Profile Based Comparative Analysis for AOSE Methodologies Evaluation

Luca Cernuzzi
1st author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
lcernuzz@uca.edu.py

Franco Zambonelli
2nd author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
zambonelli.franco@unimore.it

ABSTRACT
This study focuses on the analysis and evaluation of agent-oriented methodologies. Different studies have been proposed for the evaluation of agent-oriented methodologies adopting specifics types of evaluation and criteria. The present work proposes to adopt the Profile Analysis technique for comparing evaluations carried out by different authors (perhaps using different evaluation frameworks) with the aim of improve the acceptability of agent-oriented methodologies evaluation in the agent community. To exemplify the proposal, we present the application of the Profile Analysis technique on a case study.

Categories and Subject Descriptors
D.2.10 [**Software Engineering**]: Design – *Methodologies.* I.6.5 [**Simulation and Modeling**]: Model Development – Modeling *Methodologies.*

General Terms
Measurement, Reliability, Verification.

Keywords
Agent Oriented Methodologies, Evaluation, Profile Analysis

1. INTRODUCTION
The specialized literature presents different studies on the evaluation of Agent-Oriented Software Engineering (AOSE) methodologies [16], [4], [5], [8], [9] and [18]. However, attempting a comparison among the available studies we found some difficulties to deal with. The first one is that the majority of them are mainly based on qualitative evaluations which are difficult to compare. In this sense, the framework described in [3] represents a great improvement, starting from qualitative evaluations to obtain quantitative results. Another relevant problem is that some works present the evaluation of a specific AOSE methodology done by the same authors of the methodology. Despite the good intentions of such researchers, in general it is quite hard to objectively evaluate one's own work. If analyzer and evaluators (agent engineers) were to disregard these works, the number of acceptable evaluations would be quite reduced. Moreover, usually the evaluation proposals presented so far are based on different types of evaluations that are not totally congruent among them and, even worse, they adopt different criteria introducing some complications in comparing the different evaluations. Finally, none of these studies compare their results with those obtained by others authors and consequently each one offers a relative value that is harder to be accepted in the community of multi-agent researchers and developers.

The present paper focuses on the application of the Profile Analysis method to reach a greater reliability with regards to some evaluations of AOSE methodologies carried out by different authors. We are looking for evaluation similarities to confirm results and thus make them more acceptable in the agent community.

2. THE PROFILE ANALYSIS METHOD
The Profile Analysis [10], [11], and [13], is a multivariate statistical method which allows the processing and the evaluation of profiles in specific areas in order to obtain quantitative results [15]. It allows comparing two or more profiles in the same space. It is possible to define a Profile as "the results of the real or ideal description of a specific entity (human, project, institution, etc.) in terms of different dimensions, simultaneously considered, functionally or theoretically related, and which adopt the same scale of evaluation" [15]. For example, choosing a methodology, and considering we want to evaluate four specific criteria (e.g. Autonomy, Expressiveness, Project Phases, and Scalability), each criterion is a dimension and consequently each profile has four dimensions.

In each evaluation there exists a relationship among the plan (what is expected) and the result (what is implemented) that is conceptually expressed in terms of distance/proximity. A greater distance means that we are far from the expected or ideal value. Considering that a profile may have n dimensions, to measure the distance we need to use a n-dimensional space, where the set of values of the n-dimensions determine a position of a point in this space. This way, it is possible to determine the Euclidean distance between the obtained point and the ideal one.

By having different profiles it is possible to obtain a set of values, as estimation of each profile with regards to some ideal situation, to calculate the distance between two or more profiles, and to analyze the similarities in the high and low points between profiles establishing the Configurational Similarity. The coefficient of Congruency "C" may be used to quantitatively evaluate the Euclidean distance between profiles. It is defined as a the proximity between the expected or ideal situation and the obtained or real one. The values of C range from 0 to 1. Table 1 shows the absolute values of C and their interpretations.

