PARACONSISTENT QUALITATIVE EVALUATION: A PRACTICAL TEST

Carlos Alberto Ferreira Bispo (EESC-USP)
cafbispo@sc.usp.br
Edson Walmir Cazarini (EESC-USP)
cazarini@sc.usp.br

In continuing with the studies on paraconsistent qualitative evaluation (BISPO & CAZARINI, 2004, 2005, 2006; BISPO & GIBERTONI, 2005), the objective of this article is to reinforce concepts presented in these articles, carrying out a real eexperiment through field research. Professors from a large Higher Education Institution were studied. The research was divided into four parts allowing the observation of three distinct scenarios. By comparing them, it was possible to show in practice the effectiveness of the technique developed to detect distortions in data obtained in qualitative evaluations.

Keywords: Qualitative Data, Paraconsistent Qualitative Evaluation, Two-Value Annotated Paraconsistent Logic

1. Introduction

Qualitative evaluations are susceptible to flawed personal interpretation, according to Pereira (2004), Pidd (1998), Patton (1990), Peirce (1972), Mazzotti & Gewandsznajder (1998). In Bispo & Cazarini (2004, 2005, 2006) and in Bispo & Gibertoni (2005), the paraconsistent qualitative evaluation technique was introduced that can detect some flaws and attempt to reduce their effects. In these four articles, only simulations were executed during technical development, aimed at demonstrating its applications and effectiveness. In Bispo & Cazarini (2004), a simulation of software quality developed with Software Engineering Techniques was made. In Bispo & Gibertoni (2005) a quality evaluation of an e-commerce site was made. In Bispo & Cazarini (2005), when the technique was improved, a simulation was made of the evaluation of an Occupational Health and Safety Assessment System based on OHSAS 18001 standard. The height of the technique was achieved in Bispo & Cazarini (2006), where the simulation of the evaluation of an Environmental Management System was made.

The purpose of this article is to reinforce concepts employed in the four previous articles, carrying out a real experiment through field research.

2. The paraconsistent qualitative evaluation

The foundations of paraconsistent qualitative evaluation were derived from the Two-Value Annotated Paraconsistent Logic – 2VAPL. According to Costa *et al*. (1999) and Carvalho, Brunstein & Abe (2003), the 2VAPL is a ramification of Annotated Paraconsistent Logic, elaborated by Abe (1992), which in turn is a ramification of Paraconsistent Logic, elaborated by Costa & Subrahmaniam (1989) and improved in Costa, Abe & Subrahmaniam (1991), Costa, Vago & Subrahmaniam (1991) and Costa & Abe (2000). The 2VAPL is a bivalued logic, where the values for favorable evidence degrees and contrary evidence degrees are adopted for each variable in its annotation technique: $\boldsymbol{p(\mu_1, \mu_2)}$, where μ_1, μ_2 belong to the closed interval [0;1]. The $\boldsymbol{p}$ variable represents the item that is being evaluated; the annotation μ_1 represents the favorable evidence degree, while the annotation μ_2 represents the contrary evidence degree. After obtaining the values for the favorable evidence degree (μ_1) and the contrary evidence degree (μ_2), the Two-Value Annotated Paraconsistent Logic has a para-analyzer algorithm to classify those values in twelve distinct results: t (true); f (false); $\perp$ (uncertain); $\top$ (inconsistent); $A\top \rightarrow f$ (almost inconsistent tending towards false); $A\top \rightarrow t$ (almost inconsistent tending towards true); $A\perp \rightarrow f$ (almost uncertain tending towards false); $A\perp \rightarrow t$ (almost uncertain tending towards true); $Af \rightarrow \top$ (almost false tending towards inconsistent); $Af \rightarrow \perp$ (almost false tending towards uncertain); $At \rightarrow \top$ (almost true tending towards inconsistent); $At \rightarrow \perp$ (almost true tending towards uncertain).

