DO KNOWLEDGE EXPERIENCE PACKAGES IMPROVE KNOWLEDGE TRANSFER? RESULTS FROM A CONTROLLED EXPERIMENT

Pasquale Ardimento*, Vito Nicola Convertini*, Marcela Genero Bocco**, Giuseppe Visaggio++

* Dipartimento di Informatica – Università di Bari - Via Orabona, 4, 70126 Bari – Italy
email: pasquale.ardimento@uniba.it, convertini@di.uniba.it, giuseppe.visaggio@uniba.it

** Instituto de Tecnologías y Sistemas de Información - University of Castilla-La Mancha - Ciudad Real, Spain
email: Marcela.Genero@uclm.es

+ Daisy-Net - Driving Advances of ICT in South Italy – Net, Bari, Italy

ABSTRACT
The transfer of research results into production systems requires, among other things, to make these results explicit and understandable for stakeholders. Many researchers have been studying means to favour knowledge acquisition on the behalf of stakeholders using not only scientific papers. In fact, scientific papers have limitations in the transferability and reusability of knowledge owing to: size limitations that do not permit all the necessary knowledge to be exploited; structure and contents heterogeneity that could cause conflicts; and redundancy and incoherence between papers dealing with same topic. We therefore propose the concept of the Knowledge Experience Package (KEP) that contains both the conceptual “model(s) of the research results which make up the innovation, including all the necessary documentation ranging from papers or book chapters, and the” [42] appropriately structured experience acquired from the business processes. The KEP conceptual model has been implemented in a Knowledge Experience Base (KEB) called PROMETHEUS. The principal goal of this work is to assess whether the knowledge transferred by a KEP, and delivered by PROMETHEUS, adds comprehensibility to the knowledge transferred by scientific papers. A controlled experiment was therefore carried out with 14 IT professionals contracted in projects led by the Alarcos research group at the University of Castilla-La Mancha (Spain), all of whom have degrees in Computer Science. The comprehension of the knowledge transferred was analyzed from three different perspectives: semantic comprehension, retention and transfer. The evidence collected permits us to state that the structure of the KEP increases the ability to retain knowledge in a significant manner than occurs with scientific papers only. The results have encouraged us to carry out further studies in order to refine the KEP structure, although we are conscious that the results should be considered as preliminary.

KEY WORDS
Knowledge Experience Base; Knowledge Transfer, Controlled Experiment, Comprehension.

1. Introduction
“The ever greater pressure of competition to which firms are subjected has made product and process innovation a crucial issue” [42]. If the production and the delivery of technological innovations are to be increased, then the dynamic integration of the many competences and different knowledge produced and provided by different organisations is necessary [1, 2]. “Successful innovators must complement in-house knowledge with technologies from external sources” [42]. R&D is consequently shifting from its traditional inward focus towards a “more outward-looking management that draws on knowledge from networks comprised of universities, start-ups, suppliers and even competitors” [42].

Knowledge is a critical production factor in IT because it is human centred and is “used in order to enforce capabilities in each application domain” [42]. The knowledge needed is therefore both of a technical and a social nature. The first type includes the knowledge of methods, techniques and processes with which to build and maintain software products, in other words: the “knowledge of the technologies that apply to software development” [42]. The second type includes knowledge concerning the developer’s behaviour and stakeholder needs. Knowledge involves, at the very least, the problem of transferability and consequent reusability”. Much of the knowledge used in IT development processes is tacit, and much is hidden in processes and in products [3]. Indeed, it is known that tacit knowledge concepts consequently emerge in the face of events that are not necessarily planned. Knowledge that is hidden in processes and products is often not even readable by its authors since it is spread out and confused in many of the process or product components [3, 4]. “So, until knowledge is transferable or reusable, it cannot be considered as part of an organisation’s assets” [42].

Moreover, scientific publications as a means to transfer knowledge have common problems in appraising the quality of published research because journal articles and, in particular, conference papers, rarely provide sufficient detail on the methods used owing to the space limitations in journal volumes and conference proceedings [45].

This problem signifies that knowledge must be formalised in order for it to be comprehensible to and reusable by
others who are not the author of that knowledge. We have already proposed an approach, called PROMETHEUS (Practices Process and Methods Evolution Through Experience Unfolded Systematically) with which to formalize the knowledge that is to be transferred. The PROMETHEUS approach [5, 6, 7, 8, 9] is a Knowledge Experience Base (KEB) in the form of the Knowledge Experience Package (KEP). The KEP is the vehicle suggested for the transfer of experience obtained by analyzing and synthesizing all kinds of experience. It acts as a repository for this experience and then supplies it to various projects on demand.

