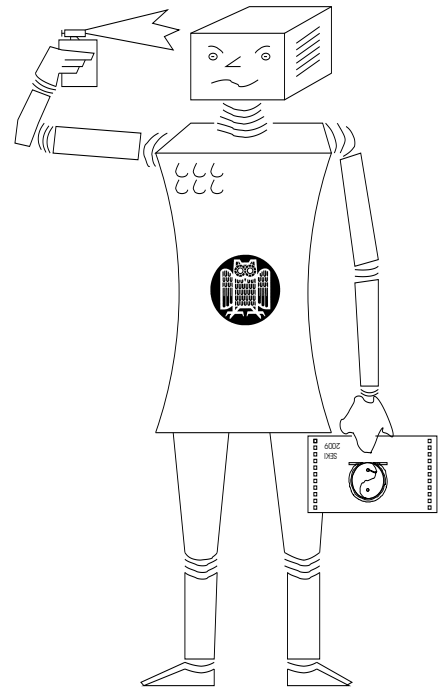


**SEKI**-Working-Paper ISSN 1860-5931

UNIVERSITÄT DES SAARLANDES  
FACHRICHTUNG INFORMATIK  
D-66123 SAARBRÜCKEN  
GERMANY  
www: <http://www.ags.uni-sb.de/>

**On Requirements for Programming Exercises  
from an e-Learning Perspective**

Carlos Loría-Sáenz  
ArtinSoft ([www.artinsoft.com](http://www.artinsoft.com))  
SEKI-Working-Paper SWP-2008-01



**This SEKI-Working-Paper was internally reviewed by:**

Jörg Siekmann  
FR Informatik, Universität des Saarlandes, D-66123 Saarbrücken, Germany  
E-mail: [siekmann@ags.uni-sb.de](mailto:siekmann@ags.uni-sb.de)  
WWW: <http://www-ags.dfki.uni-sb.de>

**Editor of SEKI series:**

Claus-Peter Wirth  
Brandenburger Str. 42, D-65582 Diez, Germany  
E-mail: [wirth@logic.at](mailto:wirth@logic.at)  
WWW: <http://www.ags.uni-sb.de/~cp>

# On Requirements for Programming Exercises from an e-Learning Perspective

Carlos Loría-Sáenz  
ArtinSoft ([www.artinsoft.com](http://www.artinsoft.com))

Submitted July 1, 2008  
Searchable Online Edition July 3, 2008

## Abstract

This report accounts for some of the main results obtained by the author during a short research visit at the Saarland University, in Germany.<sup>1</sup> In this work, we deal with the question of modeling programming exercises for novices pointing to an e-Learning scenario. Our purpose is to identify basic requirements, raise some key questions and propose potential answers from a conceptual perspective. We visualize and start delimitating a potential research area, one serving as a basis for further development and more specific work in the future. Presented as a general picture, we hypothetically situate our work in a general context where e-Learning instructional material needs to be adapted to form part of an introductory Computer Science (CS) e-Learning course at the CS1-level. Meant is a potential course which aims at improving novices skills and knowledge on the essentials of programming by using e-Learning based approaches in connection (at least conceptually) with a general *host* framework like ActiveMath<sup>2</sup>. Our elaboration covers contextual and, particularly, cognitive elements preparing the terrain for eventual research stages in a derived project, as indicated. Fully aware that this area is huge and multi- and interdisciplinary, we concentrate our main efforts on reasoning mechanisms about exercise complexity that can eventually offer tool support for the task of exercise authoring. We base our requirements analysis on our own perception of the exercise subsystem provided by ActiveMath especially within the domain reasoner area. However, we use other complementary sources, too. We enrich the analysis by bringing to the discussion several relevant contextual elements from the CS1 courses, its definition and implementation. Concerning cognitive models and exercises, we build upon the principles of Bloom's Taxonomy as a relatively standardized basis and use them as a framework for study and analysis of complexity in basic programming exercises. Our analysis includes requirements for the domain reasoner which are necessary for the exercise analysis. We propose for such a purpose a three-layered conceptual model considering exercise evaluation, programming and metaprogramming.

