

# An Extension of the Eindhoven Classification Model to the Educational Sector

Margarida Figueiredo<sup>1</sup> Lília Vicente<sup>2</sup> Henrique Vicente<sup>1</sup>

Victor Alves<sup>3</sup> José Neves<sup>3</sup>

<sup>1</sup>Department of Chemistry & Évora Chemistry Centre, University of Évora, Évora, Portugal

<sup>2</sup>Agrupamento de Escolas de Reguengos de Monsaraz, Reguengos de Monsaraz, Portugal

<sup>3</sup>Department of Informatics, University of Minho, Braga, Portugal

## Abstract

This work presents an extension of the *Eindhoven Classification Model* to sort adverse events root causes for the *Educational Sector*. *Extended Logic Programming* was used for knowledge representation and reasoning with defective information, allowing for the modelling of the universe of discourse in terms of default data, information and knowledge. Indeed, a systematization of the evolution process of the body of knowledge in terms of *Quality of Information (QoI)* embedded in the *Root Cause Analysis* was accomplished, i.e., the knowledge representation and reasoning system proposed led to a process of *QoI* quantification that allowed the study of the event's root causes, on the fly.

**Keywords:** Eindhoven Classification Model; Knowledge Representation and Reasoning; Education; School Dropout.

## 1. Introduction

Education is a powerful driver of development and one of the soundest instruments for reducing countries poverty. Although there has been great progress in the last decade, a large number of young people that finish their education did it without acquiring basic skills necessary for work and life. This is particularly det-

rimental when unemployment is high and labour markets are demanding more skills than ever before [1].

The reasons why young people have poor skills when they finish their school careers are many and diverse. Solving this problem requires detailed knowledge of the causes that lead to this situation. The school dropout is one of those causes. It is a complex phenomenon, resulting in economic and social losses, either to the individual, family or the community to which the person belongs. If the school dropout is large in a country or in a developed region, the consequences will be mainly damaging in terms of economic competitiveness and social environmental degradation [2].

In 2012 Portugal had a dropout rate of 20.8%, according to data existing in the report of the official statistical office of the EU. Spain, with 24.9%, and Malta, with 22.6%, were the only European countries that in the past year had higher values. Despite the decrease (from 38.8% to 20.8% between 2005 and 2011), this rate is still far from the national goal of a 10% dropout from secondary schools, intended for 2020 [3].

The school is a complex and multifaceted system where varied types of adverse events occur. An adverse event may be defined as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim,

and includes problems in practice, relationships, procedures and systems. The most effective way to prevent adverse events is to attack directly their causes. Preventing the adverse events' root causes improves significantly the educational success. Thus, the proposed model will focus primarily on preventing the adverse events' root causes. The model planned serves as the formal foundation to an adverse event reporting and learning computational system.

## 2. The Computational Model

An educational version of the *Eindhoven Classification Model (ECM)*, with the extensions and adaptations for the educational field and its causal tree, used to classify the adverse events' root causes in school dropout, is presented. The theoretical foundation is based on an extension to *Logic Programming*, in terms of a revision of its knowledge representation and reasoning mechanisms. Undeniably, the introduction of explicit negation in this universe allowed for the development of a process of quantification of the *Quality of Information (QoI)* of the predicates extensions that make a logic program or theory, making possible to study the event's root causes and to generate alerts and recommendations on quality improvements in the didactic process.

### 2.1. The Eindhoven Classification Model (ECM)

The *ECM* was originally developed to manage human error in the chemical process industry [4], and was then applied to various other industries, such as steel industry, energy production and in healthcare. The *Eindhoven Classification Model – Medical Version* consists of 20 (twenty) codes, divided into four categories frequently used in a medical environment to classify the underlying causes of the ad-

verse events [5, 6]. This approach assumes that humans are fallible and that errors are to be expected in every organization, so it is necessary to concentrate efforts on the conditions under which individuals work and try to build defences to avert errors or to mitigate their effects. Assigning codes to the causes of each adverse event, it is a practice that is useful for tracking and trending.

