

From scoring to orienting: the development of a compass for evaluating sustainable urban development at the neighbourhood scale

H. Vandevyvere

*Department of Architecture, Urbanism and Planning,
Katholieke Universiteit Leuven, Belgium*

Abstract

Recent tools for the evaluation of sustainable neighbourhood performance have been conceived as measuring instruments that deliver an unequivocal outcome in the form of a score. The benefits of these tools in terms of benchmarking, communication power and marketing are obvious. However, the complex trade-offs that characterize urban development projects complicate a straightforward translation of ‘sustainability aspect’ to ‘score’, not at least because quantitative and qualitative criteria must be combined into a final judgement. The quantitative data uncertainties characteristic of many urban surveying processes further add to the problem. Quantifying the unquantifiable may therefore be considered as the Achilles’ heel of these evaluation tools. A more process-oriented guidance can then offer an alternative.

The present contribution describes an approach that combines scoring with arguing as a way out of the difficulties of mere quantification. Attention is deviated from the single score to the wider panorama of contributing factors. Three elements provide the basis for the development of the method. Referring to research in political sciences, indicators are primarily conceived as ‘boundary institutions’ that facilitate information exchange between scientists, decision makers and stakeholders. A new form of radar diagram is developed for a better visualisation of complex valuations, minimizing perception bias while maximizing useful information content. Performance thresholds and qualitative accounting are finally added as indispensable components of the judgement palette. The resulting tool thus functions as a compass rather than a gauge.

Keywords: sustainable urban development, indicator system, neighbourhood, assessment, multimodal system analysis, radar diagram.



1 Introduction

Evaluating the sustainable performance of urban fragments has gained much attention in recent years. The advent of market-based instruments like *LEED for Neighborhood Development* [1] and *BREEAM Communities* [2] in 2009 is illustrative of this trend. It coincides with the emerging consciousness that aggregated sustainability benefits appear at the urban scale level compared to optimizing single buildings for sustainability, or inversely, that ignoring sustainability factors at the urban scale results in deficient overall system performance. Evaluation does not come without methodological difficulties, however. At the level of the neighbourhood or the city, the complexity of framing and assessing sustainable performance is considerable. Two major contributing factors to this complexity are the intricate mix of environmental, social and economic border conditions for sustainability on the one hand, and the difficulties of quantitative modelling and corresponding data collection on the other hand. The first factor can further be interpreted in terms of a tension between quantitative and qualitative evaluation, or between the *determinative* conditions for environmental sustainability and the *normative* conditions that define the sustainability of society as a whole [3, 4].

Starting from the observation that a mixed quantitative-qualitative evaluation will thus meet difficulties in uncontested scoring, an alternative scheme for sustainability evaluation has been developed. Elements of such strategy can also be found with other acquainted assessment tools (e.g. [5]), but it is not the approach of commercial labels and official rating instruments (e.g. [6]), as these aim at an undisputable outcome for reasons of benchmarking and regulation. The alternative approach delivers a compass rather than a gauge, and has particularly been thought of as an instrument for *reflexive governance* [7].

2 Preconditions for the assessment instrument

Starting from both a literature review and field inquiries, a set of preconditions for the assessment instrument has been determined. These are, summarizing [4]:

To provide a comprehensive and transparent way of assessing sustainability so that professional actors can go through the evaluation process with a larger group of participating stakeholders [8, 9];

- To consider sustainability components through indicators that are effective at the micro-urban scale level. Efficiency can be gained by integrating input from other evaluations carried out at the same time, e.g.: EIAs (environmental impact assessments), EPR calculations (energy performance regulation), building simulations, compulsory water management checks, life cycle assessments (LCA) and life cycle costing (LCC);
- To cover, by extension, all *modal spheres* of sustainable functioning;
- To focus on results rather than on the means to achieve these results;
- To provide a strategy to deal with uncertainties;



- To be flexible enough to allow adaptation to the unique context of a particular project;
- While not remaining so undefined or elastic that individual actors can divert the evaluation process towards their sectorial concerns, thus harming the common interest or the overall sustainability targets (e.g. 10)).

