ISAHP 2005, Honolulu, Hawaii, July 8-10, 2003

A STUDY ON THE UTILIZATION OF COMPATIBILITY METRIC IN THE AHP: APPLYING TO SOFTWARE PROCESS ASSESSMENTS

Min-Suk Yoon
Yosu National University
San 96-1 Dundeok-dong Yeosu City
Jeonnam, Korea, 550-749
msyoon@yosu.ac.kr

Ho-Won Jung
Korea University
5-1 Anam-dong Sungbuk-gu
Seoul, Korea, 136-701
hwjung@korea.ac.kr

Keywords: compatibility, group decision making, software process assessments, conflict resolution

Summary*: This study intends to extend the applicability of the Analytic Hierarchy Process (AHP) to software process assessments. Recently an AHP approach has been successfully applied to software process assessments especially with regard to boundary problems, ambiguities between two adjacent ratings. Boundary problems may cause a problem in the process of inter-rater agreement for a process attribute rating among assessors. However, the approach assumes that the assessors reach a consensus for priorities of the associated practices. When assessors cannot reach a priority consensus, a more systematic method is required to make a consensus among those assessors having different sets of priorities. In order to solve this problem, this study proposes a consolidating method among conflicting assessors using compatibility metric of the AHP, and shows its application steps.*

1. Introduction

For about three decades, the usefulness of the Analytic Hierarchy Process (AHP) has been verified by rich applications, not only in management science areas (AHP special issues, 1990), but also in software research areas such as software reliability application (Zahedi and Ashrafi, 1991), software requirement analysis (Karlsson and Ryan, 1997), reusable component evaluation (Kontio, Caldiera, and Basili, 1996), software product evaluation (Jung and Choi, 1999), software quality evaluation (Yoon, 1997), Commercial- off-the-shelf software selection (Maiden, Ncube, and Moore, 1997) and others.

Recently, Jung (2000) proposed a method that utilizes the AHP, in order to deal with boundary problems in software process assessments based on ISO/IEC TR 2 15504 (1998), which was developed to meet the needs and requirements for a standardized software process assessment. The AHP approach was helpful in objectively justifying the rating processes and in increasing the confidence of assessment results. Furthermore, the AHP approach could be useful in tackling boundary problems, which used to occur when some difficulties would arise in understanding the boundary between two adjacent ratings: “Partially" and "Largely", "Largely" and Fully", and "Not Achieved" and "Partially". Many assessors have experienced such boundary problems (El Emam and Jung, 2001).

Under the circumstances of boundary problems, determining the capability of a software process would become a predicament. The ambiguity in rating can usually lead to inter-rater disagreement; different ratings for a process attribute (PA) among assessors and an increase in assessment effort. Reaching a consensus in ratings among assessors was the most influential factor to assessors’ effort (El Emam et al.,

1998). Therefore the AHP approach seems be more effective in solving ambiguous problems that require a systematic approach, in addition to re-investigating evidence and collecting more information.

The aim of this study is to extend the applicability of the AHP, when used in solving boundary problems. The situation being focused on is inter-rater disagreement in rating a process, which is caused by disagreement on priorities [weights], in spite of agreement on achievements, of a set of associated practice indicators (hereafter referred to as PIs). Even though the AHP is employed in this kind of boundary problem, if the assessors do not easily reach a consensus on the priorities, the critical research question still arises: How can a consensus be derived from assessors with obviously different sets of pairwise comparison matrices in evaluating priorities of a set of PIs? In order to answer this question, a systematic approach is required. This study proposes a consolidation method of the inter-rater conflicting priorities, utilizing the compatibility index of the AHP for group judgment situations.

The remainder of this paper is organized as follows: Section 2 overviews software process assessments utilizing the AHP. Section 3 briefly addresses the inter-rater disagreement rationale and presents an extension of the AHP approach, which is our revised method using the compatibility of the AHP. Section 4 illustrates the application steps of our method, and Section 5 concludes this paper with final remarks.

2. Software Process Assessment Using the AHP

2.1 A brief review of ISO/IEC 15504-2

The architecture of the published ISO/IEC 15504-2 standard consists of both process and capability dimensions. Figure 1 shows the structure of the two dimensions. In the process dimension, the processes associated with software are defined and classified in the process reference models such as ISO/IEC 12207 (2004) and CMMI (2002). For example, ISO/IEC 12207 (2004) provides five categories of software processes; *Customer- Supply process, Engineering process, Support process, Management process, Organization process* categories.