The principal goal of this paper is to present a controlled experiment to assess whether the knowledge transferred by a KEP, and delivered by PROMETHEUS, make more comprehensible than the knowledge transferred by scientific papers. The controlled experiment was carried out with 14 IT professionals contracted in projects led by the Alarcos research group at the University of Castilla-La Mancha (Spain), all of whom have degrees in Computer Science. The remainder of the paper is structured as follows. Section 2 discusses related works and research activities dealing with knowledge packaging; section 3 presents PROMETHEUS, focusing particularly on the structure of the KEP. Section 4 details the experiment, while in section 5 the analysis of the data obtained in the experiment is presented. Finally, in section 6 our conclusions and future works are presented.

2. Related Works
The problem related to knowledge transfer is currently being investigated in both industrial and academic contexts. Some companies have established internal organisations whose task is to acquire new knowledge [10, 11, 12] in order to confront knowledge transfer needs. For example, Shell Chemicals has organised some groups with the scope to discover knowledge from outside sources [10]. Hewlett Packard is commercialising not only its own ideas, but also innovations from other entities [11], and Philips Research is participating in consortiums that direct one to one collaborations with innovative organisations [12]. Another approach to knowledge search and transfer is based on the use of ontologies [13, 14]. “This approach is in fact the object of many studies which currently lack tools for creation and management. Much attention is being focused on these issues, but the experimental evidence that is available is not yet sufficient for large-scale use” [42]. In this work, we propose an alternative approach to knowledge transfer based on the concepts of knowledge packaging and knowledge bases. “The problem of knowledge packaging for better use is being studied by many research centres and companies. The current knowledge bases in literature sometimes have a semantically limited scope” [42]. This is the case of the IESE base [15], which collects lessons learned, mathematical prediction models or the results of controlled experiments. In other cases, the scope is wider but the knowledge is too general and is not therefore very usable. This is the case of the MIT knowledge base [16], which “describes business processes but only at one or two levels of abstraction. Other knowledge bases that cover wider fields with greater operational detail” [9] probably exist, but we do not know much about them because they are private: for example, the Daimler-Benz Base [17, 18]. There is another recent KEB, known as DAU [19] which is not structured like our KEB but has many key performance indicators that are similar to ours. The literature analysis and study of QIP and Experience Factory issues are strictly linked to the practices and application, as a matter of the fact QIP and EF approach were born with the collaboration of research and industrial communities [26, 20]. Many EF and KEBs are developed in industrial contexts. In [26] there is a list of the most common and used KEBs such as CeBase National Science Foundation (NSF), VSEK German, ESERNET Fraunhofer IESE and many other ones. The analysis of those EBs shows that even if the EB concept is spread in the last years, there is a low number of available KEBs corresponding to KEB theorized characteristics. Moreover, the available KEBs contents, goals and stakeholders are very different and these differences probably causes a higher difficult of the stakeholders in the use and consultation and the knowledge sharing. In fact, the KEB structures are not homogeneous and sometimes are quite different from the theoretical models. Moreover, the KEP definitions are, in own opinion very hard to apply in practical context and we are still missing rigorous definition of the necessary KEP characteristics and structure.

3. Prometheus
The authors use the term KEP to refer to an organized set of: knowledge content, teaching units on the use of the demonstration prototypes or tools and all other information that may strengthen the package’s ability to achieve the proposed goal. The KEP must be usable independently of its author or authors, and the content must therefore have a particular structure: distance education and training must be available through an e-learning system. In short, the proposed knowledge package contains knowledge content integrated with an e-learning function. In PROMETHEUS, the KEP must include all the components shown in Fig. 1. “A user can access one of the package components and then navigate along all the components of the same package according to her/his needs” [9].
the user's knowledge, and can be found in conventional sources of knowledge such as technical reports, papers and books. When users need some of the basic concepts for understanding the contents of KEP they can use an educational e-learning course, and if users should need more information, they can use the "attachments" regarding reports, papers and books concerning the basic topics of KEP. As stated above, the use of these courses is flexible, to meet individual user's needs. "When a package also has support tools, rather than merely demonstration prototypes, Knowledge Content (KC) links the user to the available tool. For the sake of clarity, we should point out that this is the case when the knowledge package has become an industrial practice, signifying that the demonstration prototypes included in the archetype they are derived from have become industrial tools. “The tools are collected in the Tools Component (TL))” [42]. Each of the tools that are available is associated with an educational course, again of a flexible nature, in the use of the correlated e-learning training course. “A KEP is generally based on conjectures, hypotheses and principles. As they mature, their contents must all become principle-based. The transformation of a statement from conjecture through hypothesis to principle must be based on experimentation showing evidence of validity. The experimentation, details of its execution and relative results, are collected in the Evidence component (EV)” [9], which is duly indicated by the KEP. Should the user need support from those who have knowledge of the contents of KEP, a list of resources is provided as a reference. The list is collected in the Competence component (CM).