Among the twelve possible results, four are extreme values, and are the only ones used in traditional qualitative evaluations: true, false, uncertain and inconsistent. However, in these evaluations, when there is an uncertainty or an inconsistency, it is disregarded or forgotten, if the evaluated system can proceed without it, even if they present results a little far from reality. The eight intermediate values are exclusive to the 2VAPL. With all these values, this logic allows working naturally on inconsistency and uncertainty and obtaining more precise results, closer to reality.

Costa *et al*. (1999) created a tool to aid the graphic visualization of data in those twelve classifications. It is treated in a diagram called Unitary Picture of the Cartesian Plan – UPCP. Figures 1 to 3 use this tool to plot the data obtained in the four phases of the research.

3. Conducted field research

The objective of this study is to test the null hypothesis H_0: $(\mu_1 + \mu_2) \neq 10$; the alternative hypothesis H_1 is: $(\mu_1 + \mu_2) = 10$. In other words, we intend to prove that the sum of the favorable evidence degree and contrary evidence degree is not complemental totaling 10, when they are obtained from different sources and evaluating the same item, but rather each source measuring one of the evidences, as preconized by the paraconsistent qualitative evaluation delineated in Bispo & Cazarini (2004, 2005, 2006) and in Bispo & Gibertoni (2005).

The chosen universe (or population), due to the ease in obtaining a sample, are academic professors. According to data from 2006 obtained from the National System of Evaluation of Superior Education – SINAES site, belonging to the National Institute for Education Studies and Research – INEP of the Ministry of Education, that year there were 242,795 professors of higher education in Brazil, comprised of 58,618 (24.14%) doctors (PhD.s), 86,294 (35.54%) masters, 70,554 (29.05%) specialists and 27,289 (11.24%) bachelors. The data can be found at http://sinaes.inep.gov.br/sinais (accessed in April 24, 2007).

The research was carried out during the first semester of 2006. The sample was comprised of professors in the Administration Course of an Institution of Higher Education, located in the state of Sao Paulo. The institution had 86 professors at that time, all invited to participate in the study and given a questionnaire. According to Human Resources at the institution, the group was comprised of: 21 (24.41%) doctors (PhD.s), 45 (52.32%) masters, 9 (10.46%) specialists and 11 (12.79%) bachelor's. Of that total, 72 (83.72%) questionnaires were returned. That sample was not stratified due to the form of application of the research. The questionnaires were given personally and voluntary and participation in the study was not obligatory. The questionnaire was returned to a neutral place. Therefore, there was no control over who answered the questionnaire or not. Since the participants' percentage was high (83.72%), stratification of the sample is considered to be practically identical to the one for the group. Other data that could be used in stratification, such as age, race, income and gender, available from the SINAES site to stratify the study's population, could not be supplied by the institution's Human Resources. Therefore, the sample is not perfectly faithful to the characteristics of the chosen population. However, for purposes of this article, it is considered that the sample assists in the needs of the studies.

To discover the established significance level and the estimate error allowed, according to Richardson (1999), the sample is considered stratified and the formula to be used for those calculations is used for infinite universes (above 100,000 elements). Since the sample is small, the established significance level is 90% and estimate error allowed is 10%. Both indicators would be incompatible if the intention of the study was to make inferences on the two items addressed in thestudy, i.e. network television and printed newspapers, both in Brazil. However, since the inferences one wants to make refer to the behavior of the data obtained in relation to the traditional qualitative evaluation *versus* the qualitative paraconsistent evaluation, the obtained results are considered satisfactory for the obtained sample.

4. Data survey

Two kinds of forms were used (A and B) to carry out the study. The study was divided into four parts in both forms. The first two parts were common in the two kinds of forms, whereas the two final parts were what differentiated them. In the four parts of the study, the participants were asked to qualitatively evaluate, i.e., attribute a score to the six items of each part of the research.