The completeness in terms of *modal spheres* refers to a methodological concept based on the philosophical work of Herman Dooyeweerd [3, 4, 11–13].

By evaluating qualitative results, rather than the means to achieve these results, the method adopts a different approach compared to pure rating systems. For example, spatial quality will be assessed *in se* and will not be valued procedurally, e.g. by controlling if a spatial quality surveyor has been assigned during the development process.

The above preconditions have led to three strategic options for working out the instrument:

- The selected indicators shall function as *boundary institutions* that facilitate the exchange of knowledge between the spheres of science, policy and society [14];
- Rating will be complemented by other assessment techniques, in particular the introduction of performance thresholds and the composition of (qualitative) arguments;
- A new type of radar diagram will be developed, in particular to respond to problems of perceptible bias that occur with some commonly used radars.

Performance thresholds must guarantee that essential conditions for sustainable functioning are fulfilled, regardless of the obtained overall score.

3 Indicator system and composite tool

The present section explains the components of the evaluation.

3.1 Indicators

A set of indicators makes up the basis of the system. As mentioned higher, these indicators are thought of as boundary elements that facilitate knowledge transfer between concerned parties. Main elements can be composed of up to 4 sub-indicators. Two major inputs guide the selection of these indicators: a survey of existing indicator systems on the one hand, and multimodal control for completeness of the set on the other hand. Complementary input was obtained from an expert panel (see also further). The resulting system of main indicators with its inputs is represented in figure 1. The indicators are grouped in four categories that correspond to the classical 3P division (*planet, people, prosperity*), completed with a category that controls process quality and integrity. This may be compared with ISO 14031's distinction between *operational performance indicators* (3P categories) and *management performance indicators* (steering category) [15].



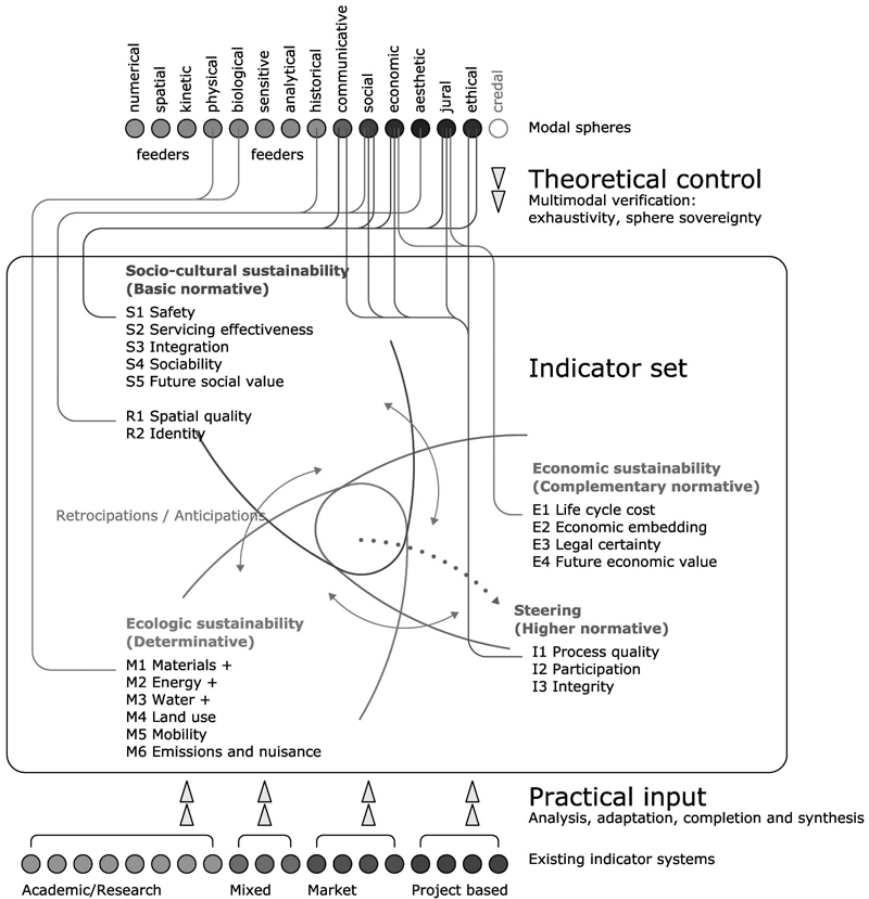


Figure 1: Inference diagram for the main indicator system.