The following sections describe KEP components, and further details are available in the following publication [20].

3.1 Knowledge Content
As can be seen in Fig. 1, the central KC is that of Art & Practices (AP). “It contains the KEP expressed in a hypermedia form in order to include figures, graphs, formulas and whatever else may help the user to understand the content. The KC is organized as a tree that starts from the root (level 0) and descends to the lower levels (level1, level2, ..., leveln) through pointers [9]. The higher the level of a node, the lower the abstraction of the content, which focuses” [9] to an ever increasing extent on operative elements. The root node of the KC is made up of the following sections: a thoughtful index that tells the reader how the package suggested will change in practice, with a list of processes and activities, the whole or a part of the process of an organization; a problem (one or more) that describes the problem faced by the KEP. A problem may belong to one of the two following types: decision and optimization. If the problem is the decision then the user should be able to make a choice. If the problem is optimization, then the resources that the user needs to improve the performance and the objective function of optimization must be indicated. The context must be defined for each problem, which is to say all the facts and circumstances that cause and condition a certain problem. The leaf nodes have the answers to the problems: the solution or solutions suggested for each problem set.

To ensure control of completeness and lack of ambiguity in the contents of KEP, the vocabulary of KEP, i.e. concepts and relations between there meanings, has been formalized through the use of the SCORM (Sharable Content Object Reference Model). SCORM “is a collection of standards and specifications for web-based e-learning. It defines communications between client side content and a host system called the run-time environment, which is commonly supported by a learning management system. SCORM also defines how content can be packaged into a transferable ZIP file called a "Package Interchange Format" “ [44]. The use of SCORM has the following advantages:

1. The full list of concepts (elements) which have to be declared with obligatorily, multiplicity and default values of the elements / concepts, relationships between elements / concepts, type of elements, attributes defined for each element, type of attributes;
2. Elimination of ambiguity, incompleteness, verbosity owing to Informal definitions;
3. Verification of the correct syntax;
4. Interoperability of the KEP, at the syntactic level, between background of experiences that share the SCORM format, leading to an independence of the software that produces them.

The research results integrated by a KEP may be contained within the same knowledge base or be derived from other knowledge bases or other laboratories. If the knowledge package being read uses knowledge packages located in the same experience base, the relations will be explicitly highlighted.

3.2 Attributes
A search within the package starting from any of its components is facilitated by the component’s Attributes. As shown in Fig. 1, each component in the knowledge package has its own attribute structure. In the case of all the components, the attributes allow a rapid selection of the relative elements in the knowledge base. The Attributes have already been defined in [3, 4]. To facilitate the research, we used a set of selection classifiers and a set of descriptors which summarize the contents. The summary descriptors include: a brief summary of the content and a history of the essential events occurring during the life cycle of the package, giving the reader an idea of how they have been applied and improved, and how mature they are. “The history may also include information telling the reader that the content of all or some parts of the package are currently undergoing improvements” [9]. The classifiers include:

- The adoption risks of the technological innovation for which it is a provider;
- The risk mitigation initiatives that ensure a better performance of the KEP in the solution of the principal problems
The impact that the KEP will have on the active processes of the production lines to which it will be applied, supposing that the problems to be solved correspond with those in the KEP.

A forecast of the Return of Investment that the new introduction will have in the company. Both the economical impact of the KEP and its impact on the value chain are therefore specified.

The acquisition plan of the KEP methods;

The history of the KEP, i.e. the set of practices that have required its use and the results obtained from their application in order to ensure a higher perception of the KEP’s reuse.

3.3 Project Component
This component contains all the information that may be “useful for project characterization: project description, project contest description” [20]; project resources used; project results; events which have occurred during the project execution. It also includes “the metrics model used for monitoring, and therefore the values collected by the measures. Of particular interest are the economical indexes: expected cost of the planned work; actual costs of the work carried out, predicted cost of the work carried out” [20].

3.4 Tool Component
This component describes the problems that the tool is able “to overcome and answer, and which problems it is appropriate for. The competences needed should be outlined, along with the adoption risks and mitigation initiatives. To speed up adoption, stakeholders should be involved” [20]. Finally, the tool maturity should be indicated by explaining the number of versions released, the motivations that have led the version to be changed, and the changes carried out. Training courses should also be planned. The attributes are: “short description, technical domain area covered by the tool, keywords, date of first release, adoption requirements; competences and adoption risks, plan for adoption, classification values for type of stakeholder, number of versions and history” [20].

