Available online at www.sciencedirect.com
ScienceDirect
Computers in Human Behavior 23 (2007) 2641–2655
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

Evaluating pedagogical classification frameworks for learning objects: A case study

Elena García-Barriocanal [a,*], Miguel-Angel Sicilia [a], Miltiadis Lytras [b]

[a] *Computer Science Department, University of Alcalá, Ctra. Barcelona km. 33600, 28871 Alcalá de Henares (Madrid), Spain*

[b] *ELTRUN, The Research Center Department of Management Science and Technology, Athens University of Economics and Business, 47A Evelpidon Str and 33 Lefkados Str, GR-113 62 Athens, Greece*

Available online 25 September 2006

Abstract

The use of toolkits and reference frameworks for the design and evaluation of learning activities enables the systematic application of pedagogical criteria in the elaboration of learning resources and learning designs. Pedagogical classification as described in such frameworks is a major criterion for the retrieval of learning objects, since it serves to partition the space of available learning resources depending either on the pedagogical standpoint that was used to create them, or on the interpreted pedagogical orientation of their constituent learning contents and activities. However, pedagogical classification systems need to be evaluated to assess their quality with regards to providing a degree of inter-subjective agreement on the meaning of the classification dimensions they provide. Without such evaluation, classification metadata, which is typically provided by a variety of contributors, is at risk of being fuzzy in reflecting the actual pedagogical orientations, thus hampering the effective retrieval of resources. This paper describes a case study that evaluates the general pedagogical dimensions proposed by Conole et al. to classify learning resources. Rater agreement techniques are used for the assessment, which is proposed as a general technique for the evaluation of such kind of classification schemas. The case study evaluates the degree of coherence of the pedagogical dimensions proposed by Conole et al. as an objective instrument to classify pedagogical resources. In addition, the technical details on how to integrate such classifications in learning object metadata are provided.

* Corresponding author. Tel: +34 91 885 66 63; fax: +34 91 885 66 46.
E-mail addresses: elena.garciab@uah.es (E. García-Barriocanal), msicilia@uah.es (M.-A. Sicilia), mdl@eltrun.gr (M. Lytras).

0747-5632/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2006.08.001

Keywords: Pedagogical classification; Learning resources; Learning objects; Classification

1. Introduction

The evolution of Web-based learning has fostered the search for methods and technologies that enable a degree of reuse of learning contents and learning activity designs (Sicilia & García, 2003). The concept of *learning object* is at the centre of a new instructional design paradigm for Web-based learning. This new paradigm emphasizes reuse as a quality characteristic of learning contents and activities. For example, the definition of learning object by Polsani (2003) explicitly includes reuse in his definition: "an independent and self-standing unit of learning content that is predisposed to reuse in multiple instructional contexts". In one of the most referenced articles on the field, Wiley (2001) also mentions the term reuse in his learning object definition: "any digital resource that can be reused to support learning". However, the concept of learning object reusability as a key quality factor for content design is difficult to characterize and measure, since it encompasses not only the evaluation of the contents themselves (Vargo, Nesbit, Belfer, & Archambault, 2003), but also a balance between their usability in specific contexts and the range of educational contexts it explicitly targets (Sicilia & García, 2003). Reusability of learning objects is thus dependant on the quality of their metadata records. Metadata enables software agents or software systems to select learning object from global repositories for some given search criteria. One of these criteria is the classification of the resources into one or several classificatory frameworks.

This paper focuses on the classification of learning objects according to their pedagogical properties. Pedagogical classification is understood here as tagging the learning objects with schemes that characterize them according to the pedagogical standpoint that was used to create them, or to the actual pedagogical orientation of their constituent learning contents and activities. From an ontological perspective, the assumptions and hypotheses of different theoretical standpoints need not be compatible or need not consider the same aspects of reality (Packer & Goicoechea, 2000). Furthermore, different pedagogical approaches are known to result in different mappings to activities (Conole, Dyke, Oliver, & Seale, 2004). In consequence, pedagogical classification is a relevant category in processes of search and browsing in digital learning resource repositories. However, pedagogical classification systems need to be evaluated to assess their appropriateness. Several aspects of such classifications can be evaluated, including its *coverage* (i.e. if the classification system covers every possible pedagogical dimension) and its level of detail (i.e. if the number of sub-classifications providing more specific categories is appropriate for a meaningful description of the resources). But these aspects have as a common requirement that of providing a degree of inter-subjective agreement on the meaning of the classification dimensions they provide. Without such evaluation, classification metadata, which is typically provided by a variety of contributors, is at risk of being fuzzy in reflecting the actual pedagogical orientations, thus hampering the effective retrieval of resources.