With the aim of constructing a working basis, several types of existing indicator system have first been screened for valuable input. The analysed systems can be distinguished as belonging to 4 major spheres: academic and research-oriented; market-based; mixed; and *ad hoc* or project-based. Mixed instruments refer to tools that result from a collaboration between research institutes, consultancies and/or (local) authorities. In total, 19 indicator systems have been analysed this way [4], among which LEED for Neighborhood Development, BREEAM Communities, the Hammarby Sjöstad Model [16] and BedZED's 21 Steps Chart [17], to name a few of the better-known examples.

The resulting pragmatic set of possible indicators has subsequently been reworked and controlled for completeness. At the top of figure 1, the modal aspects to be accounted of are represented as knowledge spheres. Two types of knowledge input, direct and indirect, may be distinguished. For example,

mathematics, while belonging to the numerical realm, has a supportive role in delivering tools for analysing other spheres, e.g. the social realm by means of statistics or the economic realm through monetizing techniques. Such indirect aspects are labelled as feeders in figure 1. An in-depth discussion of multimodal analysis and its application in sustainability evaluation is beyond the scope of the present article, but can be found in [4].

An observation should be made about the independency of the indicators. It appears that a workable set is always composed of indicators that influence each other, and so are not strictly independent. For example, the sustainability of the transport system analysed under mobility influences spatial quality and physical quality of life (absence of harmful emissions and nuisance), and thus indirectly many other environmental and social aspects. A similar reflection can be made about economic behaviour viewed as a social construct, or about any other links between modal aspects. This phenomenon is recognized in multimodal system analysis, and explained in terms of the inevitable *anticipations* and *retroicipations* that exist between the modal spheres. In fact, the only independent variable for our analysis appears to be the integrated sustainability itself. However, when assigning effects to particular indicators, it remains important to avoid double-counting. For example, energy use for transport will here be considered under the indicator energy, and so not again under the indicator mobility.

Sub-indicators make up the technical layer for the assessment. As an example, the sub-indicators for energy consumption during the use phase and for spatial quality are explained here.

Energy use is assessed by considering:

- Fossil energy consumption in buildings and infrastructures during their exploitation, expressed as surface-averaged $\text{kWh}_{\text{primary, fossil}}/\text{m}^2, \text{ year}$ or as $\text{kWh}_{\text{primary, fossil}}/\text{person, year}$;
- Environmental quality of the building and infrastructure energy use expressed as proportionally averaged Ecopoints/kWh;
- Fossil energy consumption for transport of inhabitants and users, expressed as $\text{kWh}_{\text{primary, fossil}}/\text{person, year}$.

Energy embedded in building and infrastructure materials is included in the materials indicator through an LCA-score.

The energy assessment is thus not a pure LCA-analysis, but reflects a series of actual priorities as follows: (1) to reduce the fossil energy consumption of buildings and urban infrastructures; (2) subsequently, to fill in the remaining energy demand as environmentally friendly as possible; and (3) to add the location impact of settlements on transport energy in order to complete the energy consumption figure of the urban fragment. This approach is similar to the principles of the *trias energetica*. Moreover, the subdivision allows to handle certain problems of data collection and modelling: at the neighbourhood scale it is very difficult to assess mobility effects, whereas estimating intra-building energy figures is more feasible. The sub-indicators allow thus to give relief to the information in terms of uncertainty degrees. Of course, this relief is

lost at the aggregation level of the main indicators, but the processed information remains available for in-depth analysis.