The first step when using the *ECM* based classification system is to identify the root causes that result in a specific adverse event. These root causes are subsequently classified according to the classification model. To do so, a causal tree is built and techniques of *Root Cause Analysis (RCA)* are applied. Once the root causes are identified, they may be used to provide a more realistic view of how the system really works, as well as to contribute to the creation of effective and lasting solutions.

### 2.2. The Extended Eindhoven Classification Model

The *Extended Eindhoven Classification Model (EECM)* was adapted from the *ECM*, presented in the previous section. To apply this model to the educational sector, extensions were developed for each category of the original model. These extensions allow fitting each category into the educational arena and provide a broader view of the events that may occur and the degree of complexity of this field. Thus, the classification process becomes easier and more efficient. Figure 1 presents the *EECM* flow chart and the codes assigned to classify each adverse event. For instance, in the original model, the adverse events classified as “*Human behaviour – Knowledge-based errors*” (*HKK*) occur due to “*the inability of an individual to apply existing knowledge to a novel situation*”. In the *EECM*, this definition was extended by saying that the events classified under this category are

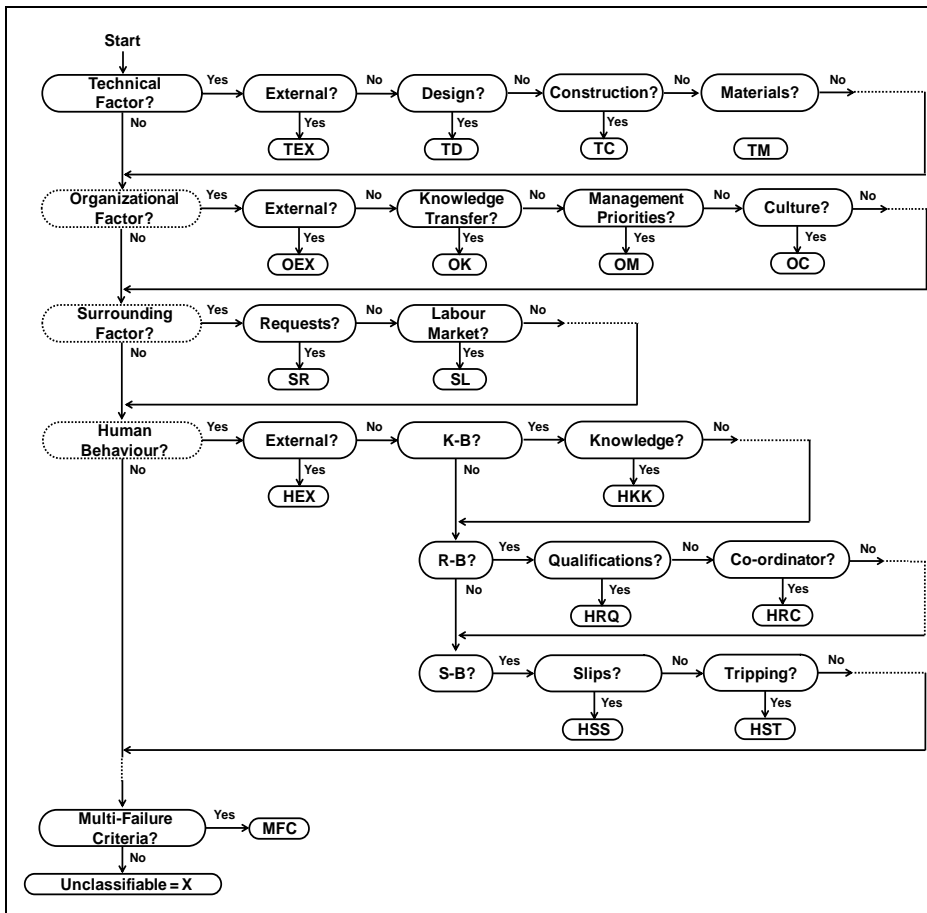


Fig. 1: The Eindhoven Classification Model for Education.

due to “difficulties in execution, interpretation or reporting procedures”. Some of the adverse events falling into this category are “poorly executed procedures, incomplete procedures and procedures poorly validated”.