The evaluation described in what follows uses rater agreement techniques, which are often used to evaluate a new rating system or instrument. Since classification systems are actually measurement instruments of some pedagogical aspects aimed at expert cataloguers, this seems to be an adequate methodological approach. However, the case study described is not intended to provide a "definitive" judgment on the best classification, but to illustrate the methodology that can be used to assess and compare different alternative classification systems, existing or future. The statistical techniques used might also be used to automatically analyze data about the agreement on classifications of different raters in an on-line system as a learning object repository. This would provide a continuous assessment on how consistently classification tags are being used during the lifetime of the system, eventually leading to some re-design of the classificatory frameworks.

The rest of this paper is structured as follows. Section 2 provides background information on learning objects and pedagogical toolkits and classifications. Then, Section 3 provides the technical details for the inclusion of pedagogical classifications in standard-compatible metadata records. Section 4 reports the use of rater agreement techniques to evaluate the coherence in the use of a concrete pedagogical classification system. Finally, conclusions and future research directions are provided in Section 5.

2. Background

The concept of learning object cannot be understood without it description using metadata records, since they enable to appropriated reuse of it in different learning management platforms and in different educational context.

Pedagogical frameworks or toolkits are useful tools that support the development of pedagogically driven approaches to e-learning. These tools can be used also as an instrument to classify learning objects, supplementing criteria of a different nature. It should be noted that the purpose of these kinds of classifications is different from that of collaborative, open annotation systems. For example, *Folksonomies*[1] are tag systems collaboratively generated, open-ended labelling system that enables Internet users to categorize content such as Web pages and Web links. However, the purpose of these systems is that of informal classification, which is very different from the purpose of controlled and validated classification regarding pedagogical properties.

2.1. Learning objects and standards for learning objects

In practical terms, a learning object is a piece of Web content – or a prescribed sequence of activities to be carried out through the Web – of arbitrary type and structure, which is described by a metadata record. Metadata records provide information about the object and its prospective educational usages. Learning object metadata is thus the key to reuse. In the last years, a number of specifications and standards that describe or make use of the learning object concept have been proposed (Friesen, 2005). Regarding metadata, the basic elements associated to learning objects have been described in the IEEE LOM standard (IEEE, 2002). This standard, conceptually compatible with the well-known Dublin Core Metadata Element Set (Dublin Core, 2003), organizes its conceptual metadata schema in nine categories: general, lifecycle, meta-metadata, technical, educational, rights, relation, annotation and classification. The *general* and *annotation* categories cover descriptions as title and coverage and general purpose comments with unrestricted purposes, respectively. *Lifecycle* and *rights* cover contributors, change control and property

[1] Thomas Vander Wal. Explaining and showing broad and narrow folksonomies. http://www.vanderwal.net/random/entrysel.php?blog=1635, February 2005.

matters. The *technical* category describes technical characteristics of the Web contents, e.g. software requirements for installation. *Meta-metadata* covers the description of the metadata record itself. *Educational* describes the envisioned educational characteristics of the object, including type of interactivity, typical educational context, typical age of the intended learners and the like. The *relation* category describes relations between learning objects, which could be seen as a form of "linking" the described learning object to educational characteristics, e.g. related learning objects that constitute prerequisites or that cover semantically related elements (Sicilia, García, Aedo, & Díaz, 2004). Finally, the *classification* category serves several different purposes, including stating the objectives of the learning object, the prerequisites of the learner and the overall classification of the contents inside taxonomical schemes or ontologies.

Another remarkable effort to describe learning objects and their use is the influential ADL sharable content object reference model (SCORM) (ADL, 2003). In contrast with IEEE LOM and IMS Learning Design, SCORM is not a different specification but "a model that reference a set of interrelated technical specifications and guidelines, designed to meet high-level requirements for learning content and systems". As part of the specifications compiled by SCORM, IEEE LOM has been adopted as the metadata language for learning resources, but it also includes specifications oriented towards achieving a degree of interoperability in the functioning of *learning management systems* (LMS).

2.2. Toolkits for e-learning activities

Toolkits for learning design are defined as model-based resources that offer a way of structuring users' engagement that encourages reflection on theoretical concerns as well as supporting the development of practical plans for action (Conole & Oliver, 2002). Its application to learning technology enables supporting two different kinds of tasks: (i) the development of learning activities programs and (ii) the evaluation of existing programs or designs according to several classifying criteria.