*Keywords:* Programming exercises, CS1, e-Learning, ActiveMath, Bloom's Taxonomy

---

<sup>1</sup>Supported by the DAAD ([www.daad.org](http://www.daad.org)), the WG of Prof. Dr. Jörg Siekmann at Saarland University ([www-ags.dfki.uni-sb.de](http://www-ags.dfki.uni-sb.de)) and ArtinSoft ([www.artinsoft.com](http://www.artinsoft.com))

<sup>2</sup>[www.activemath.org](http://www.activemath.org)

# 1 Introduction

We are interested in the general problem of how to take advantage of an e-Learning framework (like [36]) as a target for adequately coding programming exercises of the kind we might encounter at a traditional CS1 course level ([1]). In such a situation, a core problem to deal with is how *static* learning programming material (books, exercises, examples, etc.), whose *dynamics* (feedback, diagnosis, guidance, etc.) mainly takes place at classroom or labs or through personal interactions, can be implemented as active e-Learning content. In such a way that an important part of the dynamics can become an active component of the corresponding e-Learning version. At this work, and from a very particular point of view, we want to contribute to the understanding of some parts of the problem by identifying requirements and proposing solution paths at a concrete level.

In addition, we also consider important to account for reasons and issues at the contextual level because they may have an impact on our understanding of the problem helping us to reflect upon scope, relevance, feasibility and potential benefits of alternative approaches to solution paths. And even though our work is limited we want to situate it within a practical context and make explicit where it pretends to fit in and where it does not.

Let us begin saying that it is not difficult to become attracted by the potential benefits that e-Learning might offer from the point of view of curricula, in general. We may think about things like extensive and massive availability of educational material via the Web and rich electronic media, active course administration, maintenance and application among many other evident advantages which extend the scope of a more traditional classroom oriented model. Hence, given a particularly effective approach to teach programming, it seems quite reasonable to try to deliver it with the help of some e-Learning strategy, expecting this way to profit from its benefits at some appreciable degree. To define an adequate strategy for delivering programming is evidently an important and sensitive first stage to take into consideration.

When establishing such an objective at the CS1 level, indirectly and unavoidably, we come across the well-known debate around what is *the most appropriate way* to teach programming for novices at Computer Science (CS) and Information Technology (IT) related disciplines, where programming is considered in their curricula an essential skill ([1]). For answering such a question a lot of conceptions, models and tools have been presented in the past and the discussion still remains open nowadays (for a small subset where different perspectives and issues are discussed see [1, 32, 26, 18, 21, 35, 37, 38, 39]).

Issues in the dispute cover the proper abstract nature of the subject, the definition of minimal programming skills, the programming paradigm, the programming language, the convenience or not of using tools, the approach to motivation and course situation among many others. Such a wide spectrum and space for divergence shows us that the subject CS1 turns out difficult to normalize even though it is technologically, scientifically and socially quite relevant, given the CS/IT impact on many societies.

At least, there seems to exist consensus on some of the main issues associated with former and current teaching approaches. For instance, in terms of:

- Expected levels of quality from an academic perspective
- IT labor market demand and technical requirements
- Student individual motivation for taking a computing discipline as a career (student attraction and retention)

These **dimensions** of learning trigger forces which might naturally compete with each other so the particular teaching approach tends to mirror that in one way or another. Therefore, we should not completely ignore them in our analysis.

In fact, contextual considerations of that kind indicate we have enough good reasons for innovation through e-Learning techniques at any, at some or even all the mentioned dimensions. We believe that any e-Learning approach concerning some key facet of the learning how to program (especially one considered so crucial like programming exercises) should not be designed ignoring some issues related with these forces. Hence, we would consider rather important to try to understand in more detail some of the key technical issues within the debate (for instance, the question of *the right* programming paradigm and the transition CS1-CS2 under the *functions-first* model, [1]). To a certain extent, we have tried to incorporate them, at least indirectly, in the background of our analysis.

However, and given the limited scope of this research, we decided to focus on cognitive models of programming in a way that can help us understand more formally why exercise are considered '*hard*' for novices (we use [5, 15, 25, 30, 34, 32, 43, 49] and [23] as a general and standard reference). We address the question within the frame of e-Learning.