In the first part of the two types of forms, participants were asked to qualitatively evaluate, scoring from 0 (awful) to 10 (exceptional), with one decimal digit, the positive aspects of the items related to network television in Brazil, in a generic manner, and without specifying the channel, thus obtained the favorable evidences. The questions were: 1. Regarding program quality; 2. Suitability of programs for the different age groups; 3. Services provided to Society (participation in social campaigns, Public Health, etc.); 4. Assistance to education, culture and sports provided to Society as a whole; 5. Disclosure of the most varied Brazilian regional cultures; 6. The variety of programs to satisfy the most varied types of audiences.

In the second part of the two form types, they were asked to evaluate the same items in terms of their negative aspects (the bad aspects or flaws of each evaluated item), thus obtained the contrary evidences, and giving them a score from 0 (there is no negative aspect) to 10 (very negative aspect), with one decimal digit. This evaluation was to take took place completely independent from the evaluation of the first part of the form.

In the last two parts of the study, half of the forms were type A (third part) and the rest type B (fourth part). The forms were intercalated after printing in a type A form sequence, followed by a type B and so on. These were distributed to participants in the exact sequence of this combination. Therefore, in the first two phases, 72 scores were obtained for each question. In the last two phases, the 72 participants in the study were divided among the two types of forms, forming two groups of 36 people. Each group participated in a phase of the research, the third or forth phase. In the two types of forms, participants were asked to evaluate the printed media (newspapers) in Brazil, in a generic manner, and without specifying any of them. In the type A form, an evaluation of six items concerning positive aspects of this media was requested, thus obtained the favorable evidences. Type B form requested an evaluation of the same six items but regarding their negative aspects (bad aspects or flaws), thus obtaining the contrary evidences.

For the type A form a score from 0 (awful) to 10 (exceptional) was requested, with one decimal digit, for each of the six items: 1. Regarding information quality; 2. Regarding information reliability; 3. In-depth level of reports; 4. Impartiality of the news; 5. Services provided to Society; 6. Variety of published items satisfying different types of readers.

For the type B form a score from 0 (there is no negative aspect) to 10 (very negative aspect) was requested, with one decimal digit, for the same six items above.

5. Results obtained

Due to the limited article space, not all of the research steps will be presented, going straight to the presentation of results obtained, already in the UPCP.

The 72 scores given in the six questions of the first part of the research (432 scores) are the favorable evidences **(μ_1)**. For each score obtained, the respective complement of 10 was extracted **(μ_c)**, i.e. $\mu_c = 10 - \mu_1$. In Figure 1, all pair of values ($p_(\mu_1, \mu_c)$) appear, separate for each question. Seventy-two values do not appear in each picture of Figure 1, because many scores were equal. Values of $p_(\mu_1, \mu_c)$ appear aligned because the sum of their degrees equals 10, i.e. $\mu_1 + \mu_c = 10$. This is typically the behavior of data in traditional qualitative evaluations (BISPO & CAZARINI, 2004, 2005, 2006; BISPO & GIBERTONI, 2005).