Previous work has demonstrated that e-learning practitioners usually do not apply models and theories to design learning activities. This is often due to – as academics outside the field of education – they find the diverse array of theoretical perspectives alien and overwhelming (McNaught, 2003). As Conole et al. (2004) propose, the development of toolkits provides a way for non-specialists to engage with such theories in a manner that supports careful design and prompts productive reflection and engagement.

Nowadays, several toolkits for the design of e-learning technology activities exist. Most of them are concerned with media and resource selection (Conole & Oliver, 1998) and evaluation activities in learning processes, (Conole, Crewe, Oliver, & Harvey, 2001; Oliver, MacBean, Conole, & Harvey, 2002), which use qualities as 'authenticity' (the degree to which the evaluator seeks to control influences upon the focus of evaluation), 'scale' (the quantity of participants that the method is typically used to capture data from) and 'time' to differentiate between approaches to evaluation (such as experimental designs or naturalistic enquiries). With regards to learning activities design, one the most detailed toolkit is the one described in Conole et al. (2004).

The classification ranges of Conole et al. (2004) for learning activities design can be considered as a kind of "ordered categorical" scale, or better, as a type of Likert scale, if we consider that the qualifying scale of each component is a set of levels separated by "a same distance". The toolkit used as the basis for the study considers six components to determine criteria on the nature of the learning objects. These criteria enable the evaluation and are the following:

- A learning object has *individual* nature if the individual is the main object of the learning.
- A learning object has *social* nature if the learning is carried out by means of the interaction between the student and other people, like a tutor or other students.
- A learning object is of *reflection* nature if conscious reflection exists on the educative experience, by means of which the experience is transformed into learning.
- A learning object is of *non-reflection* nature if the learning is explained by means of processes like the agreement, the pre-conscious learning, memorization, or learning of capacities or abilities.
- A learning object is based on *information* if there is an external body of information (text or other elements) that conform the base of the experience and the fundamental material for the learning.
- A learning object is based on *experience* if the learning is reached by means of the direct experience, the practical application and the development of activities.

The authors of the toolkit provide classifications for a number of theories or approaches to learning design, along with concrete criteria for the design. However, the classification of learning objects is not explicitly addressed in their work. Since that classification is expected to be used in many cases after the process of learning object design, an evaluation on its degree of coherence and understandability is required.

3. Pedagogical issues in learning object classification

The classification of learning object is considered in LOM under the category "*9. Classification*". The aim of this category is "to describe where the object is focused in a specific classification system". Each classification assigned to an object is described using four information elements, shown in Table 1.

Table 1
LOM classification category

Number	Item	Description
9.1.	Purpose	The purpose of classifying this learning object. Value space: *discipline, idea, prerequisite, educational objective, accessibility, restrictions, educational level, skill level, security level, competence*
9.2.	Taxon path	A taxonomic path in a specific classification system. Each succeeding level is a refinement in the definition of the preceding level. Each path is described as a pair (Source, Taxon), where *source* is a reference to the name of the classification system and *taxon* is a reference to one or more particular terms within a taxonomy. A taxon is a node that has a defined label or term, and an ordered list of taxons creates a taxonomic path
9.3.	Description	A unique description of the learning object relative to the stated purpose
9.4.	Keyword(s)	Keywords and phrases (zero, one or more) descriptive of the learning object relative to the stated purpose

The LOM classification model is flexible and enables classifying using well-known public taxonomies as well as taxonomies created for a specific purpose. However, in terms of the research described in this article, category *9. Classification* should be extended in two dimensions:

1. Classification with regards to pedagogical design is not considered explicitly in any label of the value space recommended in the standard. An obvious extension is to include a generic purpose like "*design rationale*".
2. Several classificatory scales do not use nominal scales (sets of values) since classification is rough or fuzzy to some extent and thus it has a specific degree. This degree can be specified using ordinal labels or real/integer numeric scales.

In the following sub-sections these two aspects are described in more detail.

3.1. Pedagogical criteria representation

There is not an obvious and unique way to classify a learning object with regards to its pedagogical design process. For example, if an author applies "reflexive learning" as the main criterion to design one learning object, he/she might refer to an explicit representation of this technique as a taxonomical label, e.g. an instance `Reflective-Learning` of a class `LearningMethod` defined inside an ontology (Gruber, 1995). Pedagogical assumptions, guidelines and hypotheses could be subject to much more complex modelling (Sicilia, 2006). However, simple models based on axis as those described by Conole et al. (2004) are much more economical in terms of the effort spent in creating the metadata for the learning resources, and they do not require a complex tool or method for edition. This has an impact in the return of investment of metadata creation (a topic being addressed currently by the DCMI consortium[2]), and becomes a pragmatic criterion that must be considered in every concrete setting. Another important aspect of pedagogical classification for the purpose of creating metadata is that of economical criteria in the automated extraction of metadata. Even though some studies exist that have attempted to classify learning resources or programs using machine learning (Romero & Ventura, in press), the classifications obtained are about the topics of the material, which can to some extent be inferred from the contents. But pedagogical classification through automated means is still undeveloped, and due to its highly interpretive and theory-dependant nature, it is not likely that effective automated tools will be available in the near future.