In contrast with the quantitative assessment of energy use, spatial quality is judged in exclusively qualitative terms. The valuation is structured by a non-exhaustive series of aspects to be judged, as follows:

- gradations of public and private character, transitions between these spheres, buffer areas, and corresponding changes of scale;
- articulation of the different programmatic elements with respect to each other; in particular the relation between buildings and infrastructures and the possible barrier effects of the latter;
- meaningful integration of green structures and green-blue networks into the urban landscape;
- visual landscape quality and scenic beauty;
- legibility and permeability of the urban tissue;
- integration of different architectural concepts in a given setting or masterplan; and the articulation and integration of existing patrimony herein;
- authenticity and architectural quality of individual buildings.

The final score is normative, but could be numerically composed by assigning symbolic ratings to each of the individual aspects, and subsequently weighting them. In the present state of the instrument, such partial scoring and weighting has not been provided for spatial quality. Whether or not this is done depends on how far users want to instrumentalize the evaluation. This problem is very similar to the challenges of judgement formulation in, for example, architectural competitions.

Mixed sets of sub-indicators are used as well. This means they include quantitative (determinative) and qualitative (normative) aspects. Symbolic rating and weighting allow aggregation into the corresponding main indicator score.

3.2 Scores and weighting factors

All indicators and sub-indicators are scored on a scale from 0 to 10, independent of their quantitative / qualitative character.

To arrive at an overall score, sub-indicators and indicators are aggregated by means of a weighted sum.

For aggregating the main indicators, a set of weighting factors has been established by an expert panel. The consulted panel included 20 professionals from academia, government functions and the private sector (designers, consultants, project developers,...). The experts were selected for their authority and familiarity with problems of sustainable urban development in Flanders, the context of the research. The reference situation to be weighted was an urban development with a dominantly residential program.

Apart from assigning weighting factors, the experts were also asked for intrinsic feedback about the indicator system (cf. supra).

Table 1 resumes the averaged weightings obtained from 17 respondents.



Table 1: Rounded indicator weights and corresponding standard deviations (s.d.).

Category / Indicator	%	s.d.	Category / Indicator	%	s.d.
Environmental aspects	37,5		Socio-cultural aspects	25,5	
Materials+	5,0	1,95	Safety	3,0	1,14
Energy+	8,5	5,47	Servicing effectiveness	4,0	1,88
Water+	5,5	1,75	Integration	3,5	1,66
Land use	7,0	3,86	Sociability	3,5	1,46
Mobility	7,0	2,48	Future social value	3,5	1,89
Emissions and nuisance	4,5	1,55	Spatial quality	4,5	1,53
			Identity	3,5	1,32
Economic aspects	18,5		Process aspects	18,5	
Life cycle cost	5,0	2,17	Process quality	6,5	2,61
Economic embedding	4,5	1,65	Participation	6,5	2,98
Legal certainty	4,5	2,25	Integrity	5,5	2,73
Future economic value	4,5	1,96			

The standard deviations are fairly proportional to the weight values. As a critique, the averaged opinion of a group of experts could be judged as being 'grey' or levelled off. On the other hand, the procedure assures a higher degree of robustness while some tendencies still remain clear, in particular the dominance of the environmental category.

Particular contexts can justify a different weighting set, as far as arguments are brought in for doing so. For example, an urban development around a railway station may require increased attention for mobility, nuisance prevention and servicing effectiveness, at the cost of other factors. The assessor in charge of the evaluation (see also further) controls this contextual adaptation in consultation with the different stakeholders. The above weight factors then serve as a benchmark.

Weighting factors for sub-indicators have not yet been assigned on the basis of an expert consultation, but are derived from arguable working hypotheses.