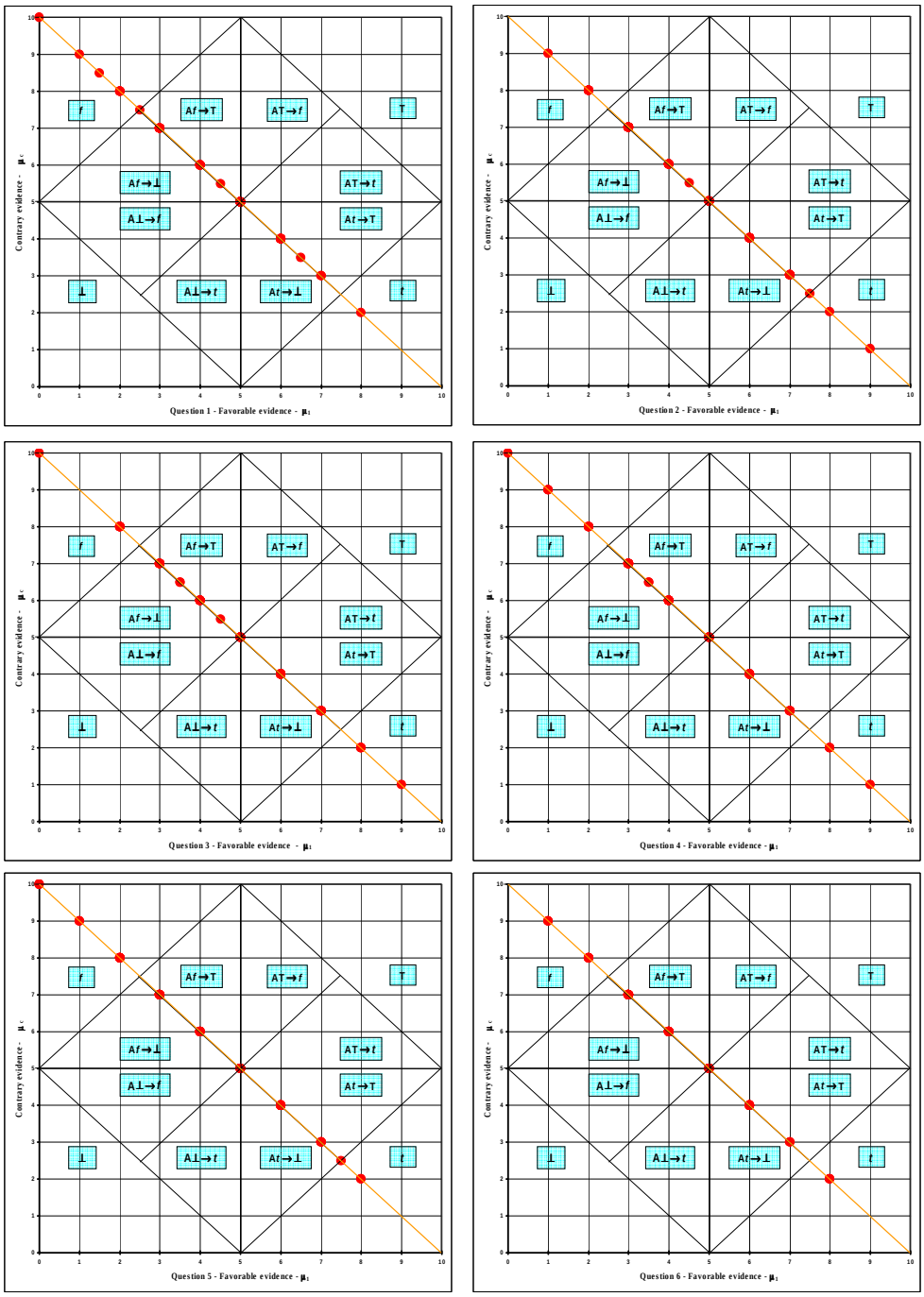

Figure 1 – Values of pairs of data $p(\mu_1, \mu_c)$ regarding each question of the first part of the study

In Figure 2, the data from the first part of the study (favorable evidences – μ_1) were crossed with the data from the second part (contrary evidences – μ_2), originating in 432 new pairs of data, now $p(\mu_1, \mu_2)$. Each pair of data was obtained from the same appraiser. As occurred in Figure 1, seventy-two values are also not shown in each picture of Figure 2 due to the fact that several identical evaluations exist, even in that part of the research. However, more values appear if comparing to the pictures from Figure 1.