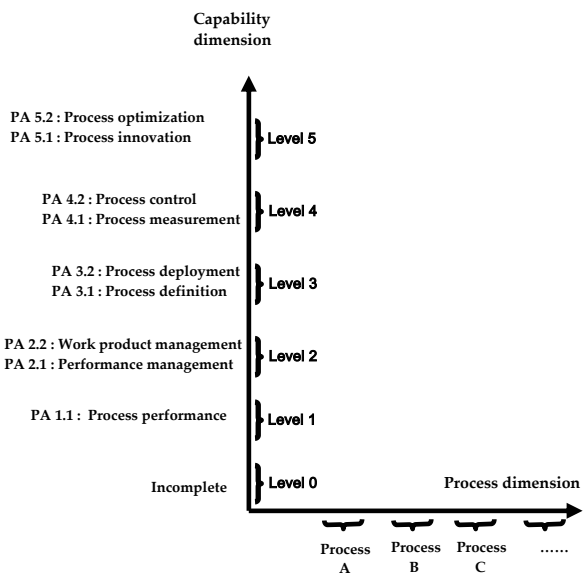

Figure 1 - Two-dimensional architecture of ISO/IEC 15504

The capability dimension is comprised of six capability levels ranging from 0 to 5. The greater the level is, the greater the process capability achieved. Each level of process capability is depicted by one or two process attributes (PAs) as shown in Figure 1, applicable to any software process, which represent

measurable characteristics necessary to manage a process and to improve its performance capability. The capability level is summarized as follows:

- **Level 0, Incomplete:** The process is not implemented, or fails to achieve its process purpose. At this level there is little or no evidence of any systematic achievement of the process purpose.
- **Level 1, Performed:** The implemented process achieves its process purpose.
- **Level 2, Managed:** The previously described *Performed process* is now implemented in a managed fashion (planned, monitored and adjusted) and its work products are appropriately established, controlled and maintained.
- **Level 3, Established:** The previously described *Managed process* is now implemented using a defined process that is based upon a standard process and that is capable of achieving its process outcomes.
- **Level 4, Predictable:** The previously described *Established process* now operates within defined limits to achieve its process outcomes.
- **Level 5, Optimizing:** The previously described *Predictable process* is continuously improved to meet relevant current and projected business goals.

The capability level is determined by measuring its PAs and each PA is measured by an ordinal rating *Fully, Largely, Partially,* or *Not achieved* as in Table 1. The ordinal rating scale is a transformation of a numerical value between 0% and 100% that represents the extent of achievement of a PA. The achievement of capability level *k* indicates that all PAs prior to level *k* satisfy the rating *Fully* and then level *k*'s attributes are rated as *Fully* or *Largely*.

Table 1 - The rating scale of the process attributes (ISO/IEC 15504-2)

Acronym	Achievement of the defined attribute
N: Not achieved	0% to 15%: There is little or no evidence of achievement of the defined attribute in the assessed process.
P: Partially achieved	16% to 50%: There is evidence of a sound systematic approach to and achievement of the defined attribute in the accessed process. Some aspects of achievement may be unpredictable.
L: Largely achieved	51% to 85%: There is evidence of a sound systematic approach to and significant achievement of the defined attribute in the accessed process. Performance of the process may vary in some areas or work units.
F: Fully achieved	86% to 100%: There is evidence of a complete and systematic approach to and full achievement of the defined attribute in the assessed process. No significant weaknesses exist across the defined organization unit.

The existence of base practices and work products provide evidence of the performance of the processes associated with them (PA 1.1). The extent of the achievement for PA 2.1 to PA 5.2 in the capability dimension have a set of associated process attribute indicators, which provides an indication of the extent of achievement of the attribute in the instantiated process. These indicators concern significant activities, resources or results associated with the achievement of the attribute purpose by a process. PIs are the principal indicators of process capability. Most of PIs are instances of generic practices defined in the exemplar model of ISO 15504-5 (2003).

