustainability measurement models that are scalable and comparable beyond specific contexts (Sala et al., 2015). Without proper reporting on all characteristics in measurement models, the results cannot be integrated into general theory; after all, different measurement models may lead to different conclusions, policy advice, evaluations of policy effects et cetera. The papers reviewed in this study all addressed sustainability through interdisciplinary measurement. The measurement models used to bridge this interdisciplinarity did not converge. Differences in measurement level of indicators and the weighing of indicators to come to composite measures of indices, dimensions or the attribute sustainability resulted in near unique measurement models between these papers. For the weighing of indicators to form a composite measure, most studies relied on expert judgement. The heterogeneity of the measurement models studied hampered answering the research question on whether the difficulties in measurement were attributable to issues in interdisciplinary research. Overall, a common understanding and operationalization of the concept sustainability is needed to make the measurement of sustainability a measure, rather than an opinion, and to allow for comparison across time and contexts, and to evaluate policy decisions.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., . . . Gebbie, K. M., 2007. Defining interdisciplinary research: Conclusions from a critical review of the literature. Health Services Research, 42, 329-346. doi:10.1111/j.1475-6773.2006.00621.x

Bergman, M. M., 2011. The politics, fashions, and conventions of research methods. Journal of mixed methods research, 5, 99-102.

Böhringer, C., & Jochem, P. E. P., 2007. Measuring the immeasurable — A survey of sustainability indices. Ecological Economics, 63, 1-8. doi:http://dx.doi.org/10.1016/j.ecolecon.2007.03.008

Borsboom, D., Mellenbergh, G.J. & van Heerden, J., 2004. The Concept of Validity. Psychological Review, 111, 1061-1071. doi:10.1037/0033-295x.111.4.1061

Clark, W. C., 2007. Sustainability Science: A room of its own. Proceedings of the National Academy of Sciences, 104, 1737-1738. doi:10.1073/pnas.0611291104

Clark, W. C., & Dickson, N. M., 2003. Sustainability science: The emerging research program. Proceedings of the National Academy of Sciences, 100, 8059-8061. doi:10.1073/pnas.1231333100

Davidson, K. M., 2010. Reporting Systems for Sustainability: What Are They Measuring? Social Indicators Research, 100, 351-365. doi:10.1007/s11205-010-9634-3

Finkelstein, L., 2003. Widely, strongly and weakly defined measurement. Measurement, 34, 39-48. doi:10.1016/s0263-2241(03)00018-6

Finkelstein, L., 2005. Problems of measurement in soft systems. Measurement, 38, 267-274. doi:10.1016/j.measurement.2005.09.002

Fischer, A. R. H., Tobi, H., & Ronteltap, A., 2011. When Natural met Social: A Review of Collaboration between the Natural and Social Sciences. Interdisciplinary Science Reviews, 36, 341-358. doi:10.1179/030801811x13160755918688

Gasparatos, A., El-Haram, M., & Horner, M., 2008. A critical review of reductionist approaches for assessing the progress towards sustainability. Environmental Impact Assessment Review, 28, 286-311. doi:http://dx.doi.org/10.1016/j.eiar.2007.09.002

Jerneck, A., Olsson, L., Ness, B., Anderberg, S., Baier, M., Clark, E., . . . Persson, J., 2010. Structuring sustainability science. Sustainability Science, 6, 69-82. doi:10.1007/s11625-010-0117-x

Kajikawa, Y., 2008. Research core and framework of sustainability science. Sustainability Science, 3, 215-239.

Kampen, J. K., & Tobi, H., 2011. Social Scientific metrology as the mediator between sociology and socionomy: a cri de coeur for the systemizing of social indicators. In Social Indicators: Statistics, Trends and Policy Development (pp. 1-26): Nova Science Publishers, Inc.

Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., . . . Svedin, U., 2001. Sustainability Science. Science, 292, 641-642. doi:10.1126/science.1059386

Liverman, D. M., Hanson, M. E., Brown, B. J., & Merideth, R. W., 1988. Global sustainability: Toward measurement. Environmental Management, 12, 133-143. doi:10.1007/bf01873382

Maul, A., Torres Irribarra, D., & Wilson, M., 2016. On the philosophical foundations of psychological measurement. Measurement, 79, 311-320. doi:http://dx.doi.org/10.1016/j.measurement.2015.11.001

Martens, P., 2006. Sustainability: science or fiction? Sustainability: Science, Practice and Policy, 2, 36-41, DOI: 10.1080/15487733.2006.11907976

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E., 2005. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Statistics Working Papers. Retrieved from http://dx.doi.org/10.1787/533411815016 doi:10.1787/533411815016

Neumayer, E., 2003. Weak versus strong sustainability: exploring the limits of two opposing paradigms: Edward Elgar Publishing, Cheltenham, UK.

Ostrom E, Janssen MA, Anderies JM, 2007. Going beyond panaceas. Proc Natl Acad Sci USA 104:15176–15178 https://doi.org/10.1073/pnas.0701886104

Parris, T. M., & Kates, R. W., 2003. Characterizing a sustainability transition: Goals, targets, trends, and driving forces. Proceedings of the National Academy of Sciences, 100, 8068-8073. doi:10.1073/pnas.1231336100

Saisana, M., & Tarantola, S., 2002. State-of-the-art report on current methodologies and practices for composite indicator development. Retrieved from https://publications.europa.eu.

Sala, S., Ciuffo, B., & Nijkamp, P., 2015. A systemic framework for sustainability assessment. Ecological Economics, 119, 314-325. doi:http://dx.doi.org/10.1016/j.ecolecon.2015.09.015

Sawyer, K., Sankey, H., & Lombardo, R., 2013. Measurability invariance, continuity and a portfolio representation. Measurement, 46, 89-96. doi:10.1016/j.measurement.2012.05.023

Schoolman, E. D., Guest, J. S., Bush, K. F., & Bell, A. R., 2012. How interdisciplinary is sustainability research? Analyzing the structure of an emerging scientific field. Sustainability Science, 7, 67-80. doi:10.1007/s11625-011-0139-z

Siedlok, F., & Hibbert, P., 2014. The Organization of Interdisciplinary Research: Modes, Drivers and Barriers. International Journal of Management Reviews, 16, 194-210. doi:10.1111/ijmr.12016

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K., 2012. An overview of sustainability assessment methodologies. Ecological Indicators, 15, 281-299. doi:http://dx.doi.org/10.1016/j.ecolind.2011.01.007

Stevens, S. S., 1946. On the Theory of Scales of Measurement. Science, 103, 677-680. doi: 10.1126/science.103.2684.677

Stock, P., & Burton, R. J. F. , 2011. Defining Terms for Integrated (Multi-Inter-Trans-Disciplinary) Sustainability Research. Sustainability, 3, 1090. doi:10.3390/su3081090

Tobi, H., 2014. Measurement in interdisciplinary research: The contributions of widely-defined measurement and portfolio representations. Measurement, 48, 228-231. doi:10.1016/j.measurement.2013.11.013

Tobi, H. & Kampen, J.K., 2018. Research Design: the Methodology for Interdisciplinary research Framework. Quality & Quantity, 52: 209-1225. doi: 10.1007/s11135-017-0513-8

Wiek, A., Withycombe, L., Redman, C., & Mills, S. B., 2011. Moving Forward on Competence in Sustainability Research and Problem Solving. Environment: Science and Policy for Sustainable Development, 53, 3-13. doi:10.1080/00139157.2011.554496