However, and irrespective of the model used to classify the learning objects, some general aspects of those classifications can be stated as guiding principles in devising classificatory approaches. In what follows, we describe some principles which guide that classification.

Principle #1. The representation of pedagogical criteria must distinguish between the classification provided by the author(s) and the classifications made by other(s).

[2] http://dublincore.org/groups/corporate/

Although both the classification stated by the author(s) of the learning object and classification made by non-authors might be arguable or misleading, the nature of the former with regards to the latter is different when a search is carried out. The rationale for this is that the author's classification reveals the original pedagogical intention, despite the quality of the actual result. This information is valuable in itself and may be used as a departure point for subsequent classification processes.

Further, the learning object can be described as a whole or described in terms of its components. For example, in a SCORM (ADL, 2003) SCO, either each asset or the SCO as a whole can be described. When an activity design compliant with IMS *learning design* (Koper, Olivier, & Anderson, 2003) is described, the method can be described as a whole or in terms of its acts or activities (or both at the same time).

> **Principle #2.** Pedagogical criteria representation must allow for different levels of aggregation in the described object.

In addition, the description of learning activities can be made at different levels of detail. Continuing the example above, an author may use the following schema to represent Dewey's *reflective learning* (Conole et al., 2004): (1) Experience: Problem situation (2) Problem identification and definition (3) Gather all necessary information (4) Reflection on information and experience. (5) Theory formation (6) Test theory in practice. If this six-step schema is used, an IMS LD method with six activities could be created as a straightforward mapping. In this case, the classification would be composite: one classification to the overall unit with *taxon path* metadata referencing to `ReflectiveLearning`, and six activity classifications making reference to instances of an ontology that represent each method part. This kind of representations requires more than a taxonomy, since relations among elements must be represented. In the example, several predicates are required to semantically describe that each step is "*part of*" the reflexive learning method or theory and the order in which the steps are arranged. The consequence of this multi-level classification is that both the classification of the independent, fine-grained elements and their arrangement determine the final pedagogical properties of the learning resource. For example, changing the order of the practical activities in relation with the exposition of the theoretical material leads to two different designs. This can be viewed as an emergent property (Bunge, 2003) coming from the systemic arrangement of different learning resources. This paper considers only the classification of resources without considering the classification of its parts, but further work will be required in multi-level, complex classifications.

> **Principle #3.** Pedagogical classification criteria must allow for different levels of abstraction in the description of learning resources.

Principle #3 refers to the complexity of the models used for the description. This paper will later deal with a multi-axis description. This can be viewed as an extension of common classificatory frameworks that are based on the assignment of resources to one or several labels from a predefined vocabulary or taxonomy. However, more complex models can be used. Ontologies representing theoretical positions or guidelines (Sicilia, 2006; Sicilia & Lytras, 2005) go beyond label assignment, and can be used to state how coherent is a given resource with each of the elements in a set of guidelines for a concrete pedagogical standpoint. Or these models can be used to create formal statements regarding the pedagogical properties explained. However, these practices are still undeveloped and require a significant amount of resources, which has led us to focus first on standard label assignment systems.

Principles #2 and #3 together entail a high complexity level that is not fully tackled with here, since the empirical study described later only uses a categorization in numerical axes with a high level of abstraction. In any case, the precondition for a classification system to be useful is that it provides a shared and understandable framework to assist experts or educators in the task of classifying. This is stated in the fourth principle.

Principle #4. Pedagogical classification criteria for learning resources should be assessed in terms of the consistency in the outcomes of classification processes.

This paper deals with a technique to assess the coherence and understandability of common label-based classificatory systems for learning resources, which is considered a precondition for improving the processes of seeking learning resources based on their educational properties.

3.2. Extension to non-categorical criteria

As previously described, category `9. Classification` of LOM is flexible enough to allow the representation of specific metadata to describe categorical pedagogical criteria, but some extensions are required.

Non-categorical elements play an important role in evaluation and design of learning contents, and several toolkits provide criteria that are evaluated using Likert scales (DeVellis, 1991). The values must be attached to the term of the classification they mention, so that an extension of *taxonpath* elements is needed to include non-categorial elements. The addition of a new label inside a taxon provides the required information for the goal pursued. In this label the Likert value corresponding to the evaluation of the criteria under study must be specified. Creation of a new label and not of an attribute of another element of `taxon`, e.g. `entry`, has been chosen since a label expresses a concept different to existing ones and enables the inclusion of new attributes that described it in detail.