Table 1. Interpretation of the Coefficient of Congruency C

Absolute Value of C	Interpretation
0.90 - 1.00	Virtually perfect Congruency
0.70 - 0.89	High Congruency
0.40 - 0.69	Moderate Congruency
0.20 - 0.39	Low Congruency
0.00 - 0.19	Virtually no Congruency

The coefficient of Simple Congruency is used when all the variables or dimensions have the same weight in the evaluation. The formula is [14]:

C_{ij} = Simple Congruency of the values

[1] $$C_{ij} = 1 - \frac{D_{ij}}{D_{max}}$$

D_{ij} = Euclidean Distance between profiles in a p-dimensional space

D_{max} = Maximum Distance lied for the common scale of the p-dimensions

[2] $$D_{ij} = \left[\sum_{k-1}^{p} \left(X_{ik} - X_{jk} \right)^2 \right]^{1/2}$$

[3] $$D_{\max} = T.p^{1/2}$$

Where:

k = 1,2,……,p

p = Number of dimensions of the profile

X_{ik} = Value of the profile *i* in the dimension k

X_{jk} = Value of the profile *j* in the dimension k

T = Difference between the maximum and the minimum values of the scale.

The coefficient of Weighted Congruency may be used when the variables or dimensions have different weights in the evaluation. Since all the evaluation studies presented so far for the AOSE methodologies give the same weigh to each criterion, we adopt the Simple Congruency formula.

A complementary coefficient is the Configurational Similarity which measures the grade of correspondence of the high and low values between different profiles in each dimension. When graphically represented, such values are easy to compare. It is worth to mention that different profiles may have great distances among them but still have great similarities in their configuration. The formula for the Configurational Similarity coefficient [15] is:

[4] $$ES2_{ij} = \frac{\sum_{k-1}^{p} \left| d_{1(k)} + d_{2(k)} \right|}{\left(\sum_{k-1}^{p} \left| d_{1(k)} + d_{2(k)} \right| + \sum_{k-1}^{p} \left| d_{1(k)} - d_{2(k)} \right| \right)}$$

Where:

k = 1,2,……,p

p = Number of dimensions of the profile

If k = 1:

$$d_{i(k)} = d_{j(k)} = 0$$

If 2 ≤ k ≤ p:

$$d_{i(k)} = X_{i(k-1)} - X_{i(k)}$$
$$d_{j(k)} = X_{j(k-1)} - X_{j(k)}$$

$d_{i(k)}$ = Difference between successive values of the profile X_i

$d_{j(k)}$ = Difference between successive values of the profile X_j

The same as the Congruency coefficient, the values of Configurational Similarity range from 0 to 1, being 1 the perfect matching in the evaluations of profiles and the interpretation of the values the same as the one presented in the Table 1. In Table 2 we show an example of the application of the Configurational Similarity coefficient.

Table 2. Values obtained considering two profiles for a methodology with 4 dimensions

Criteria or Dimensions	Profile 1 - Values	Profile 2 - Values
Autonomy	9	7
Expressiveness	6	3
Process Phases	7	6
Scalability	4	2

A graphical representation of the values of Table 2 is showed in Figure 1. In the graph it is possible to observe that there is a distance between the two profiles but both are similar in the high a low points, that is they have a high configurational similarity.

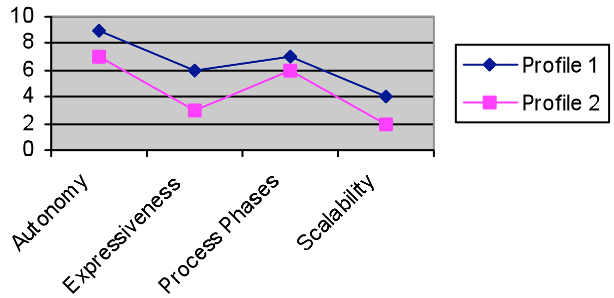

Fig. 1. Graphical representation of the profile evaluation example of Table 2

3. APPLYING THE PROFILE ANALYSIS METHOD

The Profile Analysis method, by means of the Congruency and the Configurational Similarity coefficients, allows agent-oriented engineers to easily compare different evaluations of AOSE methodologies. However, to carry out such comparison several steps are needed. First, we need to choose the evaluation of methodologies to be compared; second, we need to define a common comparative framework; and finally, we may present the obtained results and a corresponding analysis.