2.2 AHP application for boundary problems

If assessors cannot easily reach consensus of rating and they agree on the use of the AHP approach, this step is activated. The generally outlined the AHP steps are followed in order to determine the rating of a PA with associated its indicators. The AHP starts by breaking down the decision problem into interrelated decision elements (criteria, attributes, or factors). Note the decision problem corresponds to rating of a PA and the AHP criteria do to PIs belonging to the PA.

The composite value of weights and achieved values of PIs is transformed into the rating of a PA as defined in Table 1. In order to obtain weights for a set of *n* PIs, the AHP begins with construction of a

pairwise comparison matrix according to the relative importance of the PIs. The comparison matrix is defined as $A = (a_{ij})_{n \times n}$. The scale a_{ij} is an estimate for w_i / w_j , where w_i and w_j indicate the importance of the *i*th and the *j*th PI, respectively. The matrix A has positive entries everywhere and satisfies the reciprocal property, $a_{ji} = 1/a_{ij}$, which is called a reciprocal matrix. The weight vector of PIs is generated by the following equation:

$$Aw = \lambda_{\max} w \tag{1}$$

where $\lambda_{\max}$ is the principle eigenvalue of the matrix A and $w^T = (w_1, w_2, \cdots, w_n)$ is the weight vector, corresponding to $\lambda_{\max}$. To make the vector unique, let the sum of weights be 1.

If a given judgment matrix, $A = (a_{ij})_{n \times n}$ is perfectly consistent, *i.e.*, $a_{ji} = 1/a_{ij}$ and $a_{ik} \times a_{kj} = a_{ij}$ for all *i, j, k,* then the principle eigenvalue, $\lambda_{\max}$, equals *n*. However, in a real decision making environment, people's perceived relative preferences in pairwise comparisons remain inconsistent and intransitive. A small perturbation around A leads to an eigenvalue problem, $\lambda_{\max} \geq n$ from equation (1). Inconsistency throughout the matrix A can be captured by a single number, $\lambda_{\max} - n$, which measures the deviation of the judgments from the consistent approximation and leads to the consistency index (CI) as follows;

$$CI = (\lambda_{\max} - n)/(n-1) \tag{2}$$

Random index (RI) is obtained as an average over CI values from large number of randomized reciprocal matrices (RI will be seen later in Table 2). The ratio between the two indices (CI / RI) is defined as consistency ratio (CR). If CR is less than or equal to 0.1, Saaty (1980) recommends that the estimate of the weights be accepted. Otherwise, the comparison matrix re-created. This process is repeated until the threshold condition is satisfied.

Finally, the AHP approach aggregates the weights and the measured values of PIs as follows:

$$R = \sum_{j=1}^{n} w_j r_j \tag{3}$$

where R is the composite value and r_j is the measured value for the *j*th PI. If all r_j s are measured by a value between 0 and 1, then the value R also has a value between 0 and 1. Hence, the value R is automatically transformed to a rating of *Fully, Largely, Partially,* or *Not achieved*.

3. Consolidation method

3.1 Inter-rater disagreement rationale

Commonly, team-based assessors participate in rating a software process assessment. The assessors make their own preliminary process ratings based on the interpretation of their assessment record. These are then discussed during the consolidation activity, and a consensus is made by the assessment team. The consensus is formed on the final ratings, as well as the evidence and findings for the achievements of PIs.

During a team-based assessment, assessors are exposed to the same evidence. This evidence can be the result of pre-onsite questionnaires, the responses to questions during an interview, or from the document inspection (SPICE trials Report, 1999). Whereas, the importance of each PI of a PA depends on an assessed process context, such as application domain, business purpose, development methodology, organization size, etc. Hence, each PI should have a different priority depending on its impacts on the PA (Jung, 2001). In some cases, even though the team assessors agree on the achievement of each PI, the

assessors would disagree on the final ratings because of differences in thinking of priorities for a set of PIs. This aspect is more critical in the boundary problems because different priorities lead to different ratings among assessors. It is logical to expect that the more disagreement exits among the assessors, the more effort will be spent on consolidation (El Emam et al., 1998).