Hereafter an example to illustrate the use of this new label is provided. Taken Conole toolkit as point of departure, a category named `ConoleCriteria` can be chosen to represent the toolkit evaluation criteria as follows.

```
[1] ConoleCriteria
  [1.1] Individual–Social
  [1.2] Nonreflection–Reflection
  [1.3] Experience–Information
```

Metadata of a highly reflexive learning object should include in taxonpath as source Conole criteria taxonomy (**ConoleCriteria**) and as reference to the "path" in the taxonomy, the criteria Non-reflection–Reflection (**1.2**). A label **<LikertValue min** = "**0**" max = "**10**"**>** reflecting Likert scale should be introduced to specify the value corresponding to the criteria.

```
<classification>
     <purpose>
          <selection-source>author</selection-source>
          <source>
               <langstring xml:lang = "en">IE.R.U</langstring>
          </source>
          <value>
               <langstring xml:lang = "en">
               LearningDesign
               </langstring>
          </value>
     </purpose>
     <taxonpath>
          <source>
               <langstring xml:lang = "en">
               ConoleCriteria
               </langstring>
          </source>
          <taxon>
               <id>1.2</id>
               <entry>
                    <langstring xml:lang = "en">
                    Nonreflective-Reflective
                    </langstring>
               </entry>
               <LikertValue min = "0" max = "10">10</ LikertValue>
          </taxon>
     </taxonpath>
</classification>
```

It should be noted that the fragment above contains a new label <selection-source> to specify the author the pedagogical criterion (according to Principle #1), author in this case. The author himself of the learning object can be specified in LOM category 2.3Contribute, concretely using metadata 2.3.1 Role (value *author* of the metadata value space) and 2.3.2 Entity (identification of and information about the author).

4. Case study: rater agreement on a multi-axis classification

Several alternative pedagogical frameworks to classify learning objects can be devised. However, in order to obtain an appropriate categorization these two requirements must be fulfilled: (i) the existence of a theoretical framework defined for it, which could be used to compare the framework with similar ones (or at least with other learning theories or learning approaches) and (ii) an assessment of the coherence in use of the classificatory system.

The first requirement (i) enables the comparison of the framework according to existing consensual knowledge. A possible way to elaborate such framework is to connect it with public ontologies of learning (Sicilia, 2006). Although nowadays a single commonly accepted ontology – in a formal way (Gruber, 1995) – does not exist, Sicilia and Lytras work (2005) has revealed some of the difficulties that must be overcome to achieve it.

In the rest of the paper a study on the second requirement (ii) is provided. Concretely, the coherence of classifiers following Conole et al. criteria (2004) is studied in a concrete setting. The study does not aim to obtain a definitive statement about the goodness of the toolkit used, but to be useful as an example of the kind of methodology and research required with respect to pedagogical classification frameworks. Pedagogical classification systems for learning resources are currently used without a concern for their quality, which makes the resulting classifications questionable. Here we provide a first attempt to improve the practice of classification by inquiring on the quality of the coherence and understandability of classification frameworks.

4.1. Case study description

According to the original scale of the authors, the evaluator must provide assessments in the range [0, 10] for each pair of criteria described above. Thus, a value corresponding to this scale for each criterion-axis will have to be specified: *Individual–Social*, where 0 means completely "individual" and 10 "purely social"; *Nonreflection–Reflection*, where 0 means "absolutely nonreflective" and 10 "purely reflective".

Following this method, a total of 12 learning objects extracted from the learning objects repository MERLOT (http://www.merlot.org) have been evaluated (see Table 2). The selection of the objects was carried out by participants, who did not take part in the evaluation, and it covered learning objects of different subjects and approaches. Table 2 provides their names and the MERLOT ratings for them when available. MERLOT was used as a reliable source of resources with good quality, considering their built-in quality control mechanisms (Cafolla, 2002).

Table 2
Learning objects used in the study with evaluations at 15th August 2005

#	Name in MERLOT	Avg. peer reviews	Avg. member comments
O1	Information security management concepts	4.33	–
O2	Recursion explained	–	4.0
O3	Binary number conversion	–	–
O4	Applied computing course material	–	–
O5	Integers	5.0	5.0
O6	Web activities: an introduction to baroque art	–	4.0
O7	Intro to computer engineering	–	–
O8	Upgrading and repairing PC's video archive	–	–
O9	Introduction to level of measurement module	4.33	5
O10	Graphic derivative	4.0	–
O11	Entrepreneurship classroom activities	5.0	–
O12	Teen travel experts	4.0	–

An example of the outcomes of an evaluation is provided in Table 3. The first column provides the evaluator identifier and the other three columns contain the evaluations for the learning object labeled like *O12* ("Teen travel experts", http://www.Nevada.edu/~brunelle/travelexperts.html).