3.1 Choosing the Evaluations and the Methodologies

As we have seen, there are some evaluations that cover different AOSE methodologies. Some of them focus on qualitative evaluations [16], [2], [5]. However, to apply the Profile Analysis method we need quantitative evaluations. The more interesting in this category are [8], [4], [9], [18]. Among these, we select the works of Hoa Dam and Winikoff [9] and Sturm et al. [18] for diverse reasons. They are quite often referenced in the specialized literature. Moreover, both evaluate common methodologies applying sometimes the same evaluation methods (Feature Analysis, Survey, etc.). Finally, they present a partial similarity with regards to the evaluation criteria adopted.

The evaluated methodologies in the selected works are: MaSE [6], Tropos [7], the original Gaia version [19], Prometheus [12], MESSAGE [1], and OPM-MAS [17]. Among them, three are common to both works: MaSE, Tropos, and Gaia. However, both works have considered Gaia in its first version even though there is an official extension that brings substantial improvements [20]. For this reason, we decided to take into account just MaSE and Tropos for our purposes. It is relevant to mention that no one of the selected methodologies has the authorship of the evaluators.

Hoa Dam and Winikoff [9] apply the Feature-based Analysis, Survey, and Case Study types of evaluation while Sturm et al. [18] apply Feature-based Analysis, Survey, and Structured Analysis. Thus, in a first attempt we focused our study on Feature-based Analysis and Survey types of evaluation which are common for both works. However, the evaluations based on the Survey type present some limitations. Sturm et al. based their work on the experience of a 15 students course in which each methodology was used by a group of 2 or 3 students. This experience, despite the interesting information it provides, is not relevant from a statistical point of view. Moreover, the results suffer the lack of expertise of the students engaged in the experiment. In the work of Hoa Dam and Winikoff, the survey is based on the evaluation done by the authors of the methodologies complemented by an experiment with 5 students (one per each methodology to be evaluated). For the student experiment it is possible to assert the same considerations of the work by Sturm et al. On the other hand, the form fulfilled by the authors of these methodologies may offer interesting information considering the high expertise of the evaluators, but the results may be strongly influenced by the bias of the judgment. For all these reasons, we decided to focus just on the Feature-based Analysis evaluation of both works.

3.2 Defining a Common Comparative Framework

In the definition of a comparative unified evaluation framework it is necessary to do some steps: i) select the common criteria; ii) transform the qualitative attributes into quantitative values; and iii) convert the values of the attributes to a common scale previously defined. Let us analyze each step in depth.

For the first step, both evaluation works classify the attributes in four major aspects or categories which are: Concepts and Properties; Modeling and Notations; Process; and Pragmatics. However, in the attributes definition we found many discrepancies. There exist criteria with the same definition but different names, and vice versa, criteria with the same name but different semantics. Moreover, there are some attributes that in a work have been directly defined as individual while in the other comprise a set of criteria; in these cases we grouped the comprised criteria in a unique more general criterion assigning it the mean of the values of the comprised criteria. The next table shows the final set of common attributes which are the basis for the application of the comparative framework. Please observe that * means a change in the original name.