The causal trees taken on by the original ECM, set that the recognition of the event’s root causes and its mental picture, is done under a hierarchical structure. On the other hand, once one has to deal with incomplete and even contradictory information, an *Extension of Logic Programming (ELP)* was used to knowledge representation and reasoning, in order to get a truth value in the interval 0...1 as a

measure of confidence in any qualification process susceptible to be handled by the system. Since an event may only occur due to the combination of more than one cause, and a different event may come about due to two or more causes, taken separately, in the original model *AND-gates* and *OR-gates* are used to embody these two possibilities in the causal tree. The usual situations may also include the case where only one cause leads to the occurrence of a certain event. In any case the adverse events’ origins are known, i.e., there are sure about the events’ grounds. Beyond these situations, it may happen that the causes of an event,

action or decision are unknown; may be known that certain views are the source of a given event, but may not be sure what are the event grounds; or it is not allowed to know the origin of a given event (e.g. due to internal policies of the schools). Therefore, it is proposed the use of “unknown” and “forbidden” operators, to allow the representation of unknown values of an infinite set of values, unknown values of a given set of values, and values not allowed or forbidden. The information contained in each causal tree is then represented in *ELP* by the extensions of a predicates set, being also used as a formalism to quantify the causal tree § *QoI* (see Section 2.4). The *QoI* allows the identification of the causes that should be taken into account, in first place, and how this hampers all the classification process. The information obtained in this way to the *RCA* enables automatic report generation with improvements in the recommendations. Figure 2 presents the application of the *EECM* to the adverse event *School Dropout*. In the source of this event there is a great diversity of reasons. It is possible that only one situation might be enough for the event to occur or, perhaps, it may be necessary a combination of several factors. The causal trees should include all possible causes and aim to be a generic representation of the problem. For a particular occurrence of the event, its causes will fall on a branch of the tree.

### 2.3. Knowledge Representation and Reasoning

In the past few decades, many non-classical techniques for modelling the universe of discourse and the reasoning procedures of intelligent systems have been proposed, in addition to classical ones [7]. Of particular interest to this work are the techniques to deal with incomplete, inconsistent, contradictory, default and forbidden information [8]. Intelligent systems require the

ability to reason with incomplete information, since in the real world complete information is hard to obtain, even in the most controlled situations. The idea behind default information is the ability to make assumptions or to jump to a plausible conclusion, derived from a knowledge base in the absence of information to the contrary. The derived information is defeasible, because in light of new information the conclusion may need to be retracted, i.e., we are in the presence of non-monotonic reasoning [8]. A suitable logic is needed, one that permits the representation of incomplete, inconsistent and default information and supports non-monotonic reasoning. In a classical logical theory or logic program, the proof of a theorem (here understood as a question submitted to the classification system) the outcome is a truth value, namely true or false [9]. *ELP* introduces another kind of negation, strong negation, represented by the classical negation sign  $\neg$ . In most situations, it is useful to represent  $\neg A$  as a literal, if it is possible to prove  $\neg A$ . In *ELP*, the expressions *A* and *not A*, being *A* a literal, are extended literals, while *A* or  $\neg A$  are simple literals. Intuitively, *not A* is *true* whenever there is no reason to believe *A*, whereas  $\neg A$  requires a proof of the negated literal.