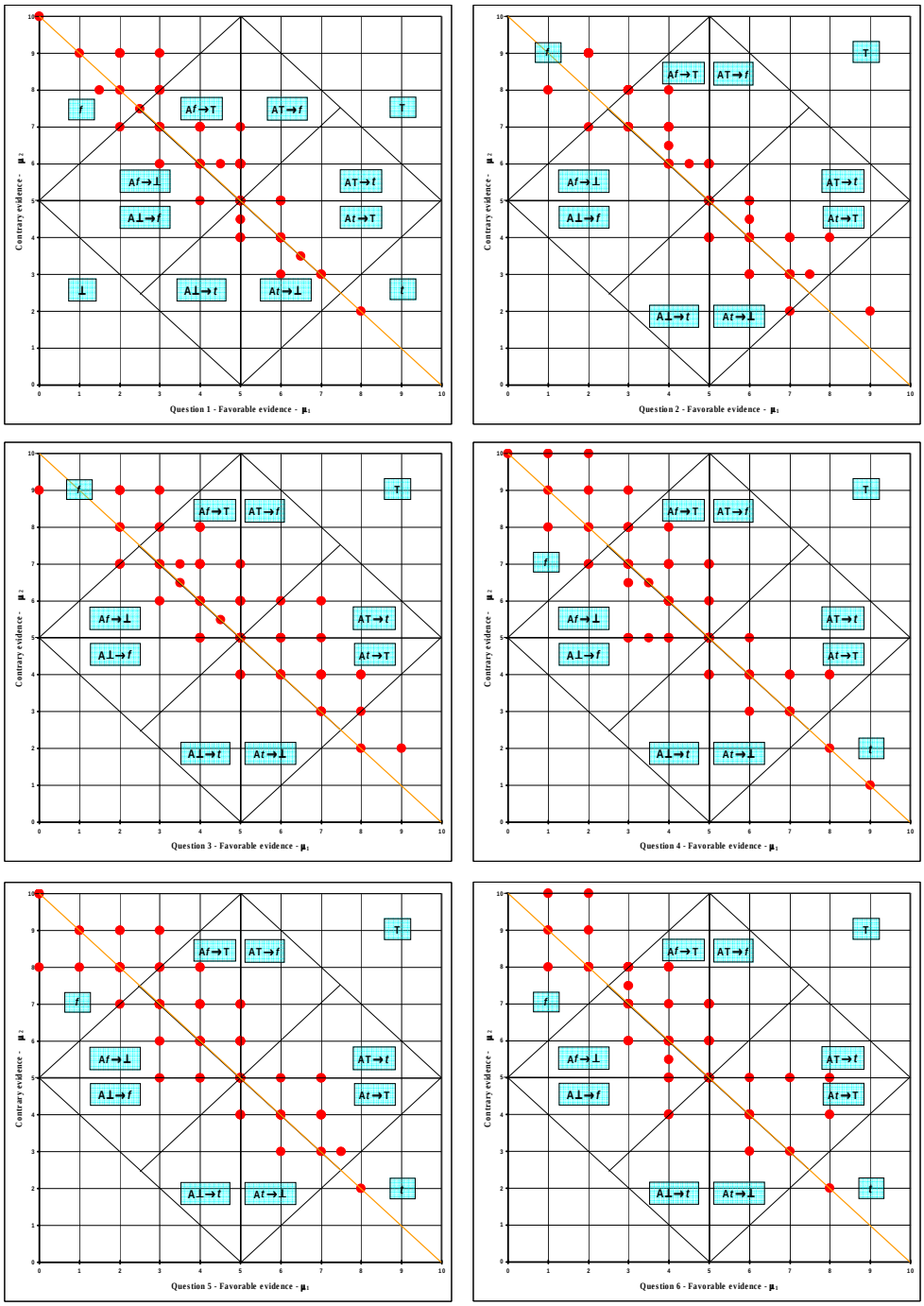

Figure 2 – Values of the pairs of data $p(\mu_1, \mu_2)$ regarding the combination of data from first and second part of the study

In Figure 3, the data from the 36 answers of each of the six questions in the third part of the type A form (favorable evidences) were crossed with the 36 answers for each of the six questions in the third part of the type B form (fourth part of the study, contrary evidences – μ_2) totaling 216 pairs of data $p(\mu_1, \mu_2)$. Each pair of data was obtained from two different people. As occurred in Figures 1 and 2, seventy-two values are not plotted in each picture of Figure 3, because some identical evaluations also occurred even in this part of this study.