The same schema was applied for each of the learning objects cited in Table 2. The evaluators were homogeneous in their background knowledge. All of them were educators with more than four years of experience in instructional design.

Techniques for evaluators' agreement analysis (*rater agreement*) are commonly used to statistically analyze the degree of coherence of several evaluators. In the work described in this paper, the aim of the research is not assessing a particular group of evaluators, but finding some initial empirical evidence for the coherence of the classification model in the described categories (Individual–Social, Experience–Information, and Reflection–Nonreflection). Therefore, the instrument of measurement in this case is the scale for the criteria used in the toolkit. The description provided to the evaluators was limited to the definitions and the examples that are described in the article of Conole et al. (2004).

The selection of the appropriate rater agreement statistical technique for the study depends on the type of scale. In our case, we are dealing with a Likert scale, which represents data based on intervals where the categories (the scale from zero to ten) can be considered as equally spaced. Once determined the type of scale, the basic premise of the study is that different components in the agreement exist, and they must be separated to improve the coherence of the classification. Concretely, we will analyze here the bias of the evaluator (*rater bias*) and the association between evaluator (*rater association*).

4.2. Case study: result analysis

Table 4 provides the standard deviation of the evaluations that each evaluator has provided for each evaluated criterion, which can be used to evaluate rater bias.

Deviations in the data provided in Table 4 are not significant in most of the cases. The exception is a higher deviation in the seventh evaluator. Such deviations require further

Table 3
Example of evaluation for a learning object

	Individual (0)–Social (10)	Nonreflection (0)–Reflection (10)	Experience (0)–Information (10)
Evaluator 1	8	8	1
Evaluator 2	6	7	4
Evaluator 3	8	7	3
Evaluator 4	0	5	5
Evaluator 5	6	6	3
...	...	...	...

Table 4
Standard deviation by criteria and evaluator

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Individual–Social	0.68	0.40	0.23	0.23	0.23	0.15	1.18	0.23	0.40	0.02
Reflection–Nonreflection	0.65	0.32	0.27	0.60	0.43	0.77	1.98	0.43	0.10	0.35
Experience–Information	0.56	0.02	0.21	0.18	0.02	0.33	2.28	0.64	0.25	0.48

inquiry to determine the reasons of the deviation, which roughly may fall in a deviation in the understanding of the scale, or other factors not directly related to the measurement scale. One option might be that of removing the evaluator from the study, but this requires some form of informed input to avoid arbitrary changes in the evaluator set. The solution adopted was that of initiating an interview with the rater to gather more information in the divergent assessments, and another in parallel with one of the other raters, for the sake of contrast. The technique used for that was a variant of the "thinking aloud" protocol (Lewis & Reiman, 1993) that is used commonly in user interface design. The conclusions of the interview were that the evaluator did not understand the axis scales in the same form than the others. The ratings were consistently higher in each of the axis in the direction of "social", "reflective" and "experience". The difference in the divergent rater was that these points in the axis were considered: (a) "positive" in the sense of being considered as the "correct" points from the pedagogical viewpoint, and (b) the ratings were not considered in a comparative sense with the rest of the learning objects. Regarding (b), the divergent rater was not able to correctly explain the rationale of the divergences between the ratings of different learning objects in the same scale, while the other rater was.

This leads to several important recommendations for these rater bias assessment procedures:

- A more clear explanation of the "relative" nature of the ratings must be provided. The arrangement of an *a priori* scan of a set of learning objects before starting the rating helps in realizing the potential divergences between different learning objects in each of the axis.
- The text or explanations of the different extremes of the axis must be carefully devised to avoid any bias towards one extreme. For example, in the "Social–Individual", the descriptions will easily go in the direction of presenting the social approach as the "right" one from the viewpoint of pedagogy.

With respect to the association between evaluators, factor analysis has been used as base tool, considering the association of each evaluator with a latent factor. Table 5 shows the loadings of each evaluator with the latent factor including E7 ratings (Table 6 shows loadings with evaluator E7 exclusion). These measures can be interpreted like superior levels of the validity of the evaluations (Uebersax, 1992). A high load for a evaluator does not provide very useful information, but low loadings must be interpreted as an evidence of little correlation between the evaluation of the evaluator and the latent factor.