Even though the AHP method is employed to determine the ratings, there still remains consolidating different pairwise judgments among the team assessors that lead to inter-rater disagreement on the priorities. A well-known method for such group decision-making in the AHP is to take the geometric mean of individual judgments because the reciprocal of the geometric mean of the judgments becomes the geometric mean of the reciprocals (Saaty, 1980). In this method, removing outliers and taking the geometric mean of the rest helps to avoid maleficence of the outliers. However, since not many assessors participate in rating a software process, removing outliers is not proper; there is not enough judgment data to waste. This method, even if possible, requires a tool to check outliers.

In this regard, we propose an alternative method to make iterative individual judgments toward a group consensus after feedback of the group outcome in each iteration stage. This method also requires a tool to measure how close an individual outcome is to the group outcome and a stopping point of iteration to insure that individual outcomes reach a consensus. The compatibility index of the AHP can satisfy those requirements.

3.2 Compatibility Metric

Compatibility in the AHP is concerned with two different vectors derived from two judgments matrices. In accordance with a group judgment of the AHP, Saaty (2001) developed the compatibility metric that measures how mutually close two pairwise reciprocal matrices of two ratio vectors are. The metric was analytically derived from the relation between a matrix of judgments and the matrix of corresponding eigenvector ratios. At the beginning on compatibility metric, let **Error! Objects cannot be created from editing field codes.** be the Hadamard product in this paper and its multiplication between the two given matrices W and V is defined as follows:

$$W \circ V = (w_{ij} \cdot v_{ij}) \text{, where } W = (w_{ij}) \text{ and } V = (v_{ij}) \tag{4}$$

Measuring compatibility between the two ratio scales $w^T = (w_1, w_2, \cdots, w_n)$ and $v^T = (v_1, v_2, \cdots, v_n)$ is defined below.

$$e^T W \circ V^T e \text{, where } e^T = (1,1,\cdots,1) \text{, } W = (w_i / w_j) \text{, and } V = (v_i / v_j) \tag{5}$$

If two matrices are exactly the same, then $e^T W \circ V^T e = n^2$.

Compatibility between two ratio vectors can be determined using compatibility index (S.I.) which is derived from the relation between compatibility and consistency. Consistency is concerned with the compatibility of a matrix of the ratios constructed from a principal right eigenvector with the matrix of judgments from which it is derived. Compatibility is concerned with two different vectors. Let W be the matrix of ratios of the principle right eigenvector of the positive reciprocal matrix A, and $\lambda_{\max}$ be the corresponding principal eigenvalue of A. Using the two matrices, W and A, the Compatibility Index (S.I.) is defined as follows;

$$S.I. = n^{-2} \cdot e^T A \circ W^T e \tag{6}$$

S.I. becomes 1 if and only if the two matrices are exactly the same (i.e., matrix A is perfectly consistent). Otherwise, *S.I.* goes beyond 1.

Since equation (6) equals $\lambda_{\max} / n$ (Saaty, 2001 p 67), the right hand side of (6) can be replaced by $1 + CI(n-1)/n$ using equation (2). From the acceptance level of *C.I.*, we can derive the significance level of *S.I.* to assure that two matrices of ratio vectors are compatible. Table 2 gives information on compatibility and consistency for different size judgment matrices.

Table 2 - Relationship between Consistency and Compatibility (Saaty, 2001 p 68)

Size (n)	*C.R.*	*R.I.*	*C.I.*	λ_{max}	*S.I.*
3	0.05	0.52	0.026	3.052	1.017
4	0.08	0.89	0.071	4.214	1.053
5	0.10	1.11	0.111	5.444	1.089
6	0.10	1.25	0.125	6.625	1.104
7	0.10	1.35	0.135	7.810	1.116
8	0.10	1.40	0.140	8.980	1.123
9	0.10	1.45	0.145	10.160	1.129

3.3 Consolidation Method

Our method can be triggered at the stage of consolidation after assuring inter-rater disagreement in rating a PA, in spite of agreement on the achievements of the associated PIs, among assessors. Assuredly, this kind of boundary problem is caused when individual assessors have different sets of PI priorities.