Every program is associated with a set of *abducibles*, which may be seen as *hypotheses* that provide possible solutions or explanations of given queries, being given here in the form of *exceptions* to the extensions of the predicates that make the program. The issue is providing expressive power for representing explicitly negative information, as well as to directly describe the closed world assumption for some predicates, also known as predicate circumscription [10]. Three types of answers to a given question are then possible, i.e., *true*, *false* and *unknown*. The representation of null values will be scoped by the *ELP*. It is possible to consider three

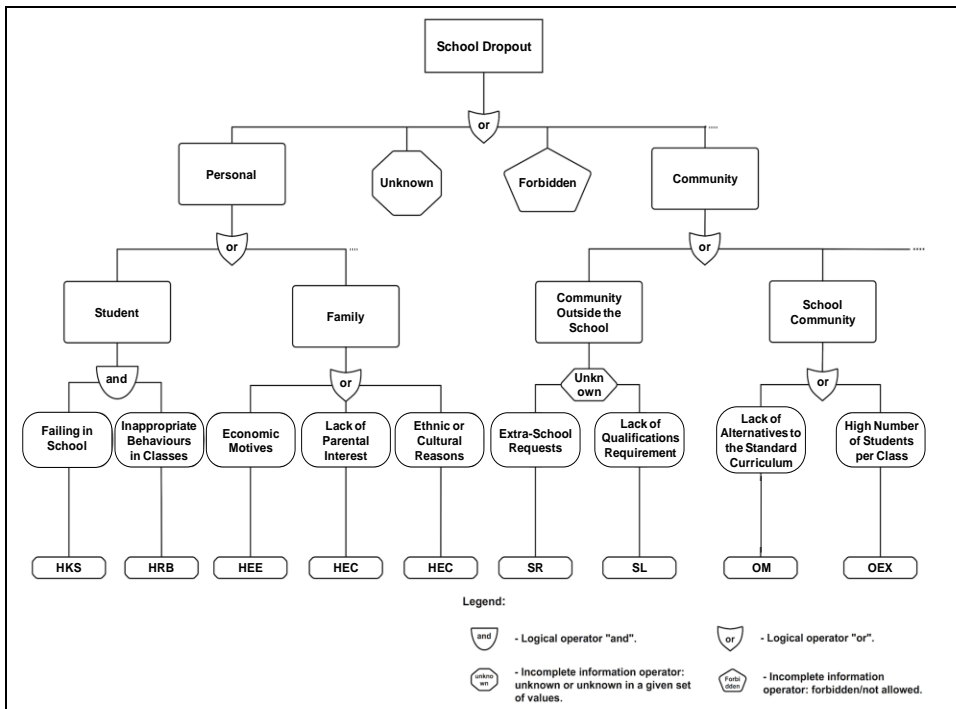


Fig. 2: Extended Causal Tree for the Adverse Event *School Dropout*.

types of null values: the former will allow for the representation of unknown values, not necessarily taken from a given set of values, the middle one will represent unknown values taken from a given set of possible values, and the latest will define values that are not allowed or forbidden. Taking the example of the adverse event *School Dropout* (Figure 2) it might represent all the possible situations according to the following setting:

- It is known that the student is in school dropout because he/she is failing in school – *known value*;
- It is only known that the student is in school dropout. In this case who reported the adverse event did not know which actions or decisions led to the event occurrence – *unknown value*;
- The professional that recorded the school dropout situation only reported family reasons. It is not possible to be constructive, concerning the action or truth-value to consider. However, it is false that the action or decision could be different. This situation suggests that the lack of knowledge may be associated to a set of possible known values – *unknown value in a finite set of values* (in this case there are three possibilities, i.e., *economic motives*, *lack of parental interest* or *ethnic/cultural reasons*);
- And finally, namely due to internal policies of the educational institution, it is not permitted to know the causes of a given event – *forbidden or not allowed values*.

Using *ELP*, as the logic programming language, it is now possible to set a procedure given in terms of the extension of a predicate called *demo*: *question, answer*  $\rightarrow [0,1]$ . Given a question (*Q*), it returns a solution based on a set of assumptions, where question indicates a theorem to be proved and answer denotes a truth value (see Program 1; *true* (1), *false* (0), being *unknown* (U) in the range of the truth values ]0,1[).

```
demo(Q,T) ← Q
demo(Q,F) ← ¬Q
demo(Q,U) ← not Q ∧ not ¬Q
```

Program 1: Extension of meta-predicate demo.