In the second step we need to transform eventual qualitative evaluations (which represent the differences between values but not their magnitudes) into quantitative values. Moreover, we need to unify the scale of evaluation. In effect, in Sturm et al. [19] the authors adopt a metric ranging from 1 to 7 for each attribute, being 1 the minimum and 7 the maximum value; while in Hoa dam and Winikoff [9] the authors adopt different metrics according to the attribute. These metrics are:

- "High - Medium - Low - None" for the attributes in the category Concepts and Properties
- "Strongly Disagree - Disagree - Agree - Neutral - Strongly Agree" for the attributes in the category of Modeling and Notations; and Pragmatics (Scalable and Maturity)
- "Yes - No" for the attribute Domain Specific in the Pragmatics category
- "0 - 1 - 2 - 3 - 3'- 4" for the attribute in the Process category.

Table 3. Selected attributes for comparing the evaluations

Hoa Dam and Winikoff, 2003	Sturm et al., 2004
Concepts and Properties	
Autonomy	Autonomy
Reactive	Reactiveness
Proactive	Proactiveness
Mental Attitudes	Mental notions (*)
Teamwork	Organization (*)
Protocols	Protocol
Modeling and Notations	
Consistency check	Analyzability
Refinement	Complexity Management
Language adequate and expressive	Expressiveness
Easy to use + Easy to learn + Clear Notation	Accessibility
Process	
Lifecycle coverage (*)	Lifecycle coverage
Pragmatics	
Domain specific	Domain applicability
Scalable	Scalability
Maturity (Quality) (*)	Resources

It is possible to observe that some of the metrics adopted by Hoa Dam and Winikoff are qualitative; therefore we propose a scale for each metric to be converted to a quantitative scale. Being such scales independent, the same value for different attributes may have different meanings. The proposed conversions are:

- None-Low-Medium-High -> 0-1-2-3
- Strongly Disagree-Disagree-Neutral-Agree-Strongly Agree -> 0-1-2-3-4
- Yes-No -> 0-1 (considering it is a reverse scale)
- 0-1-2-3-3'-4 -> 0-1-2-3-4-5

Finally, for the third step it is necessary to convert all the values to a comparable metric to obtain meaningful inferences from them. In doing such conversion it is possible to introduce some alteration in the results that may influence the Congruency and Configurational Similarity coefficients. The definition of a common scale has been amply treated in the specialized literature and there are no strong reasons to prefer a metric over the others, except in particular cases, none of which apply to our situation. Therefore, we decided to adopt a common scale varying from 0 to 10. For the conversion we calculate the percentage of the value into its original scale and then convert it to a percentage over the scale of 10. In case of a fractional result, we round it to the closest integer. Table 4 presents the conversion for the work of Sturm et al., while Table 5 presents the conversion for the scales presented in Hoa Dam and Winikoff.

Table 4. Conversion for the scale of the work of Sturm et al.

Scale Sturm et al.	Common Scale
1	0
2	2
3	3
4	5
5	7
6	8
7	10

Table 5. Conversion for the scale of the work of Hoa Dam and Winikoff

Scales Hoa Dam - Winikoff	Common Scale
Concepts and Properties	
None (0)	0
Low (1)	3
Medium (2)	7
High (3)	10
Modeling and Notations	
Strongly Disagree (0)	0
Disagree (1)	3
Neutral (2)	5
Agree (3)	8
Strongly Agree (4)	10
Process	
0	0
1	2
2	4
3	6
3'	8
4	10
Pragmatics	
Yes (0)	0
No (1)	10

3.3 Results of the Profile Analysis

Hereinafter, we present the Congruency and Configurational Similarity coefficients, the graphs with the comparisons, and a brief analysis of the results.

In Figure 2 and Figure 3 it is possible to compare the two evaluations (and the distance from the ideal profile) corresponding to MaSE and Tropos respectively.