3.5 Evidence Component
The evidence of a research study cannot be interpreted with any confidence unless it has been considered together with the results of other studies addressing the same or similar questions [38]. The comparison and contrasting of evidence is necessary to build knowledge and reach conclusions about the empirical support for a phenomenon. “The EV contains the description of all the empirical investigations that validate the cause-effect relation between the research results the innovations have proposed and the answers to the problems. In particular, this component describes the data used, the controls carried out on them to ensure their accuracy and the mechanisms used to carry them out, the experimental design, the statistical analysis and the results obtained” [20]. The contents of the EV are closely related to the Project and AP Components. The main attributes are: Brief Evidence Description, brief abstract synthesizing the Evidence Content; Evidence Kind, Empirical evidence, Industrial context.

3.6 Competence Component
This component “collects people or organizations with competences in the first case, and capabilities in the second” [20], which “support the adoption of a KEP or of a Tool. Each subject referring to this component is characterized by a description of the combination of knowledge, skills and behaviours that are part of the professional assets” [20]. The attributes for this component are made up of the list of professional profiles according to the standards used.

4. Experiment Description
In this section the main details of the experiment will be detailed. The experiment material is available at [35]. The main goal of this experiment is to assess whether the knowledge transferred by a KEP, and delivered by PROMETHEUS, is more comprehensible than the knowledge transferred by scientific papers. We used the GQM template [21, 22] to define the goals of our experiment. This is defined as follows:

“Analyze the comprehension of KEPs for the purpose of comparing it with the comprehension of papers with regard to the effectiveness and efficiency from the viewpoint of IT researchers in the context of professional IT stakeholders from the University of Castilla-La Mancha”.

This experiment was run and reported by following the recommendations provided in [23, 24].

4.1 Planning
Various issues related to the planning of the experiment are introduced in this section.

Subjects Selection. The experiment was carried out by 14 IT professionals contracted in projects led by the Alarcos research group at the University of Castilla-La Mancha (Spain), all of whom have degrees in computer science. The KEP selected, which was used during the experimentation, referred to basic software engineering knowledge such as, software maintenance, the software development lifecycle and so on. The topic was therefore comprehensible to each single experimental subject. To avoid social threats caused by evaluation apprehension, the subjects were not graded on their performance.

Sample assignment. To compare the comprehension of a KEP with the comprehension of scientific papers, it was necessary to select more than one experimental group: one that would use the KEP and another that would use the scientific papers. When there is more than one group in an empirical experimentation it is important that the groups are balanced as regards a set of specific characteristics whose values could affect the empirical results. If the groups are not equal, then it is not possible to know whether the independent variable has caused changes in the subjects or whether this has occurred as a result of the inequality. The most appropriate method
when a sample has to be, or is, small is that of matched sample assignment. This method ensures the distribution of control variables by matching pairs of subjects in the different groups, i.e., it is necessary to find pairs of subjects who are very similar to each other as regards control variables, such as age, sex, race, etc., and it is then possible to randomly assign one of the pair to one group and the other to the second and so on. In our investigation, we identified two control variables: the level of instruction, and knowledge of the English language. The matched sample was applied by asking the subjects to fill in a background questionnaire. The results obtained were used to select two groups whose standard deviation was the lowest with regard to the two control variables. **Independent and dependent variables.** In accordance with certain suggestions concerning the measurement of comprehension [25, 26, 27], we have used the Cognitive Theory of Multimedia Learning (CTML) [46]. This choice was made for several reasons. First, it focuses on words and graphics, which are the elements used in papers and in KEPs. Second, it provides principles for the design of effective multimedia presentations which can be tested empirically. Third, it has evolved through years of work and development of experimental instruments and methods related to model comprehension [28, 29]. By following CTML, the comprehension has been defined through three dependent variables:

- **Semantic Comprehension:** the ability to comprehend the semantic material being presented.

- **Retention:** the comprehension of the material being presented, and the ability to retain knowledge from it.

- **Transfer:** the ability to use the knowledge gained from the material to solve related problems which are not directly answerable from it.

The quality model, defined according to the GQM approach [30, 31] is presented in Fig. 2 which also indicates the experimental hypothesis defined for each measure used.