According to the results, approximately a 56% of agreement in the Reflection–Nonreflection axis and a 50% of rater agreement in the Experience–Information axis have been found. This fact can be explained due to the latent factor identified (proportion of total variance 56.7% without evaluators exclusion and 55.4% excluding E7 in Reflection–Nonreflection criteria and proportion of total variance 49.4% without evaluators exclusion and

Table 5
Loadings with respect to the common factor for criteria without evaluator exclusion

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Reflexivo–No reflexivo	0.719	0.195	0.834	0.740	0.637	0.592	0.818	0.907	0.921	0.880
Experiencia–Información	0.647	0.204	0.716	0.531	0.683	0.649	0.662	0.700	0.998	0.934

Table 6
Loadings with respect to the common factor for criteria with evaluator exclusion

	E1	E2	E3	E4	E5	E6	E8	E9	E10
Reflection–Nonreflection	0.693	0.198	0.824	0.741	0.633	0.608	0.927	0.917	0.875
Experience–Information	0.648	0.206	0.715	0.529	0.683	0.649	0.699	0.998	0.934

50% excluding E7 in Experience–Information criteria). Results fairly explain the agreement ($p = 0.119$ for Reflection–Nonreflection criteria and $p = 0.406$ for Experience–Information criteria excluding E7 in both cases).

Strong loadings in all the evaluators except in E2 can be observed for the criteria analyzed. This seems to point out that divergences cannot be attributed to the scale, but to the interpretation of the evaluator. The analysis of the case of this second evaluator was carried out by means of a technique structured as a simplified Delphi process (Linstone & Turoff, 2002) focused on learning objects O2, O8 and O11 to Reflection–Nonreflection criterion and O4, O5 and O6 to Experience–Information criterion, in which evaluator E2 has a bigger disagreement with respect to the other evaluators.

It should be noted that the loadings for the Individual–Social criterion resulted in a system that was computationally singular. The correlation matrix for the Individual-Social axis shows strong correlations (minimum value = 0.814 and average = 0.943 including E7 and minimum value = 0.857 and average = 0.959 excluding E7). In consequence, no negative evidence about the coherent understanding of the Individual–Social criteria has been found.

In general terms, the results of the case study point out that the pedagogical classification framework under study is understandable and provides a good ground for coherent classifications.

The assessment technique just described is useful for the assessment of newly created classification systems, and also to evaluate how some existing ones are actually being used. However, learning object classifications will eventually be collected in the context of a large learning object repository (Nash, 2005). This provides opportunities for the continuous assessment of classificatory processes, taking into consideration the continuous creation of classifications. In fact, collaborative filtering systems have yet used some kind of rater bias analysis for producing recommendations (Anderson et al., 2003).

5. Conclusions and future work

The classification of learning objects may serve different qualifying intentions, as it is reflected in the standard IEEE LOM of metadata. However, the standard explicitly does not address the classification according to pedagogical design issues. These kinds of criteria are fundamental in the search of objects when users (humans or software systems) attempt to retrieve learning objects that have certain properties regarding the standpoints of the designers about pedagogy and human learning.

A flexible way to extend metadata schemes in order to integrate this kind of information is as a first contribution of this paper. Secondly, the validity of the classification task has been approached. An initial but important condition of classificatory frameworks is that they are understood by different experts consistently, thus being able to result in valid and coherent classifications. A case study regarding this last aspect has been reported in this paper, describing a general method for such kind of evaluations based on rater agreement assessment. The technique is based on factor analysis and could be used not only for concrete assessment processes as the one reported, but also as a tool for continuously evaluating the coherence in the assessments inside learning object repositories.

Further work should deal with the assessment of more complex classificatory frameworks, not necessarily based on rating a set of pedagogical axis. However, the formal expression of such richer and more expressive frameworks still requires a great deal of attention, especially in the field of ontologies of learning.

Another direction for future research is that of the computational properties of the automated assessment of databases of classifications of learning objects in large repositories. For the technique to be scalable, some kind of optimization similar to those used in collaborative filtering approaches would be needed (Sarwar, Karypis, Konstan, & Riedl, 2000), since the volume of users and learning resources in open Web repositories is potentially so large that conventional statistical techniques are not computationally efficient.

Acknowledgements

This work has been supported by project LUISA (Learning Content Management System Using Innovative Semantic Web Services Architecture), code FP6-2004-IST-4 027149 and by project PAWSEL funded by the IV PRICYT programme (CAM-UAH2005/046).

References

ADL (2003), Sharable Courseware Object Reference Model (SCORM) Version 1.3, Advanced Distributed Learning (ADL). Available from http://www.adlnet.org.