The key idea of our method is drawn from the Delphi method, which is one of the qualitative group decision making techniques. The Delphi method takes iterative group processes, which use central tendency to make individual decision makers gradually converge toward a position. Our method consolidates individually different judgment outcomes until all individual outcomes are compatible with the group outcome. The latter can be obtained by taking the geometric mean of the composite outcome of several individuals. In detail, after assessing the compatibility of the matrix of ratios of individual with that of the group, one can suggest to each individual which of his ratios is the most incompatible with that of the group and propose changes in his thinking to make it more compatible. Through such revision and recalculation of the group outcome, one may be able to obtain a group decision that is compatible with each member. The compatibility test can be used objectively to assess how a pairwise comparison matrix of an individual is close to that of group judgments.

We use derivatives of a given judgment matrix. Harker (1987) used the derivatives of a given incomplete pairwise comparison to determine an element that has the greatest influence on the weight vector. Following notations describe our method and present an example in the next section.

$P = (p_{ij})$: The matrix of ratios of individual assessor

$W = (w_{ij})$: The matrix of ratios of a group judgment

M : The compatibility matrix, $M = P \circ W^T = [p_{ij}(w_j / w_i)] = [p_{ij} \cdot w_{ji}]$

f(M) : The compatibility, $f(M) = e^T M e = e^T [p_{ij} w_{ji}] e$.

Following steps describe our method.

- **Step 0:** Create the matrices of ratios of each individual assessor and the group.
- **Step 1:** Compute compatibility (f) between each matrix of individual judgment ratios (P) and the matrix of group judgment ratios (W).
  If there is any P that gives compatibility beyond predetermined level, then go to the next step. Else terminate processes.
- **Step 2:** For such a P in the previous step, find the (i, j) element that gives the largest absolute

value of the coefficient of gradient of *f* with respect to p_{ij}.

- **Step 3:** Guide the decision maker of P to adjust the (i, j) element considering w_{ij}
- **Step 4:** Calculate new vector p' and, go back to Step 0.

Concerning step 2 above, the derivatives of M with respect to a matrix element, p_{ij} is an $n \times n$ matrix of the form

$$[\partial M / \partial_{ij}]_{kl} = \begin{cases} w_{ji} & if\ k = i, l = j \\ -w_{ij} \cdot p_{ij}^{-2} & if\ k = j, l = i \\ 0 & otherwise \end{cases}$$

Then derivatives of compatibility with respect to each p_{ij} is also an $n \times n$ matrix of the form,

$$[\partial f / \partial_{ij}] = [w_{ji} - w_{ij} \cdot p_{ij}^{-2}] \tag{7}$$

In this matrix, we can find the largest gradient element that can reduce the current compatibility value. The choice of the cell to adjust toward group judgment is *(i, j)* that gives the largest absolute value of the coefficient of gradient of *f* with respect to p_{ij}, i.e.,

$$(i, j) = \underset{k,l}{\operatorname{argmax}}(|\partial f / \partial_{kl}|) \tag{8}$$

After step 3, one can construct the judgment matrix P using new priority vector (p'), and compute compatibility between new P and W as follows:

$$n^{-2} \cdot e^T P \circ W^T e \tag{9}$$

According to whether or not the measured value is less than a predetermined value such as *S.I.* in Table 2, one can stop or continue the consolidation processes.

Concerning the computational complexity, the following equation lightens the computational workload.

$$\begin{aligned} e^T P \circ W^T e &= e^T \left[\frac{p_i}{p_j} \cdot \frac{w_j}{w_i} \right] e = e^T \left[\frac{p_i}{w_i} \cdot \frac{w_j}{p_j} \right] e \\ &= \sum_i \sum_j \left(\frac{p_i}{w_i} \cdot \frac{w_j}{p_j} \right) = \left(\sum_i \frac{p_i}{w_i} \right) \cdot \left(\sum_j \frac{w_j}{p_j} \right) \end{aligned} \tag{10}$$

4. An illustrative example

In the context of AHP, we will illustrate an example based on the rating of a PA with four PIs. The situation assumes that three assessors participate in the assessment. In the initial rating of Table 3, the first and the second assessors' ratings are *Largely* and the other's is *Partially*, which indicates inter-rater disagreement. The achievement scores for four PIs are agreed among assessors and given as 0.75, 0.60, 0.25, and 0.40 respectively. This kind of situation is regarded as a boundary problem.