<table>
<thead>
<tr>
<th>GOAL</th>
<th>QUESTION</th>
<th>METRICS</th>
<th>EXPERIMENTAL HYPOTHESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>The semantic comprehension of a KEP is more effective/efficient than a paper?</td>
<td>SCEffec</td>
<td>H1.1,0,1,2</td>
<td></td>
</tr>
<tr>
<td>The retention of a KEP in a more effective/efficient than a paper?</td>
<td>RetenEffec</td>
<td>H2.1,2</td>
<td></td>
</tr>
<tr>
<td>The transfer of a KEP in more effective/efficient than a paper?</td>
<td>TransEffec</td>
<td>H3.1,2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Quality Model used and Experimental Hypotheses

Each dependent variable was measured in terms of:

- **Effectiveness:** the proportion of correct answers provided in each test (number of correct answers / number of questions). This measure reflects the ability to understand the material presented correctly. It is important to note that in our analysis, the unanswered questions are considered to be wrong answers.

- **Efficiency:** the proportion of correct answers divided by time (Effectiveness / Time).

In this work, we therefore defined the following measures: SCEffec/SCEffic (Semantic Comprehension Effectiveness/ Semantic Comprehension Efficiency), TransEffec/TransEffic (Transfer Effectiveness/ Transfer Efficiency), and finally, RetenEffec/RetenEffic (Retention Effectiveness/ Retention Efficiency).

**Design of experiment.** The design used was a “Between-subject design” in which each group is subjected to one treatment in one run. The knowledge transferred refers to the topic of Reiterative reengineering. All the subjects were assigned to 2 groups (A and B). The experiment consisted of one round. Two different diagrams were presented to every subject in each group. The design is shown in Table I.

<table>
<thead>
<tr>
<th>Table 1. BETWEEN SUBJECTS DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
</tr>
<tr>
<td>Iterative reengineering</td>
</tr>
</tbody>
</table>

**Experiment objects.** There were two experimental objects, related to the two values of the independent variable (KEP, papers). One consisted of a set of four papers [42] on the “Iterative Reengineering” topic, whilst the other was a KEP [40] made up of one AP component related to one EV component. Both the AP and the EV components were built by analyzing the 4 previously mentioned papers. All the papers were written by the same authors to avoid certain considerations, information and results conflicting with each other in the papers and not in the KEP. It is important to note that we selected the topic so that no experimental subject had any specific knowledge of it.

**Hypotheses.** For Semantic Comprehension we wished to test the following hypothesis:

- Null hypotheses
  - H1.1: KEPs do not improve SCEffec with regard to those subjects using papers.
  - H1.2: KEPs do not improve SCEffic with regard to those subject using papers.

- Alternative hypotheses:
  - H1.1 = ¬H1.0
  - H1.2 = ¬H1.2

We analogously formulated a set of hypothesis H2 for Retention measures (RetenEffec and RetenEffic), and another set, H3, related to the Transfer measures (TransEffec and TransEffic) (see Fig. 2).

**Instrumentation.** In the experiment we used three questionnaires [35] for both the treatments, one for each dependent variable. For each questionnaire the
subjects were asked to write down the starting time and the finishing time. The questionnaires are described as follows:

- Questionnaire 1: made up of 5 closed questions concerning the semantic comprehension of the iterative reengineering. Each question had 4 answers with one and only one correct answer. Questionnaire 1 was used to collect the SCEffic and SCEffic measures.
- Questionnaire 2: made up of 5 open questions. Questionnaire 2 was used to collect the TransEffec and TransEffic measures.
- Questionnaire 3: made up of 5 "fill-in-the-blanks" sentences. Each sentence had a blank space to be filled in with one or more words that were semantically consistent with their content. Questionnaire 3 was used to collect the RetenEffic and RetenEffic measures.
- Post-experiment survey: made up of 6 assertions using a 5-point Likert scale [34] to capture the intensity of the subjects’ feelings as regards specific items. Finally, a blank note field was added to permit each subject to provide feedback in terms of suggestions, observations and criticisms.

4.2 Post-Experiment Survey

The goal of this survey (see Table 2) was to obtain feedback about the subjects’ perceptions of the execution of the experiment. The 5-point Likert item ranges from strongly agree (1) to strongly disagree (5), and the range thus captures the intensity of their feelings for a given item [34].

The items concerned the perceived level of difficulty of: each single questionnaire, whose corresponding items are A1, A2 and A3, as shown in Table 2; the adequateness of the training received, A4; the clearness of the material provided during the experiment execution, A5; and the clearness of all the questions provided during the experimental execution, A6.