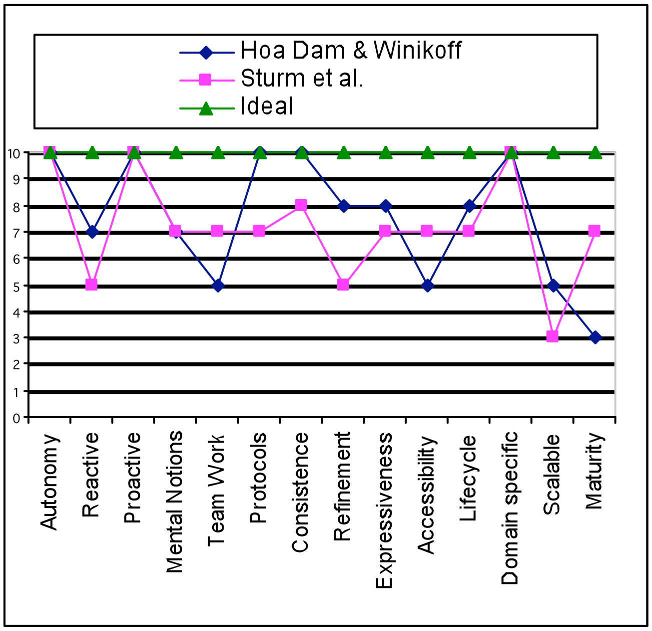

Fig. 2. Comparison for the MaSE methodology.

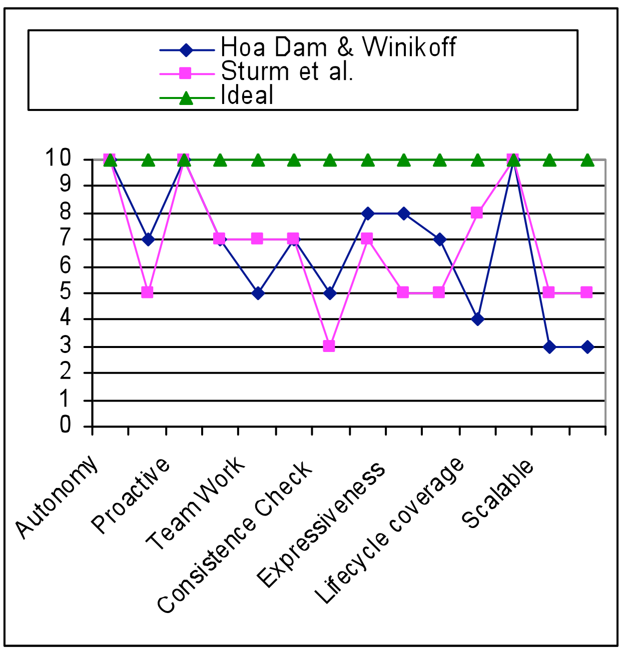

Fig. 3. Comparison for the Tropos methodology

In the next table we present the obtained results considering the entire set of criteria.

Table 6.

	MaSE	Tropos
Cwi =	0.59556	0.53478
Csi =	0.655658	0.6090305
Cws =	0. 731406	0.739506
ES2 =	0.64286	0. 66667

Where:

Cwi represents the Congruency between the ideal profile and the profile obtained from the evaluation of Hoa Dam and Winikoff [9];

Csi represents the Congruency between the ideal profile and the profile obtained from the evaluation of Sturm et al. [18];

Cws represents the Congruency between the profiles from both works;

ES2 represents the Configurational Similarity between both profiles.

Based on these results it is possible to infer some considerations:

- Both works coincide in that MaSE is more congruent than Tropos with regard to the ideal profile; moreover, the differences of the coefficient are similar for both works.
- The evaluations are consistent, that is, both works in general evaluate MaSE better than Tropos, but the distance is quite reduced between the methodologies, as reflected in the value of the Congruency coefficient (Cws) between the two profiles which reach an high level.
- The Configurational Similarity for both methodologies is moderated to high, which means that independent evaluators reach similar conclusions; this increment the reliability of the results.

For a deeper analysis it is interesting to calculate the value of Congruency and Configurational Similarity for each dimension.