We do not consider socio-motivational ([18, 21]) or market related concerns ([37]), and in order to understand a relative small part of the entire problem, we do it in the context of a particular CS1 instructional material just as a case study. Through this input, we try to raise some questions on how such *complexity* is reflected and can be recognized on exercise instances by using a hopefully cognitively adequate and sound set of reasoning criteria.

For such a purpose, we mainly follow approaches based on or quite related to Bloom's Taxonomy ([2]) because it is a simple yet powerful framework which has been applied in CS1 related questions with apparently quite interesting results ([27, 33, 32, 18, 45]).

We interpret such experiences found in the literature as a sound basis to understand the complexity of some class of exercises as a part of an e-Learning environment. Through our work, we identify interesting questions concerning representation and formal reasoning and contribute with conceptual proposals to deal with them. We consider the ActiveMath model of exercise ([17, 50]) as a point of reference to derive from there our requirements on representation and manipulation languages in terms of a potential implementation.

Consider the following code fragment:

```
int[] x1 = {1, 2, 4, 7};
int[] x2 = {1, 2, 5, 7};
int i1 = x1.length-1;
int i2 = x2.length-1;
int count = 0;

while ((i1>0) && (i2>0)){
    if (x1[i1] == x2[i2]) {
        ++count;
        --i1;
        --i2;
    }
    else {
        if (x1[i1] < x2[i2])
            --i2;
        else
            --i1;
    }
}
```

After the above while loop finishes, 'count' contains what value?

- a) 3
- b) 2
- c) 1
- d) 0

Figure 1: Java Version of Leeds Question 2 taken from [34]

## 2 An Illustrating Example

In order to illustrate some of the questions around exercise complexity, we present an example taken from [34] as shown in Figure 1. The example is referred to as the Java version of Question 2. It was used in the "Leeds"-Group study as explained by the authors in the context of testing program reading and understanding capabilities. Some studies in the past have revealed weaknesses in programmers on relation with these basic skills ([35]). The example constitutes a nice case of reasoning about exercise complexity in the way we are interested to formalize.

According to the cited work the steps indicated in Figure 2 represent a reasonable solution plan that can be expected to be followed by some responder. The answer a) is the result of following this plan.

1. Read through the code
2. Infer that code counts the number of common elements in the two sorted arrays
3. Count manually that number in the two given arrays, namely 3
4. Conclude that option a) is correct

Figure 2: Plausible Solution Plan to Question in Figure 1

We notice that counting common elements in sorted arrays is basically the intended function which is implemented by the `JAVA` code. As we can judge by seeing the plan of Figure 2, to be able to reach such kind of inference is recognized as an *advanced* skill for programmers at the level of a CS1 course. Apparently, such a skill is here being exercised.

However, the question hides a little trick, a further difficulty which might be easy to overlook: the array indexes `i1` and `i2` do not reach the value 0 which in `JAVA` is the first index of an array, as known. So the correct answer is actually b). According to the authors 65% answered correctly, but 21% chose a). This is an extremely interesting empirical result which is further analyzed in detail by the source using the SOLO Taxonomy, which is worthy to read but is currently beyond our scope.

Let us recapitulate some key ideas from the example. The solution procedure shown above would match the solving-plan that a relative experienced programmer would apply by means of concept/role assignment techniques in order to recover the intended function (see [5, 30]). In contrast, another less experienced responder would probably use a strategy of a lower cognitive level, essentially by executing the fragment and obtaining the right answer more or less mechanically ([48]). Probably, without having *grasped* the intended function, so without needing to understand the code at the abstract level. Thus, the less experienced programmer would get a better performance than the more experienced one, who fell into the trap of the trick.

The question is which level of complexity is mainly being exercised: recovering the intention or applying the concrete operational semantics of arrays and loops. If recovering was the goal, maybe 65% of correct answers is not properly reflecting it. If both area were aimed at by exercise, we learn that adding such a small piece of complexity might yield a potentially unexpected result. Tricks like these seen here are known as *distractors*. They turn out useful in exposing some misconceptions or uncovering incomplete learning. They should be consistent with the complexity of the test, however.