<table>
<thead>
<tr>
<th>Id</th>
<th>Assertion</th>
<th>values</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>During the experimental run question. 1 was easy</td>
<td>(1-5)</td>
</tr>
<tr>
<td>A2</td>
<td>During the experimental run question. 2 was easy</td>
<td>(1-5)</td>
</tr>
<tr>
<td>A3</td>
<td>During the experimental run question. 3 was easy</td>
<td>(1-5)</td>
</tr>
<tr>
<td>A4</td>
<td>The training received in the two-hour lecture which took place the day before the experiment was sufficient to execute the experimental run</td>
<td>(1-5)</td>
</tr>
<tr>
<td>A5</td>
<td>During the experimental run, all the materials provided were clear</td>
<td>(1-5)</td>
</tr>
<tr>
<td>A6</td>
<td>During the experimental run, all the questions were clear</td>
<td>(1-5)</td>
</tr>
<tr>
<td>Five-level Likert scale: 1 = strongly agree; 2 = agree; 3 neutral; 4 = disagree; 5 = strongly disagree</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 Experiment Execution

Before the experiment execution, we first carried out a pilot study and then a training session with the experimental subjects involved in the experiment execution.

The pilot study was executed in a three hour session in December 2011, two weeks before the experiment execution. It involved 29 first-year Master’s Degree students at the University of Bari in Italy, in order to obtain an initial insight into the experimental design. The results of the pilot study revealed that too much time was spent answering all the questionnaires. We observed that the average time spent answering all the questionnaires was about two and a quarter hours. We therefore decided to reduce the number of questions, observing the average amount of time spent on answering each questionnaire. The questions were removed using the following unordered criteria: time spent answering the question, feedback provided by students about the level of clearness perceived per question, number of answers not given per question.

The two hour training session took place the day before the experiment was carried out. This session had several goals:

- to introduce the concept of KEP and the use of PROMETHEUS to the experimental subjects;
- to show the experimental subjects an example similar to the material that would be used during the experiment. The topic used in the example was “Balance scorecard”, a topic not previously known by each subject;
- to explain how to solve the experimental tasks; to assign each experimental subject his own credential to access the PROMETHEUS platform [36];
- to ask the subjects to fill in a background questionnaire.

In the training session, the subjects do not know to which group they would be assigned, and did not therefore know if they would use PROMETHEUS or Papers. Their level of attention was therefore maintained.

The day after the training session, the experiment took place in a two-hour session. In the first 30 minutes we explained how to perform the experiment and assigned the subjects to each group using the matched sample assignment based on the data collected from the background questionnaire. The experiment was conducted in a laboratory, where the experimental subjects were supervised, and no communication among them was allowed. Both groups were located in the same room.

The experiment was executed as follows:

1. The subjects in group A downloaded the four selected papers. After logging onto PROMETHEUS, the subjects in group B selected the KEP concerning the Iterative reengineering topic. Immediately after all the subjects had finished these operations, the supervisor gave each one of them questionnaire 1.
2. After a subject stated that s/he had finished questionnaire 1, s/he received questionnaire 2.
3. After a subject stated that s/he had finished questionnaire 2, s/he received questionnaire 3. Only in this experimental case were the subjects unable to use the KEP or Papers, previously received, to answer the questions.
4. After a subject stated that s/he had finished questionnaire 3, s/he received the post-experiment survey.
5. Each subject immediately handed in each questionnaire to the supervisor upon its completion.

5. Data Analysis and Interpretation

In this section, we present, for each dependent variable, the descriptive statistics, the results [35] of hypothesis testing and the analysis of the data collected from the post-experiment survey.

5.1 Descriptive Statistics

Table 3 shows the descriptive statistics of all the Comprehension measures. The bold cells indicate the best group’s performance. In all cases, the mean values obtained are better for KEP, except in the case of the Semantic Comprehension measure in which the values are very close to each other.

<table>
<thead>
<tr>
<th>Measures of</th>
<th>Papers</th>
<th>KEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCEffec</td>
<td>0.8857</td>
<td>0.8571</td>
</tr>
<tr>
<td>SCEffic</td>
<td>0.0369</td>
<td>0.0714</td>
</tr>
<tr>
<td>TranEffec</td>
<td>0.5143</td>
<td>0.5714</td>
</tr>
<tr>
<td>TranEffic</td>
<td>0.0024</td>
<td>0.0026</td>
</tr>
<tr>
<td>RetenEffec</td>
<td>0.2286</td>
<td>0.4857</td>
</tr>
<tr>
<td>RetenEffic</td>
<td>0.0178</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

In all cases, except for Semantic Comprehension Effectiveness, the mean values are greater for the KEP group, although in the case of Semantic Comprehension Efficiency and Retention Efficiency the values are not very close.

5.2 Hypothesis Testing

The Mann-Whitney U Test was used to test the formulated hypotheses because all its assumptions were satisfied [33]:
- each dependent variable measured at least on an ordinal measurement scale;
- the independent variable had only two levels;
- a between-subjects design was used;
- the subjects were not matched across conditions.