Table 3 presents three pairwise comparison matrices and the geometric mean matrix of the three matrices. The weight vector for each matrix is obtained through the eigenvector method. The composite value for a

rating is calculated by the multiplication of the weight vector with the achievement score vector of four PIs as in equation (3). It is logical to expect that a rating by the holistic evaluation should be consistent with that by the decomposed evaluation utilizing the AHP.

Let $SI_{\ell G}$ be the compatibility between the matrix of assessor l and that of the group. According to equation (9), the compatibility indices between each individual matrix and the group matrix appear as SI_{1G} =1.023, SI_{2G} =1.112 and SI_{3G} =1.210. The third assessor shows the greatest incompatibility with the rest assessors in overall thinking of relative importance. The corresponding significance level of compatibility is 1.053 in Table 2. The different priorities among assessors cause different ratings.

At first iteration, according to equation (8), we can find the element that gives the largest absolute value of the coefficient of gradient to compatibility. For the judgment matrix of assessor 2, the element in the (4, 1) position is recommended to change. Assume that the new judgment is 1/4 from 1/5, considering the value of group mean (1/2.71). For the case of assessor 3, assume the current value of (1, 3) changes to 1 (See Table 4). Calculate the new priorities vectors for both assessor 2 and assessor 3, and go back to step 1.

Since SI_{3G} (= 1.119) > 1.053 in Table 4, the second iteration is activated only for assessor 3. According to equations (7) and (8), the (1, 4) position in the matrix for assessor 3 is to be revised. Let it be 2 from 1 in Table 5. Then, since all the compatibility indices between each matrix and the group matrix satisfy the stopping condition (SI_{1G} =1.015; SI_{2G} =1.040; SI_{3G} =1.050 $\leq$ 1.053), we can determine the final rating. According to the final step of our method, the composite value becomes 54% and thus, rate the illustrated process attribute as *Largely*.

Table 3 - Initial rating information

	Assessor 1	Assessor 2	Assessor 3	Geometric Mean
Matrix	1.00 2.00 2.00 4.00 1.00 1.00 2.00 1.00 2.00 1.00	1.00 1.50 3.00 5.00 1.00 2.00 3.33 1.00 1.67 1.00	1.00 1.00 0.50 1.00 1.00 0.50 1.00 1.00 2.00 1.00	1.00 1.44 1.44 2.71 1.00 1.00 1.88 1.00 1.88 1.00
Weights of PI	0.44 0.22 0.22 0.11	0.46 0.30 0.15 0.09	0.20 0.20 0.40 0.20	0.36 0.25 0.25 0.13
Measures of PI	PI 1 = 0.70, PI 2 = 0.60, PI 3 = 0.25, PI 4 = 0.40			
Composite Value (Rating)	54% (Largely)	57% (Largely)	44% (Partially)	
SI_{lG}	SI_{1G} = 1.023, SI_{2G} = 1.112, SI_{3G} = 1.210			

Table 4 - Revised rating information (1st iteration)

	Assessor 1	Assessor 2	Assessor 3	Geometric Mean
Matrix		1.00 1.50 3.00 4.00 1.00 2.00 2.67 1.00 1.33 1.00	1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00	1.00 1.44 1.82 2.52 1.00 1.26 1.75 1.00 1.39 1.00
Weights of PI	0.44 0.22 0.22 0.11	0.44 0.30 0.15 0.11	0.25 0.25 0.25 0.25	0.38 0.26 0.21 0.15
Measures of PI	PI 1 = 0.70, PI 2 = 0.60, PI 3 = 0.25, PI 4 = 0.40			
Composite Value (Rating)	54% (Largely)	57% (Largely)	48% (Partially)	
SI_{lG}	SI_{1G} = 1.034, SI_{2G} = 1.053, SI_{3G} = 1.119			

Table 5 - Revised rating information (2nd iteration)

	Assessor 1	Assessor 2	Assessor 3	Mean
Matrix			1.00 1.00 1.00 2.00 1.00 1.00 2.00 1.00 2.00 1.00	1.00 1.44 1.82 3.17 1.00 1.26 2.20 1.00 1.75 1.00
Weights of PI	0.44 0.22 0.22 0.11	0.44 0.30 0.15 0.11	0.29 0.29 0.29 0.14	0.39 0.27 0.22 0.12
Measures of PI	PI 1 = 0.70, PI 2 = 0.60, PI 3 = 0.25, PI 4 = 0.40			
Composite Value (Rating)	54% (Largely)	57% (Largely)	50% (Largely)	54% (Largely)
SI_{lG}	SI_{1G} = 1.015, SI_{2G} = 1.040, SI_{3G}= 1.050			