## 2.4. Quality of Information

The *Quality of Information (QoI)* with respect to the extension of a generic predicate *p* may be analysed in different contexts and measured in the interval [0,1]. When the information is *known*; when the information is *unknown*; when the information is *unknown* but can be taken from a set of values. If the information is *known* the *QoI* for the extension of predicate *p* is 1. For situations where the value is *unknown* the *QoI* is given by:

$$\lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0)$$

Finally, if the information is *unknown* but can be derived from a set of values, the *QoI* is set in terms of  $1/Card$ , where *Card* denotes the cardinality of the abducibles set for *p*.

As an example the *QoI* associated with the information about the *RCA* of the adverse event *School Dropout*, for the first three cases present in the previous section, is depicted in Figure 3 and is given in the form:

```
Vaction_or_decision (former case) = 1;
Vaction_or_decision (middle term case) = 0;
Vaction_or_decision (latest case) = 0.33.
```

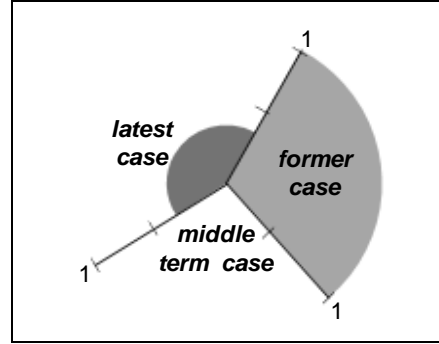


Fig. 3: The embedded *QoI* with respect to the question *Which are the actions or decisions that led to the adverse event occurrence?*.

It is now possible to measure the *QoI* associated to the question put in context, in terms of a logic program that endorses procedures of *action\_or\_decision*, which may be given in the form *Which are the actions or decisions that led to the adverse event occurrence?*. The shaded *n*-slices (here *n* is equal to 3 (three)) of the circle denote the *QoI*.

## 3. Discussion

Based on the formal approach referred to above, an adverse event reporting and learning system was introduced. Indeed, to the professionals of education and mostly for all the education institutions, this approach may bring some advantages. After the adverse events being registered, similar to what happens in other reporting systems, the analysis process becomes easier, more expedite and reliable. Undoubtedly, with the recourse to *ELP*, leading to an on the fly measurement of the *QoI* of the logic terms used in the process of judgement (in terms of a theorem to be proved), the human intervention in the analyze process is only necessary to approve the recommendations, causes and events that may need attention. It also caters for the credibility and the

measurement of the efficacy of the implemented strategies and actions.

Although the causal classification of events is sometimes time-consuming and difficult, with the development of a generic causal tree for each possible event, the increase in time consuming is on the initial phase of the model enforcement. The *QoI* allows the ordering of causes, identifying the ones that should be taken into account in the first place. In the generic tree it is necessary to consider all possible causes, rather than the most probable or usual ones. The information obtained is useful in identifying possible trends and areas requiring further investigation.

The conceptualized logic model offers the means for knowledge extraction, providing the identification of the most significant causes and suggestions of changes in the organization policies and procedures of schools, subject to formal proof. Indeed, the creation of an inference system in support of the logical model enables the generation of reports with strategies for quality improvement on the fly, where a quality measure of the system is on one's confidence on the results, in terms of the *QoI*.

#### 4. Conclusions

The main contribution of this work is to be understood in terms of the evaluation of the *QoI* in the *RCA* and the possibility to address the issue of incomplete information, through the use of an *Extension to Logic Programming (ELP)* in the construction of causal trees. *ELP* was used for knowledge representation and reasoning with defective information, catering for the modelling of the universe of discourse in terms of incomplete, inconsistent, forbidden and default data, information and knowledge. A systematisation of the body of knowledge's evolution about *QoI* embedded in the *RCA* was accom-

plished. A way to solve the representation problem of defective information was presented, adequate for evaluating the *QoI* in such situations. It was also presented a computationally feasible formal tool to measure the value of *QoI*. With this approach to *RCA* and classification it was possible to identify the causes, actions and decisions that may lead to the adverse events and define the strategies to prevent them.