3.3 Exclusion thresholds

In order to avoid compensation effects whereby a satisfactory overall score is obtained while severely underperforming on one or more indicators, a set of minimum requirements for sustainable functioning has been defined. A development can thus not be considered sustainable if one of the following requirements is not met:

- an appropriate location (expert judgement);
- space heating and cooling demand in new buildings under 50 kWh_{primary}/m², year (European moderate climate);
- an acceptable service level of public transport (calculated sub-indicator score of minimum 5/10);
- sustainability investments with a simple pay-back time of 5 years are automatically carried out (compared to common practice on basis of legal minimum standards).



3.4 Radar diagram

A new type of radar diagram representing the indicator scores has been developed, with the aim of reducing perception biases that are characteristic of common radar types. The biases originate from two effects:

- a strong angular and surface distortion near the centre of the circular representation;
- a different shape of score-connecting lines and figures, depending on the sequence in which the scores are represented.

The solution for this problem consists of keeping all graphical elements away from the circle centre.

For maximum legibility of the aggregated score, the radar diagram is set up as follows (see figure 2):

- the surface of a score block is proportional to the indicator's weight factor in the overall score;
- the colour of a score block represents the actual indicator score (going from red for 0, over orange for 4.1-5, to dark green for 9.1-10);
- the actual score is annotated in the score block;
- the 4 indicator categories are marked by a graphical segmentation.

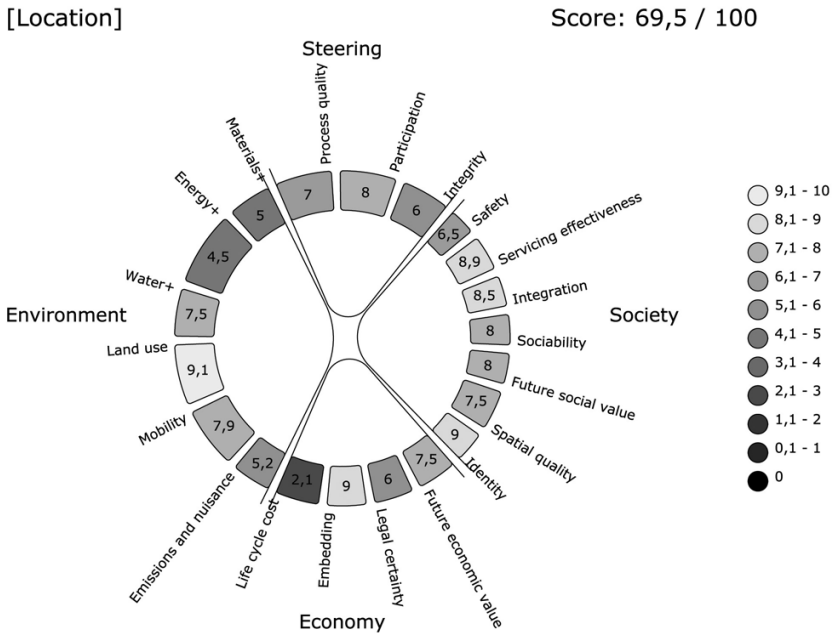


Figure 2: Radar diagram. Colour tones from red over orange to green have been replaced by shades of grey (black = 0; light grey = 10).

3.5 Argumentation

As stated higher, the present evaluation is intended to serve as a compass that facilitates an approach of reflexive governance, rather than as a gauge for awarding labels. This implies that actors and stakeholders acquire a deeper insight into the sustainable performance of an urban project and do not solely focus on end scores. Eventual barriers and shortcomings that hinder sustainable development need to be surfaced and understood, in order to enable targeted remediation.

Providing a consolidated integrity check is of particular importance. Actors and stakeholders have to verify if the sustainability claims they forward stand the proof of integer, holistic sustainable development. If this is not the case, the sectorial agendas of concerned parties may need adjustment.

Because of its debatable nature, the entire evaluation process will preferably be managed by an independent assessor. This 'quality control' partner assures that quantifiable inputs are correctly handled, qualitative judgements properly balanced, and the aspirations of the different actors and stakeholders considered by referring to a solid definition of sustainable development.