5. Concluding Remarks

In this study, we proposed a consolidating method that could systematically draw a group consensus on a set of priorities of practice indicators in software process assessments, based on ISO 15504 utilizing the AHP. Our method uses an iterative mode of judgments, which gradually narrows differences among assessors' evaluation, in order to reach the point where the differences are so trivial that individual judgments can be compatible. The compatibility metric of the AHP is an indicator to represent the degree of compatibility between individual judgments with a group judgment. This study also illustrated the application steps of our method with an appropriate example in a software process assessment.

The proposed method can be activated when inter-rater disagreement arises through different pairwise comparison matrices for occasions such as boundary problems that require careful and elaborate assessments. This method would be used to draw an agreement with the Organization Unit in presenting assessment results, as well as to reach a consensus among assessors. This would ultimately contribute to quality assessments.

The proposed method is obviously a time-consuming technique. However, considering the trade-off between the evaluation accuracy and quality, and the time required to apply this method, we propose that this method is worthy of operating for such occasions as boundary problems.

References

AHP special issue (1990) - Decision Making by the Analytic Hierarchy Process: Theory and Applications, *European Journal of Operational Research*, 48/1.

CMMI (2002), *Capability Maturity Model Integration*, SE/SW/IPPD/SS, V 1.1, CMU/SEI-2002-TR-011, Software Engineering Institute, Carnegie Mellon University.

El Emam K., Simon J. M., Rousseau S., Jacquet E. (1998), “Cost Implications of Interrater Agreement for Software Process Assessment,” *Proceedings of the 5th. International Symposium on Software Metrics*, 38-51.

El Emam K., Jung H.-W. (2001), "An Evaluation of the ISO/IEC 15504 Assessment Model", *Journal of Systems and Software*. 59/1, 23-41.

Harker, P. T. (1987), “Incomplete Pairwise Comparisons in the Analytic Hierarchy Process,” *Mathematical Modelling*, 9/11 837-848.

ISO/IEC 12207 AMD (2004), *Information Technology - Software Life Cycle Processes*, ISO.

ISO/IEC TR 2 15504-2 (1998); 5 (1999), *Information Technology - Software Process Assessment*: Part 2; Part 5, ISO.

ISO/IEC 15504-2 (2003); 5 (2004), *Information Technology - Process Assessment*: Part 2; Part 5 (FCD), ISO.

Jung H.-W. (2001), "Rating the Process Attributes Utilizing AHP in SPICE-based Process Assessments", *Software Process Improvement and Practices: International Journal* 6/2, 111-122.

Jung, H. W. and Choi, B. J.(1999), "Optimizing Models for Quality and Cost of Modular Software Systems, *European Journal of Operational Research,* 112/3, 138-144.

Karlsson J. and Ryan, K.(1997), "A Cost-Value Approach for Prioritizing Requirements," *IEEE Software*, Sep./Oct., 67-74.

Kontio, J., Caldiera, G., and Basili, V. (1996), "Defining Factors, Goals, and Criteria for Reusable Component Evaluation", *CASCON '96 Conference*, Toronto, Canada.

Maiden, N., Ncube, C., and Moore, A. (1997), “Lessons learned during requirements acquisition for COTS systems,” *Communications of the ACM*, 40/12, 21-25.

Saaty, T. L. (2001*), The Analytic Network Process (2nd Edi.)*, RWS Publications, Pittsburgh.

Saaty, T. L. (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York.

SPICE Trials Report (1999), *SPICE Phase 2 Trials Final Report*, Vol. 1, ISO/IEC JTC1/SC7/WG10.

Yoon, M. S. (1997), *S/W Quality Evaluation Model Usig the AHP - Developing a New Judgments Aggregation Method*, Ph. D. Dissertation, Korea University.

Zhedi, F., and Ashrafi, N. (1991), "Software Reliability Allocation Based on Structure, Utility, Price, and Cost," *IEEE Transactions on Software Engineering*, 17/4, 345-356.