The Mann-Whitney U Test [41] involves the calculation of a statistic, usually called U, whose distribution under the null hypothesis is known. In the case of small samples, the distribution is tabulated, but for sample sizes above ~20 there is a good approximation using the normal distribution. There are two methods with which to calculate the U value, the first of which involves:
1. arranging all the observations into a single ranked series that correspond in order to rank all the observations without regard to which sample they are in;
2. choosing the sample for which the ranks seem to be smaller (The only reason for doing this is to make computation easier). This is called "sample 1," and the other is called "sample 2";
3. taking each observation in sample 1, counting the number of observations in sample 2 that have a smaller rank (counting a half for any that are equal to it).

The sum of these counts is U.

The alternative hypothesis, however, states that one population (Y) equals the probability of an observation from the second population (X) exceeding an observation from X, that is, there is a symmetry between the populations with regard to the probability of the random drawing of a larger observation. The alternative hypothesis, however, states that one distribution is stochastically greater than the other. Table 3 summarizes the results of the Mann-Whitney analysis in order to establish whether a statistically significant difference exists between the group that used KEP and the group that used the scientific papers. The values of the rank sum, shown in the second and in the third columns, were used to calculate the U value, which was then used to calculate the p-value. Finally, the response concerning the significance of the statistical difference is shown in the last column.

Table 4. MANN-WHITNEY U TEST FOR ALL DEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>Measures of dependent variables</th>
<th>Rank Sum KEP</th>
<th>Rank Sum papers</th>
<th>U</th>
<th>p-value</th>
<th>Statistical difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCEffec</td>
<td>31.50</td>
<td>23.50</td>
<td>8.5</td>
<td>0.4033</td>
<td>NO</td>
</tr>
<tr>
<td>SCEffic</td>
<td>37.00</td>
<td>18.00</td>
<td>3</td>
<td>0.0472</td>
<td>YES</td>
</tr>
<tr>
<td>TranEffec</td>
<td>29.50</td>
<td>25.50</td>
<td>10.5</td>
<td>0.6761</td>
<td>NO</td>
</tr>
<tr>
<td>TranEffic</td>
<td>32.00</td>
<td>23.00</td>
<td>8</td>
<td>0.3472</td>
<td>NO</td>
</tr>
<tr>
<td>RetenEffec</td>
<td>38.50</td>
<td>16.50</td>
<td>1.5</td>
<td>0.0215</td>
<td>YES</td>
</tr>
<tr>
<td>RetenEffic</td>
<td>39.00</td>
<td>16.00</td>
<td>1</td>
<td>0.0162</td>
<td>YES</td>
</tr>
</tbody>
</table>
A statistically significant difference exists between the Semantic Comprehension Efficiency and Retention Efficiency variables. The efficiency, which represents the proportion of correct answers divided by time, of Semantic Comprehension and Retention shows a statistically significant difference between the KEP and the scientific papers. This result is probably owing to the fact that the scientific papers contain more and redundant information, in addition to the fact that they are not as structured as KEPs. Nevertheless, Transfer Efficiency, measured by the questionnaire made up of open questions, has no statistical difference. This measure was collected using a questionnaire made up of open questions for which we chose questions that were as simple as possible owing to the time available: only about one hour and forty-five minutes to answer all the questionnaires. This consideration is also supported by the post-experiment analyses in which the subjects agreed that this questionnaire was easy. With regard to the Effectiveness, there is no statistically significant difference between KEP and the papers. This result is probably owing to the fact that the selected papers were written by the same authors and did not therefore contain incoherent information. In general, scientific papers written by different authors could contain different and incoherent information despite dealing with the same topic. The presence of contradictory statements makes it more difficult to formulate a proper response to a problem or question addressed. All the statistical analyses presented in this section were performed using Statistica 7.0 developed by StatSoft [32].

5.3 Post-Experiment Survey
Fig. 3 shows the box plot related to each of the questions in the post-experiment survey. We used a post-experiment survey to capture the subjects' perception of the following items:

- **Easiness of Questionnaire 1, 2 and 3:** the box plot shows that the median values for questionnaire 1, 2 and 3 are respectively: “agree”, “neutral” and “disagree”. The perceived level of easiness of the questionnaires therefore increases from questionnaire 1 to questionnaire 3.
- **Sufficiency of the training time received:** the box plot shows that the median value is “agree”, and the lower and upper values of the quartile are 2 and 3. The experimental subjects therefore judged that the training time received was sufficient. This result is in accordance with our expectations since the training time was defined based on the training time given to the students in the pilot study executed in Bari.
- **Clearness of the materials provided during the experimental run:** the box plot shows that the median value is “agree”, and the lower and upper values of the quartile are 2 and 3. The experimental subjects therefore agreed that the materials received during the experimental run were clear. This result was also positively influenced by the feedback from the pilot study.
- **Clearness of questions proposed during the experimental run:** the box plot shows that the median value is “neutral”, and the lower and upper values of the quartile are 2 and 4. The experimental subjects therefore considered the proposed questions to be neither clear nor unclear.