Intrinsic quality discussions hereby present the obvious risk of conflicting valuations, but also the opportunity of reaching the core issues of sustainable development. Such strategy coincides with the goal of a result-oriented instrument, rather than a means-oriented one. It avoids the pitfalls of trying to quantify the unquantifiable.

4 Feedback from the application in three case studies

In its present state, the instrument has been tested in three case studies as part of a doctoral research project. The selected case studies are an urban brownfield redevelopment (Antwerp docklands), a peri-urban infill project (Mechelen) and a greenfield development at the city fringe (Sint-Niklaas), all in the northern part of Belgium.

Some preliminary conclusions about the methodology can be drawn from these applications.

First, even if data uncertainties prevent to construct a clear image of sustainable functioning, it is still possible to obtain a useful *quick scan* of a development project. In this case the evaluation delivers a concise SWOT-analysis. In particular, a lack of data may in itself reflect a weakness of the sustainability targets, and so provide indications for improvement. The same holds for detecting misleading sustainability claims.

Second, using the tool has indicated that a lot more research about sustainable urban functioning is needed, particularly in terms of life cycle assessments at the urban scale level. In a similar way, clients need to be methodologically supported from the early programming stages so as to assure the *deep sustainability* of the development, since a common pitfall is to try to 'green up' a traditional project.



Third, a tension between local context factors and general sustainability goals may be expected to occur in any practical case. There are however no standard rules to resolve this conflict.

And finally, qualitative judgements can be made more robust if the instrument is used to compare projects or development scenarios with each other. Similar reasoning can then be applied in a systematic and comparative way to effectively differentiate between developments and their possible variants. In the case study for Mechelen, a different development scenario was proposed and subsequently analysed as part of the PhD research. This proved that an alternative approach could drastically improve sustainable performance, even while having recourse to market-conform solutions.

5 Conclusions

The compass for evaluating sustainable urban development at the neighbourhood scale presented in this paper has set out to adopt an alternative approach to sustainable performance evaluation, as compared to instruments that adhere to a strictly quantitative setup. So an important challenge was to accommodate for qualitative judgements in a structured, but non-mechanistic way.

We need now to question if the initial border conditions for the tool have been met.

A first aspect to be reconsidered concerns manipulation: can a tool be qualitative and context-sensitive without becoming an idle measure? We venture to conclude that such can be the case under certain conditions: there must be a solid frame of reference with sufficient built-in controls, and the instrument must be used by an instructed and independent assessor, however working in close collaboration with concerned actors and stakeholders for the sake of optimum transparency. This coincides with an approach of reflexive learning. The introduction of a steering category for evaluation and integration of the underlying 3P aspects is an essential component of this strategy.

Even if qualitative judging may, from a theoretical point of view, be considered as intangible by definition, it should be noted that it does occur as a common practice. Architectural or urban planning competitions provide us with an excellent instance. The qualitative choices made by competition juries are accepted even if the economic consequences for the participants are far-reaching.

The accessibility of the evaluation results to actors and stakeholders is increased by using a concise set of main indicators divided into four categories that are easy to grasp.

Practical application has indicated that data uncertainties allow, if not to compose a reasonable score, at least to perform a SWOT-analysis or quick scan. This analysis can be particularly instructive within the development process, because it indicates gaps in the frame of reference that has been put into place for a given project.

Further adaptations to the evaluation instrument can consolidate the response to its initial claims. If desirable, minimum thresholds can be turned into quantitative-only criteria by developing a set of measurable location conditions,



so that a ‘minimum label’ can be accorded in an undisputable way. An expert panel can further evaluate and weigh the sub-indicators, while new and more LCA-data could allow to support more informed calibrations of the indicators and sub-indicators. Some qualitative indicators could be more structurally assessed by an explicit enumeration and weighting of the different aspects to be judged.

Further use of the instrument in practice is currently pending, but would allow to effectively engage with its refinement.