Perhaps the source of such a potential inconsistency, if there is any, could be recognized during the authoring phase of an exercise. Popular cognitive models, like Bloom's Taxonomy, and their interpretations in the programming field, are useful frameworks that can help instructors and teachers to reason about that kind of issue. Hence, in a CS1 or similar e-Learning context, it would certainly be interesting that the computational representation of exercises would allow semiautomated reasoning at that level. The situation we have in mind would be one where some sort of analysis of exercises could be performed in a *static* way during the authoring stage; eventually warning for potential *cognitive inconsistencies*, as in example in Figure 1.

At this point and just briefly, we like to relate our goal of checking consistency with the approach of the so-called *buggy rules* ([50]). As we will explain later on, our goal to have a reasoning procedure about exercises can be seen as a technique that would allow to suggest whether something like buggy rules might be required at some point in a given exercise specification. Another related case employing something similar to the notion of buggy rules is [40]. In both cases, we realize that a key to take into account in our analysis is a domain reasoner on programming exercises. Such a reasoner should allow us to express exercise playing as in [50] but also additional requirements like reasoning on the exercise itself would then be possible. For instance, in terms of exercise complexity, design and reuse, as we explain in more detail later on.

### 3 Structure of the Rest of the Paper

The rest of this report is organized as follows. In Section 4 we present the background on which we base the study. In general, the idea is to review the material looking for requirements according to the scope and goals of the work. We cover definitions and models for CS1 in order to obtain contextually relevant feedback. We also analyze e-Learning and the features and facilities of ActiveMath which are most related with the exercise subsystem. We pay special attention to the idea of a domain reasoner from a conceptual perspective. Then, we address Bloom's Taxonomy and focus on its computational interpretation as a framework for semantic based informal reasoning. We present some cases of study where the taxonomy has been applied for CS1 courses. In Section 6 we recapitulate the identified requirements and reformulate them as a kind of *agent interaction system*, in a simplified but useful way. This allows us to derive a conceptualization by means of a three-layered reasoning model. We then show how this architectural model could be used to represent exercise operations and its relation to the requirements of a domain reasoner adequate for programming. Finally, in Section 7 we illustrate how our model can be used for reasoning about complexity and also we summarize some open problems for further research. And finally, in Section 8 we summarize and give conclusions pointing to future work.

## 4 Background

We go through the main material we use in this work pointing to its relevance with respect to our goals, especially looking for requirements concerning e-Learning. Our intention is that this report could serve as a source for further research. In the following, we focus on programming-in-the-small for learning purposes, additional forms of expressing programs ([39]) are briefly mentioned but really not considered here.

### 4.1 CS1 View under ACM/IEEE

We take a very brief look at the ACM/IEEE Computing Curricula 2001 (CC2001 for short; [1]). It serves here simply as a particular sample model for realizing what a CS1 course might thematically cover and we employ it as a well-known and widely referred standard where some relevant aspects deserve special attention. Given the scope of this work we just account for two of them in order to identify potential research opportunities for e-Learning.

#### 4.1.1 Multiple Profiles and Expected Performance

We first recall that CC2001 considers 5 disciplines related to computing, namely Computer Science (CS), Computer Engineering (CE), Information Systems (IS), Information Technology (IT) and Software Engineering (SE).

Such a division turns out relevant in our context because CC2001 conceives different profiles with respect to expected performance on programming capabilities, even though CC2001 assumes a *common core* for the first year and values programming as an essential skill.

To see that, let us take a look at Table 1 which is an small excerpt from Table 3.3 of [1] just showing the portion named 'Algorithms'. We notice that columns CS and IS are mostly opposed to each other, according to this model. Moreover, a uniform and interesting difference between CS and SE is clearly expressed, too. Although these are models of expectation on competences from a curricular point of view, they actually may coincide with students expectations and professional aspirations.

In fact, due to organizational aspects in some CS Departments, it happens that students from different computing disciplines attend the same course and thus get exposed to the same programme at the first-year level<sup>3</sup>. It is interesting to remember that known studies like [35] showed a bimodal<sup>4</sup> distribution of the student performance in programming evidencing the lack of normality on assimilation. Thus, first year course implementations may need to cope with a similar situation, which means to pay careful attention to assessment and grading systems, and consequently to exercise models, as a consequence (as done in [33, 32] cognitively sustained on Bloom's Taxonomy).