![Figure 3. Box plot for post-experiment survey](image)

Given these results, it is possible to argue that the items considered in the post-experiment survey have not affected the experimental results.

5.4 Threats to Validity
The results of the empirical study may have been affected by the following threats: The threat to internal validity was not completely avoided because each experimental group involved in the experimentation worked with only one treatment value: KEP or scientific papers. Nevertheless, this threat was mitigated owing to the fact that the participants had about the same cultural background as regards software maintenance and the software development lifecycle. Moreover, all the participants found the material provided, the tasks, and the goals of the experiment to be sufficiently clear as is shown by the results of the post-experiment survey questionnaire. Another issue concerns the exchange of information among the participants. The participants were not allowed to communicate with each other, and this was prevented by monitoring both groups during the run. The threat to external validity was diminished by the fact that the experimental subjects were not students but 14 IT professionals contracted in projects led by the Alarcos research group at the University of Castilla-La Mancha (Spain). Nevertheless, the threats concerning external validity are the size and complexity of the experimental objects. The fact that the participants were not students but 14 IT professionals contracted in projects led by the Alarcos research group at the University of Castilla-La Mancha (Spain).
social threats. The post-experiment survey questionnaire was designed using standard forms and scales. Social threats (e.g., evaluation apprehension) were avoided as much as possible. For example, the professionals were not graded on the results obtained in the experiments. Threats to conclusion validity concern the issues that affect the ability to draw a correct conclusion. Statistical tests were used to reject the null hypotheses. In particular, we used the Mann-Whitney non-parametric statistical test because of its robustness and sensitivity. This validity could also have been affected by the reduced number of observations: only 14. Further replications on larger datasets are thus required to confirm or contradict the results.

6. Conclusion & Future Works

This paper continues our research into supporting knowledge transfer between different stakeholders. More precisely, in this paper we have analyzed whether the knowledge formalized and transferred in the form of KEPs is more comprehensible than the same knowledge when formalized and transferred in the form of scientific papers. We have therefore carried out a controlled experiment, from which the following findings have been obtained:

- KEPs are more effective in retaining knowledge than scientific papers. It is important to remember that this measure was collected using the answers provided for questionnaire 3 without the help of any material, KEP or scientific papers. This result was probably affected by the reduced size of a KEP in comparison to the size of scientific papers. A KEP, as opposed to scientific papers that deal with the same research problem, does not contain redundant information. This absence, in our opinion, favours knowledge retention.

- With regard to semantic comprehension and the transfer of knowledge, the KEP are not more effective than the scientific papers. This result was probably affected by the papers selected, since they were all written by the same authors and the information and concepts expressed were therefore homogenous and coherent between each paper. Moreover, only four papers were selected owing to the time available to carry out the experimentation. If there had been much more than four papers the difference would, in our opinion, have been greater. It might now be interesting to build a KEP from papers that report different and also opposite results and compare its effectiveness with one of the papers.

- KEPs are more efficient at transferring knowledge than the scientific papers. In other words, it is faster to find the concepts, the empirical evaluations and the empirical data expressed in a KEP than in the scientific papers. This result was probably obtained because the KEP structure provides a set of attributes, defined for each specific component, to allow a rapid selection of knowledge content. Each attribute, moreover, is further detailed in the section content of which each KEP component is made up. In other words, the KEP structure probably makes it easier to recognize and find knowledge contents in comparison to the unstructured form of a paper.

These results have encouraged us to make further studies in order to refine the KEP structure, although we are conscious that the results should be considered as preliminary. Further replications of this experiment are needed to:

- analyze other contexts, such as industrial or educational contexts; these contexts are not equal because, for instance, the industrial context has more pressures than the educational context while the latter requires more attention to educational aspects.

- validate a KEP built from papers by different authors after a “systematic localization, evaluation, synthesis, and interpretation of the evidence of past research studies” [37], in literature called a Systematic Review. In our opinion, in this case KEP might not only be more efficient but also more effective than scientific papers because the authors of KEP would solve any eventual incoherence and conflicts existing in the papers, or would highlight them.

For the sake of completeness, we should state that we are aware that the structure of KEPs requires a supplementary effort in comparison to the papers because it is produced from papers. Nevertheless, if further research demonstrates this to be true, then the KEP structure could be useful in transferring knowledge in the field of IT, as an alternative to the not so well defined structure of scientific papers.

References