References

- [1] Congress for the New Urbanism, Natural Resources Defense Council, U.S. Green Building Council, LEED 2009 for Neighborhood Development Rating System, USGBC, www.usgbc.org/DisplayPage.aspx?CMSPageID=148
- [2] BRE, BREEAM Communities SD5065B Technical Guidance Manual, BRE Global Ltd, [www.breeam.org/filelibrary/breeamcommunities/SD5065B - BREEAM Communities Technical Guidance Manual.pdf](http://www.breeam.org/filelibrary/breeamcommunities/SD5065B_-_BREEAM_Communities_Technical_Guidance_Manual.pdf)
- [3] Brandon, P. & Lombardi, P., *Evaluating Sustainable Development in the Built Environment*, Wiley-Blackwell: Oxford and Ames, 2011.
- [4] Vandevyvere, H., *Strategieën voor een verhoogde implementatie van duurzaam bouwen in Vlaanderen. Toepassing op het schaalniveau van het stadsfragment / Strategies Towards Increased Sustainable Building in Flanders. Application on the Scale of the Urban Fragment*, PhD dissertation, K.U.Leuven: Leuven, 2010.
- [5] Blum, A., HQE²R – Research and demonstration for assessing sustainable neighbourhood development (Chapter 20). *Sustainable Urban Development Volume 2: The Environmental Assessment Methods*, eds. M. Deakin, G. Mitchell, P. Nijkamp & R. Vreeker, Routledge: London and New York, pp. 412-428, 2007.
- [6] IVAM (2005), *DuurzaamheidsProfiel van een Locatie (DPL)*, www.ivam.uva.nl
- [7] Voss, J-P., Bauknecht, D. & Kemp, R. (eds). *Reflexive Governance for Sustainable Development*, Edward Elgar: Cheltenham, 2006.
- [8] Rotmans, J., Methods for integrated assessment: the challenges and opportunities ahead. *Environmental Model Assessment*, **3(2)**, pp. 155-179, 1998.
- [9] Lombardi, P. & Brandon, P., The Multimodal System Approach to Sustainability Planning Evaluation (Chapter 3). *Sustainable Urban Development Volume 2: The Environmental Assessment Methods*, eds. M. Deakin, G. Mitchell, P. Nijkamp & R. Vreeker, Routledge: London and New York, pp. 47-64, 2007.
- [10] Nijkamp, P., The role of evaluation in supporting a human sustainable development: a cosmomic perspective (Chapter 5). *Sustainable Urban Development Volume 2: The Environmental Assessment Methods*, eds. M.



- Deakin, G. Mitchell, P. Nijkamp & R. Vreeker, Routledge: London and New York, pp. 94-109, 2007.
- [11] Lombardi, P. & Brandon, P., Toward a multi-modal framework for evaluating the built environment quality in sustainability planning (s.n.). *Evaluation of the Built Environment for Sustainability*, eds. P. Brandon, P. Lombardi & V. Bentivegna, E&FN Spon: London, pp. 7-24, 1997.
- [12] Basden, A., The critical theory of Herman Dooyeweerd? *Journal of Information Technology*, **17(4)**, pp. 257-269, 2002.
- [13] Lombardi, P. & Basden, A., Environmental Sustainability and Information Systems: The Similarity. *Systems Practice*, **10(4)**, pp. 473-489, 1997.
- [14] Bauler, T., *Indicators for Sustainable Development: A Discussion of their Usability*, PhD dissertation, Université Libre de Bruxelles: Brussels, 2007.
- [15] ISO, International Standard ISO 14031: Environmental management - Environmental performance evaluation – Guidelines, ISO: Geneva, 1999.
- [16] Stad Stockholms/GlashusEtt, Hammarby Sjöstad – a unique environmental project in Stockholm, GlashusEtt: Stockholm, 2007.
- [17] Simmons, C., Ecological Footprint Analysis: a useful method for exploring the interaction between lifestyles and the built environment (Chapter 11). *Sustainable Urban Development Volume 2: The Environmental Assessment Methods*, eds. M. Deakin, G. Mitchell, P. Nijkamp & R. Vreeker, Routledge: London and New York, pp. 223-235, 2007.