Table 7. Concepts and Properties

	MaSE	Tropos
Cwi =	0.55653	0.53993
Csi =	0.705608	0.705608
Cws =	0.678545	0. 702791
ES2 =	0.65000	0.66667

Table 8. Modelling and Notations

	MaSE	Tropos
Cwi =	0.71277	0.67596
Csi =	0.657217	0.480385
Cws =	0.787868	0.787868
ES2 =	0.62500	0.71429

Table 9. Process

	MaSE	Tropos
Cwi =	0.80000	0.40000
Csi =	0.70000	0.80000
Cws =	0.90000	0.60000
ES2 =	---------	---------

Table 10. Pragmatics

	MaSE	Tropos
Cwi =	0.50334	0.42845
Csi =	0.560303	0.591752
Cws =	0.74180	0.836701
ES2 =	0.63636	0.85714

- From the Congruency coefficient for the Concepts and Properties dimension, we may observe that the evaluation of Hoa Dam and Winikoff [9] reaches the moderated level, while the evaluation of Sturm et al. [18] reaches the high level; the reason is that the authors of the first work consider that the attribute Organization is not covered by the methodologies.
- In the Concepts and Properties dimension the coefficients are basically the same (see also the graphics in Figures 2 and 3); this situation indicates that authors agree in that methodologies cover the set of criteria in a similar way.
- In the Process dimension it is possible to observe the main differences in the coefficients; being the Lifecycle coverage the only criterion which determine the values of the coefficients, it is impossible to calculate the Configurational Similarity (according to formula 4 if we try to calculate it we find a division by zero).
- In all the dimensions the Configurational Similarity is varying from moderate to high, which means we may conclude that both evaluations agree in some high and low points of the methodologies, giving more reliability to the studies. It is relevant to mention that based on more criteria it would possible to obtain more accurate results.

4. CONCLUSIONS AND FUTURE WORKS

This study focuses on the evaluation of agent-oriented methodologies. The existing evaluations do not stress in comparing their results with others carried out from different authors. It is possible to observe that such kind of comparison can improve (in the case of similar results) the reliability of the evaluations of AOSE methodologies. For carrying out this type of comparison we proposed the adoption of the Profile Analysis method. We are looking for evaluation similarities to confirm results and thus make them more acceptable in the agent community. Moreover, the application of such analysis method may help in the selection of a specific methodology with a specific level of Congruency with regards to an expected profile, being it general or limited to a particular dimension. To illustrate the application of the Profile Analysis method, we compared the evaluation done by different authors of MaSE [6] and Tropos [7]. To compare the evaluations, we have selected the common evaluators' criteria and converted each of them to a common scale.

Finally, it is possible to envision different future works to reach a greater reliability level with regard the evaluation results. Among them, it is possible to carry on similar studies to compare other independent evaluations as well as to consider further AOSE methodologies. Moreover, it may be useful to include an analysis of the Congruency and the Configurational Similarity for the evaluation of different methodologies carried out by the same evaluation group.

5. REFERENCES

[1] Caire, G., Chainho, P., Evans, R., Garijo, F., Gómez Sanz, J., Kearney, P., Leal, F., Massonet, P. and Pavón, J., 2001. Agent Oriented Analysis Using MESSAGE/UML. Proceedings of Agent-Oriented Software Engineering – AOSE 01, May 2001, Montreal (Canada), pp. 101–107

[2] Cernuzzi, L. and Giret A., 2000. Methodological Aspects in the Design of a Multi-Agent System. Proceedings of Workshop on Agent Oriented Information Systems. VII National Conference on Artificial Intelligence (AAAI-2000). Austin, USA, pp. 21–28

[3] Cernuzzi, L. and González, M., 2002. A Framework for Evaluating WIS Design Methodologies. In: Filipe, J., Sharp, B., Miranda, P. (Eds.): Enterprise Information Systems III. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 198–207.