We identify another opportunity to employ e-Learning techniques to help students remain more uniform as a group. Using appropriate exercises in those areas where cognitive differences could negatively influence individual performance. This may avoid, in many cases, the loss of motivation and attrition problems at an early stage. We believe e-Learning at CS1 level can be

---

<sup>3</sup>In our personal experience, SE, IS and CS. In student proportions that agree with data [26], in our opinion.

<sup>4</sup>A big group of *weak* students and a small group of *strong* students.

aligned with a *funnel* strategy ([1]) which is probably more adequate than a *filter* strategy in case of existing different student orientations toward computing disciplines or when an orientation is not yet mature enough at this level<sup>5</sup>.

Performance Capability	CE	CS	IS	IT	SE
Prove theoretical results	3	5	1	0	3
Develop solutions to programming problems	3	5	1	1	3
Develop proof of concepts programs	3	5	3	1	3
Determine if faster solution possible	3	5	1	1	3

Table 1: Expected Performance Capabilities on Algorithms according to Discipline (Scale: 0=no expectation-5=highest expectation)

#### 4.1.2 Multiple Paradigms

With respect to the theme of programming paradigms CC2001 proposes 6 different course models in the first year depending on the paradigm choice: the so-called procedural-first object-first, functions-first, bread-first, algorithms-first, hardware-first strategies. Among such a variety, a valid choice<sup>6</sup> is functions-first, where the standard suggests two courses, CS1 using functional programming (FP) to be followed by CS2 using object-oriented programming (OOP).

Such a transition is interesting, no special indications are given by CC2001. An e-Learning bridge FP  $\rightarrow$  OOP appears interesting to implement in a CS1-CS2 context. We are aware that the associated problems have been extensively covered in the past as OOP became a widely-spread programming standard at the software industry. However, some new forces can be able to reactivate it. On the one hand, OOP industrial languages (Java and C# mainly) have been adopting known FP features like generics (parameterized polymorphic typing) and lambda-expressions (delegates, continuations) and in some cases interesting academic languages exist where pattern-matching is naturally incorporated ([10]). As a result of these trends we can figure out e-Learning techniques helping with these multiparadigmatic trends matter, too.

## 4.2 On e-Learning Elements Notions

Based on [47], we see that e-Learning covers those forms of learning in which instructor, learning material and student might be separated by space or time where the gap between parts is bridged through the use of a wide spectrum of online technologies, mainly Web-based ones, and a planned and methodic teaching/learning experience. Then, e-Learning systems serve to complete or complement more traditional forms of learning in environments which are more appropriated to particular individual conditions, according to [36].

Precisely, this last work provides us with a rich contextual point of reference where, as a general motivation, we are interested in creating content for a CS1-like course and specifically

<sup>5</sup>In [18], for instance, more emphasis is paid to psychological and motivational involved aspects by means of active learning likewise under a Bloom's Taxonomy sustained methodology.

<sup>6</sup>Which is well known in our personal experience using Scheme and Java.

Consider the following program:

```
val x = 7 + 4
val y = x*(x-1)
val z = ~x*(y-2)
```

- a) Which identifiers, constants, operators and keywords occur?
- b) Which value is bound to the corresponding identifier?

Figure 3: Version of Exercise 1.1 in [42]

developing exercises to support feedback and assessment insight an e-Learning environment. Consequently, a key issue is that instructional material usually needs to be complemented and implemented in such a way that guidance, coaching and feedback, normally provided by the instructor in person, now need to be part of the learning interaction strategy and the content available in the learning environment. Let us consider this a little bit more in detail with attention to exercises.

#### 4.2.1 Elements of Exercise Model

Source content to be adapted as e-Learning material is usually taken from some proven and standard course material like the main book, labs projects and similar sources like exercise banks. Such a task can involve a lot of detailed and careful work. For example, the methodology of the source material might need to be adapted, as already mentioned, it also requires a formalization of the domain model using some formal language and similar tasks.