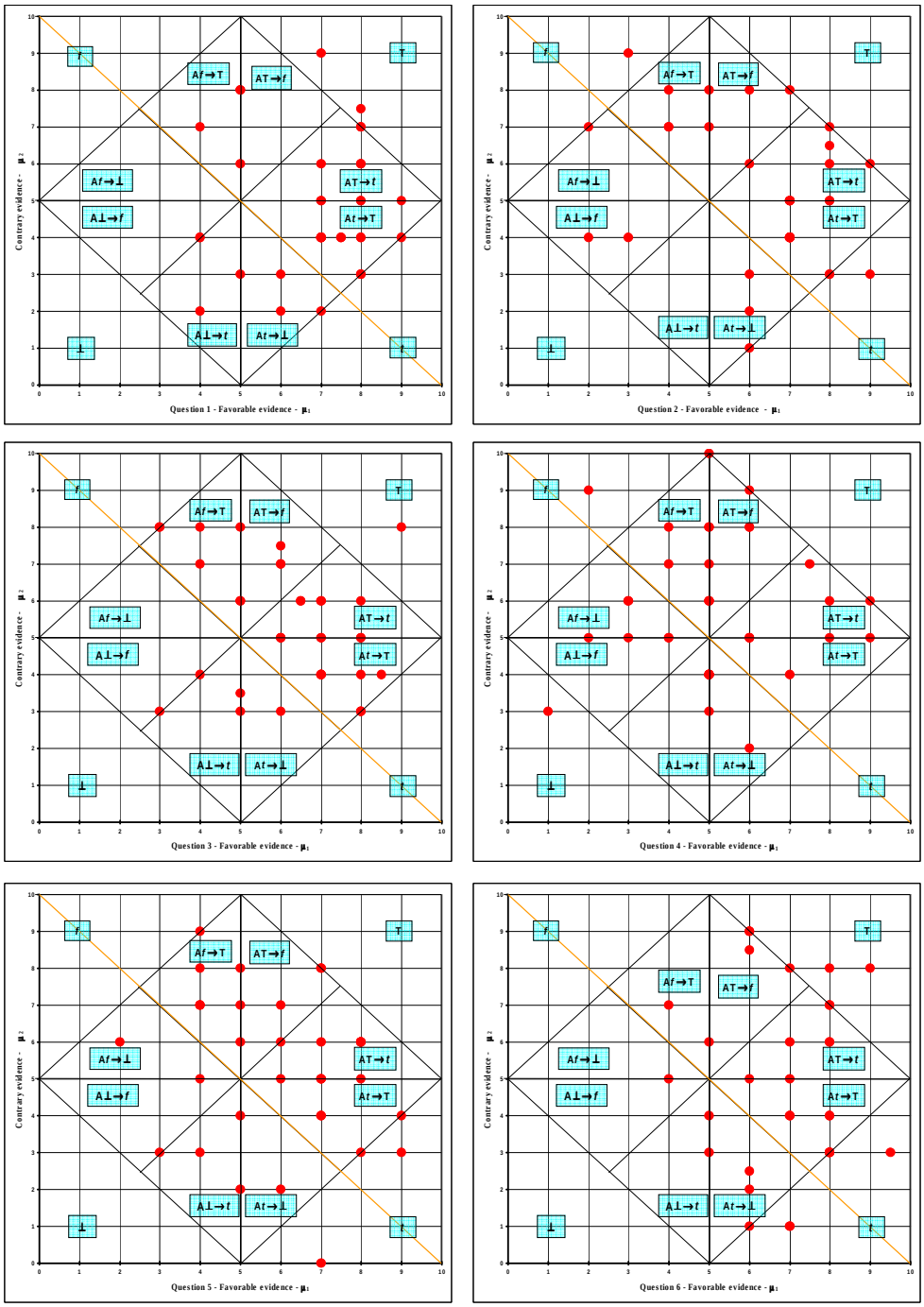

Figure 3 – Values of pair of data $p_{(\mu_1, \mu_2)}$ regarding the combination of data from the third and fourth part of the study

6. Analysis of results obtained

By comparing the three figures, it is possible to clearly observe that the behavior of the data is very different, showing three different settings, despite involving the same people who participated in the study.