[4] Cernuzzi, L. and Rossi, G., 2002. On The Evaluation Of Agent Oriented Methodologies. Proceedings of the OOPSLA 02 - Workshop on Agent-Oriented Methodologies, November 2002, Seattle (USA), pp. 21-30

[5] Cuesta, P., Gómez, A., González, J.C. and Rodríguez, F.J., 2003. A Framework for Evaluation of Agent Oriented Methodologies. Actas del Taller de Agentes Inteligentes en el tercer milenio (CAEPIA'2003), San Sebastián, 11 de Noviembre

[6] DeLoach, S., Wood, M. and Sparkman, C., 2001. Multiagent Systems Engineering. *International Journal of Software Engineering and Knowledge Engineering*, vol. 11, No. 3, pp. 231–258

[7] Giunchiglia, F., Mylopoulos, J., and Perini A., 2002. The Tropos Software Development Methodology: Processes, Models and Diagrams. Proceedings of Agent-Oriented Software Engineering (AOSE-2002), July 2002, Bologna (Italy), pp 63–74

[8] Hoa Dam, K., 2003. Evaluating Agent-Oriented Software Engineering Methodologies. M.Sc. Thesis, School of Computer Science and Information Technology, RMIT University, Australia

[9] Hoa Dam, K., and Winikoff, M., 2003. Comparing Agent-Oriented Methodologies. Proceedings of Agent Oriented Information Systems-AOIS'03, July 2003, Melbourne (Australia), pp. 78–93

[10] Morrison, D. F., 1976. *Multivariate Statistical Methods*. Mc Graw Hill, NY (USA)

[11] Nunnally, J., 1978. *Psychometric Theory*. McGraw Hill. NY (USA)

[12] Padgham, L. and Winikoff, M., 2002. Prometheus: A Methodology for Developing Intelligent Agents. Proceedings of the First International Conference on Autonomous Agents and Multi-Agent Systems - AAMAS '02, Third International Workshop on Agent-Oriented Software Engineering AOSE-2002, July 15, 2002, Bologna (Italy), pp. 135–146

[13] Reis, E., 1997. Estatística Multivariada Aplicada. Ediçoes Sílabo, Lisboa (Portugal)

[14] Serafini, O., 1981. Indicadores Cuantitativos de la Distancia Evaluativa: Coeficientes y Congruencia Simple (C) y Ponderada (Cp). Brasilia (Brazil)

[15] Serafini, O., 1988. Análisis de Perfiles en Ciencias de la Educación: Coeficientes ES1 y ES2 de Similaridad Configuracional entre Perfiles Cuantitativos. *Revista Paraguaya de Sociología*, n. 72, pp. 193–200

[16] Shehory, O. and Sturm, A., 2001. Evaluation of Modeling Techniques for Agent-Based Systems. Proceedings of Autonomous Agents, Montreal (Canada), pp. 624–631

[17] Sturm, A., Dori, D. and Shehory, O., 2003. Single-Model Method for Specifying Multi-Agent Systems. Proceedings of the Second International Conference on Autonomous Agents and Multi-Agent Systems - AAMAS'03, July, 2003, Melbourne (Australia), pp. 121–128

[18] Sturm, A., Dori, D., and Shehory, O., 2004. Comparative Evaluation of Agent-Oriented Methodologies. Methodologies and Software Engineering for Agent Systems, Kluwer, pp. 127–149

[19] Wooldridge, M., Jennings, N. R., and Kinny, D., 2000. The Gaia Methodology for Agent-Oriented Analysis and Design. *Journal of Autonomous Agents and Multi Agent Systems*, Vol. 3, No. 3, pp. 285–312

[20] Zambonelli, F., Wooldridge, M. and Jennings, N. R.: Developing Multiagent Systems: The Gaia Methodology. ACM Transaction on Software Engineering and Methodology, vol. 12, No. 3, (2003) 417-470