Anderson, M., Ball, M., Boley, H., Greene, S., Howse, N., Lemire, D., McGrath, S. (2003). RACOFI: a rule-applying collaborative filtering system. In *Proc. IEEE/WIC OLA'03*, Halifax, Canada, October.

Bunge, M. (2003). Emergence and convergence: qualitative novelty and the unity of knowledge. Toronto: University of Toronto Press.

Cafolla, R. (2002). Project merlot: bringing peer review to web-based educational resources. In *Proceedings of the USA society for information technology and teacher education international conference* (pp. 614–618).

Conole, G., Crewe, E., Oliver, M., & Harvey, J. (2001). A toolkit for supporting evaluation. *The Association for Learning Technology Journal, 9*(1), 38–49.

Conole, G., Dyke, M., Oliver, M., & Seale, J. (2004). Mapping pedagogy and tools for effective learning design. *Computers and Education, 43*(1–2), 17–33.

Conole, G., & Oliver, M. (1998). A pedagogical framework for embedding C&IT into the curriculum. *The Association for Learning Technology Journal, 6*(2), 4–16.

Conole, G., & Oliver, M. (2002). Embedding theory into learning technology practice with toolkits. *Journal of Interactive Educational media, 8*. Available fromhttp://www.jime-open.ac.uk/.

DeVellis, R. F. (1991). *Scale development: Theory and applications. Applied social research methods series* (26). Newbury Park, CA: Sage.

Dublin Core Initiative (2003). *Dublin Core Metadata Element Set, version 1.1: Reference description.*

Friesen, N. (2005). Interoperability and learning objects: an overview of e-learning standardization. *International Journal of knowledge and learning objects, 1*, 23–31.

Gruber, T. (1995). Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human–Computer Studies, 43*(5/6), 907–928.

IEEE Learning Technology Standards Committee (2002). Learning object metadata (LOM), final draft standard. In *IEEE 1484.12.1-2002*.

Koper, R., Olivier, B., & Anderson, T. (2003). *IMS learning design information model, version 1.0*. IMS Global Learning Consortium, Inc.

Lewis, C., & Reiman, J. (1993). *Task centered user interface design: a practical introduction*. Boulder, CO: University of Colorado.

Linstone, H. & Turoff, M. (Eds.). (2002). *The Delphi method: techniques and applications*. Available from http://www.is.njit.edu/pubs/delphibook/.

McNaught, C. (2003). Identifying the complexity of factors in the sharing and reuse of resources. In A. Littlejohn (Ed.), *Reusing online resources – a sustainable approach to e-learning* (pp. 199–211). London and Stirling: Kogan.

Nash, S. (2005). Learning objects, learning object repositories, and learning theory: preliminary best practices for online courses. *Interdisciplinary Journal of Knowledge and Learning Objects, 1*, 217–228.

Oliver, M., MacBean, J., Conole, G., & Harvey, J. (2002). Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning, 18*(2), 199–208.

Packer, M., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: ontology, not just epistemology. *Educational Psychologist, 35*(4), 227–241.

Polsani, P. R. (2003). Use and abuse of reusable learning objects. *Journal of Digital Information, 3*(4).

Romero, C. & Ventura, S. (in press). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*.

Sarwar, B.M., Karypis, G., Konstan, J.A., & Riedl, J. (2000) Analysis of recommender algorithms for e-commerce. In *Proceedings of the ACM e-commerce 2000 conference*, October 17–20 (pp. 158–167).

Sicilia, M. A. (2006). Semantic learning designs: recording assumptions and guidelines. *British Journal of Educational Technologies, 37*(3), 331–350.

Sicilia, M. A., & García, E. (2003). On the concepts of usability and reusability of learning objects. *International Review of Research in Open and Distance Learning, 4*(2).

Sicilia, M. A., García, E., Aedo, I., & Díaz, P. (2004). Using links to describe imprecise relationships in educational contents. *International Journal for Continuing Engineering Education and Lifelong Learning, 14*(3), 260–275.

Sicilia, M. A., & Lytras, M. (2005). On the representation of change according to different ontologies of learning. *International Journal of Learning and Change, 1*(1).

Uebersax, J. S. (1992). A review of modeling approaches for the analysis of observer agreement. *Investigative Radiology, 27*, 738–743.

Vargo, J., Nesbit, J., Belfer, K., & Archambault, A. (2003). Learning object evaluation: computer mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications, 25*(3).

Wiley, D. A. (2001). *The instructional use of learning objects*. Association for Educational Communications and Technology, Bloomington.