The setting of Figure 1 is represented by traditional qualitative evaluation, which measures only the positive aspects or favorable evidences. The distortions among the values are solved by calculating the average for obtained values. Eventual errors can made in the evaluations

for more as well as for less, and therefore, these would be cancel each other out, with the average being the most appropriate value to represent the final result of collected data, corroborated by the statistics that give it due support. However, not all of the types of possible distortions are captured in that way.

The setting presented by Figure 2 already uses the double evaluation created by the paraconsistent qualitative evaluation, i.e., a pair of data is obtained, first the favorable evidence (μ_1) and after the contrary evidence (μ_2) for each evaluation. Usually, if there is no stimulus for this double evaluation to be conducted entirely independently, each evaluator will attribute complementary scores of 10 for each evaluation made, i.e. $\mu_2 = \mu_c = 10 - \mu_1$, obtaining similar results to those presented in Figure 1. Encouraging the two evaluations to be conducted entirely independently, although for the same person, it is possible to obtain the setting presented in Figure 2. The sum of the values for favorable and contrary evidences degrees of each item evaluated no longer results in 10, i.e., not all the values for the favorable evidence (μ_1) and contrary evidence (μ_2) of each appraised item result in 10 ($\mu_1 + \mu_2 \neq 10$). Thus, in this setting, we already have the partial non-rejection of the null hypothesis H_0, and the values shown are already closer to the real properties or characteristics of the appraised "object", according to the four previously published articles.

The setting presented by Figure 3 employs exactly what is proposed by the paraconsistent qualitative evaluation technique, as extolled in Bispo & Cazarini (2004, 2005, 2006) and Bispo & Gibertoni (2005). For each pair of attributed scores ($p_{(}\mu_1, \mu_2)$), the favorable evidence (μ_1) and contrary evidence (μ_2) are obtained by two distinct people, respectively, comparing their positive and negative interpretations concerning the appraised "object". Therefore, the values presented in Figure 3 are different from the two previous figures and are quite dispersed by the UPCP. No point plotted in the UPCP is plotted on the central line, as occurred in the Figure 1, i. e. $\mu_1 + \mu_2 = 10$. In all 216 pair of data $(p_{(}\mu_1, \mu_2))$ occurred $\mu_1 + \mu_2 \neq 10$, definitively providing the non-rejection of the null hypothesis H_0. Thus, it is believed that the presented values are much closer to the real properties or characteristics of the appraised "object", reinforcing in practice the studies carried out in the four previously published articles.

7. Conclusions

Comparing the three presented settings, it was possible to prove in practice the theory that paraconsistent qualitative evaluation allows to capture distortions in qualitative evaluations that the qualitative evaluation called traditional does not get accomplish. Thus, it is believed that the results obtained by the paraconsistent qualitative evaluation arrive much closer of real properties and characteristics of the appraised "object". That theory was raised by studies made by Bispo & Cazarini (2004, 2005, 2006) and Bispo & Gibertoni (2005) and corroborated in practice for the current article through the non-rejection of the null hypothesis H_0 in the abbreviation accomplished experiment.

The discovery of distortions in the pair of evaluations $(p_{(}\mu_1, \mu_2))$ permits the investigation of their own evaluations, to find out if distortions happened in the survey of favorable evidences or of contrary evidences. Both cases provided, for instance, the maximization or excessive minimization of the attributed value. Thus, error in the appraised "object" as well as flaws in their own evaluations can be ascertained. That is the differential in the paraconsistent qualitative evaluation.

Other real experiments need to be carried out to validate the paraconsistent qualitative

evaluation definitively, and they will be carried out soon. Researchers and managers are invited to that test, as well as to prove in theory and in practice the advantages of that type of qualitative evaluation.

References

ABE, J. M. *Fundamentos da lógica anotada*. Sao Paulo, School of Philosophy, Literature and Human Sciences of University of Sao Paulo, 1992. 139 p. PhD Thesis (Doctorate in Philosophy).