#### 5. Future Work

In the future an *Adverse Event reporting and learning System applied to the Educational Sector (AESES)* will be developed. The *AESES* will comprise 3 (three) core modules, making it not only a system for adverse event registration, but also a learning system. The *Adverse Event Reporting Forms in the Educational Sector (AERFES)* module will provide a *Web* interface for adverse event registration. The effort on this interface will be focused in its usability. The event registration will be made by professionals of the educational institutions (e.g. teachers, administrative staff, auxiliary staff), by the students and by their parents, through predefined forms adapted to each user profile. The *Adverse Events Manager Reports in the Educational Sector (AEMRES)* module will be also *Web* based and aims to enable the analysis of the adverse events recorded by *AERFES*, based on the *Extension of the Eindhoven Classification Model (EECM)*. The system will provide an individual report for each adverse event recorded, which will include all its details and the extended causal tree obtained using the *EECM*. The *AEMRES* module will also provide charts with statistical information about the impact, place of occurrence, type of form and type of event recorded. Finally, the *Adverse Events Knowledge Manager in Educational Sector (AEKMES)* module

will use the data from the system database to create a *Knowledge Base (KB)*, which although had been given in terms of *ELP*, will be rewritten to productions in the logic programming language *PROLOG* [11], based on the *EECM*. From the *KB* other reports relevant to the improvement of the scholar environment may be generated, always with the assurance of data reliability and credibility, by taking into account its *QoI*.

### Acknowledgements

This work is funded by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980.

### References

- [1] [http://siteresources.worldbank.org/EDUCATION/Resources/ESSU/Education\\_Strategy\\_4\\_12\\_2011.pdf](http://siteresources.worldbank.org/EDUCATION/Resources/ESSU/Education_Strategy_4_12_2011.pdf).
- [2] T. Andrei, D. Teodorescu, B. Oancea, "Characteristics and causes of school dropout in the countries of the European Union", *Procedia – Social and Behavioral Sciences*, vol. 28, pp. 328-332, 2011.
- [3] [http://europa.eu/rapid/press-release\\_IP-13-324\\_en.htm](http://europa.eu/rapid/press-release_IP-13-324_en.htm).
- [4] T.W. van der Schaaf, "Near miss reporting in the chemical process industry: An overview", *Microelectronics Reliability*, vol. 35, no. 9-10, pp. 1233-1243, 1995.
- [5] T.W. van der Schaaf, "PRISMA-Medical: a brief description", Eindhoven University of Technology, Faculty of Technology Management, Patient Safety Systems, 2005.
- [6] S. Rodrigues, P. Brandão, L. Nelas, J. Neves, V. Alves, "A logic programming approach to medical errors in imaging," *International Journal of Medical Informatics*, vol. 80, no. 9, pp. 669-679, 2011.
- [7] F. Sheridan, "A Survey of techniques for inference under uncertainty" *Artificial Intelligent Review*, vol. 5, no. 1, pp. 89-119, 1991.
- [8] M.L. Ginsberg, *Readings in Non-monotonic Reasoning*, Morgan Kaufman Publishers Inc, USA, 1987.
- [9] U. Hustadt, "Do we need the closed-world assumption in knowledge representation?", In: Baader, Buchheit, Nutt (Eds.), *KI'94 Workshop*, Saarbrücken, Germany, 1994.
- [10] J. Neves, "A logic interpreter to handle time and negation in logic data bases", *Proc. of the ACM'84 on the fifth generation challenge*, pp. 50-54, 1984.
- [11] I. Bratko, *Prolog Programming for Artificial Intelligence*, Pearson Education Canada, Canada, 2011.