BISPO, C. A. F. & CAZARINI, E. W. A paraconsistent evaluation of software quality according to the McCall factors. In: X International Conference on Industrial Engeneering and Operation Management – X ICEOM, 2004. Florianópolis: Brazilian Association of Industrial Engineering, 2004. *Proceedings*, p. 227-234.

BISPO, C. A. F. & CAZARINI, E. W. Paraconsistent qualitative evaluation of occupational health and safety assessment system. In: XI International Conference on Industrial Engeneering and Operation Management – XI ICEOM, 2005. Porto Alegre: Brazilian Association of Industrial Engineering, 2005. *Proceedings*, p. 163-170.

BISPO, C. A. F. & GIBERTONI, D. Uma avaliação paraconsistente da qualidade de um sítio de comércio eletrônico. In: Conferência IADIS Ibero-Americana WWW/Internet 2005. Lisbon (Portugal): International Association for Development of the Information Society, 2005. *Proceedings*, p. 442-445.

BISPO, C. A. F. & CAZARINI, E. W. Avaliação qualitativa paraconsistente do processo de implantação de um sistema de gestão ambiental. *Revista Gestão & Produção*, v. 13, n. 1, p. 117-127.

CARVALHO, F. R. & BRUNSTEIN, I. & ABE, J. M. Um estudo de tomada de decisão baseado em lógica paraconsistente anotada: avaliação do projeto de uma fábrica. *Revista Pesquisa e Desenvolvimento em Engenharia de Produção*, 2003, n. 1, p. 47-62. Scholar Google Database. Can be found at: http://www.revista-ped.unifei.edu.br/n1_art04.pdf. Accessed in Dec. 11, 2006.

COSTA, N. C. A. & SUBRAHMANIAM, V. S. Paraconsisntent logics as a formalism for reasoning about inconsistent knowledge bases. *Journal of Artificial Intelligence in Medicine*, 1989, v. 1, p. 167-174.

COSTA, N. C. A. & ABE, J. M. & SUBRAHMANIAM, V. S. Remarks on annotated logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, Berlin, 1991, v. 37, p. 561-570.

COSTA, N. C. A. & SUBRAHMANIAM, V. S. & VAGO, C. The paraconsistent logic Pτ. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, Berlin, 1991, v. 37, p. 139-148.

COSTA, N. C. A. & ABE, J. M. & SILVA Fº, J. I. & MUROLO, A. C. & LEITE, C. F. S. *Lógica paraconsistente aplicada*. São Paulo: Atlas, 1999. 214 p. ISBN: 85-224-2218-4.

COSTA, N. C. A. & ABE, J. M. Paraconsistência em informática e inteligência artificial. *Estudos Avançados*, 2000, v. 14, n. 39, p. 161-174. ISSN 0-103-4014. Scielo Database. Doi: 10.1590/S0103-40142000000200012 Can be found in: http://www.scielo.br/scielo.php Accessing in Dec. 11, 2006.

MAZZOTTI, A. J. A. & GEWANDSZNAJDER, F. *O método nas ciências sociais: pesquisa quantitativa e qualitativa*. São Paulo: Pioneira , 1998. 204 p. ISBN 85-22101-3-37.

PATTON, M. Q. *Qualitative Evaluation and Research Methods*. Newbury Park: Sage Publication, 1990. 532 p. ISBN 0-8039-3779-2.

PATTON, M. Q. *Qualitative Evaluation and Research Methods*. 3. ed. California/EUA, Sage Publication, 2002. 598 p. ISBN: 0-7619-1971-6.

PEIRCE, C. S. *Semiótica e filosofia*. São Paulo: Cultrix, 1972.

PEREIRA, J. C. R. *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais*. 3. ed. São Paulo: Editora da Universidade de São Paulo – EDUSP, 2004. 156 p. ISBN: 85-314-0523-8.

PIDD, M. *Modelagem empresarial: ferramentas para tomada de decisão*. Porto Alegre: Bookman, 1998. 316 p. ISBN: 85-730-7352-7.