In particular, with respect to exercises, we may need to design and implement a dialog-based orchestration using GUI-gestures and corresponding events and associated feedback, hints responses, etc. Such kind of elements need to be assembled in an adequate way. This is naturally an extremely simplified view, we refer to <http://www.examat.org> and to [17] for appreciating specific details and available tools that aid the design process by hand.

Object representation in ActiveMath relies on OMDoc which is specifically designed for the mathematical domain ([29]). Currently there is a version for representing programs in the same direction: [28]. We will be referring to the issue as a requirement, later on.

An exercise is essentially modeled by means of a finite state automaton (FSA), according to [17]. States in FSA represent interaction stages during exercise playing and edges refer to event-based transitions associated with possible user answers to GUI-elements, those offered by the corresponding state. An example of such an event occurs when the user selects an option from list representing a multiple choice question (MCQ), which is an interaction element that is available in the design tool.

As a simple illustration, let us consider the code snippet shown in Figure 3. It is taken<sup>7</sup> from the book [42] which is currently used in an introductory course to programming at the University of Saarland. The book uses functional programming and Standard ML language.

Assume such an exercise would need to be reexpressed in the exercise model of the ActiveMath environment. Requirements are clear: Example of Figure 3 tests some previously introduced ML basic concepts like 'program', 'value', 'identifier', 'operator', etc. (that is concepts

<sup>7</sup>The original is written in German, the presented version is our free and slightly changed translation.

near to the lower syntax level). We also notice that some initial notions are required proper of the ML evaluation strategy: <sup>8</sup> basic operational semantics like statement sequence (composition, ordering) and definition and binding of identifiers. Those concepts and corresponding relations obviously need to be part of a formal domain model so that the student interacting with the exercise is able to consult (recall) and reinforcement them. Let us assume such concepts are available.

The exercise implementation is a matter of design: we may decide using a MCQ in each case a) and b) and/or textboxes to gather the input of the required answers, among others interaction and presentation elements. FSA stages need to be designed correspondingly. Thus, for instance, in case b) a state can be related to the expression `val z = x*(y-2)` asking for the value of both `z` and `y`; this can done in order to test the misconception of reading a variable <sup>9</sup>. As can be easily realized, many choices are possible, even if we are dealing with a relatively small exercise, in appearance. A task like this one needs to be performed and evaluated following some parameters and criteria concerning adequacy, alignment and quality. We are interested in studying some of such criteria but not directly addressing the design of the interaction/presentation level.

#### 4.2.2 Static Exercise Complexity

We have seen that an exercise can be implemented as a kind of interaction system whose design and implementation is actually driven by a problem as its 'leitmotif'. Thus, before or as a part of such an implementation, but at a more abstract level, we might be interested in asking how can we analyze the inherent complexity of the problem but in terms of a general cognitive model, (i.e. independently of the UI-interaction-model). In other words, our question is:

How can we represent an exercise so that, as much as possible, it can be *statically analyzed* in terms of its complexity? The complexity should be derived from its problem specification and surrounding domain model.

How complex an exercise can be is something that naturally depends, in the end, on the person that will be trying to solve it as well as on some environmental conditions. From that point of view, complexity is relative and clearly a *dynamic* value. In fact, it is a function of the student model component to help approximate and diagnose a personal profile as a result of his/her interaction with the problem. However, exercises have to be correctly *aligned* with *prototypical* student profiles, in order to be effective ([2]). So we may want to judge complexity relative to some intentionally targeted static profile, under some ideally static conditions.

For instance and using once again Example 3, we may intuitively reason and anticipate that question a) could be considered simpler than b); because, for a prototypical student, identifiers and operators notions relate very well to variables and basic arithmetic operations (up to the use of '~' in ML instead of a prefixed minus). But data-binding and control as well as sequential data-flow can be something new or not yet assimilated by such a student. See the interesting experiments evidencing some misconceptions that novices might initially develop with respect to binding (variable assignment, [30]) <sup>10</sup>.

<sup>8</sup>For instance, assuming the notion of a typical read-eval-print loop based interpreter has been introduced.

<sup>9</sup>Some students tend to think after a variable is assigned it might loose its value ([30]).

<sup>10</sup>We personally observed following case: in a similar context, using an eager language, we asked a student to evaluate something like `x*(x-1)` step by step after a definition of `x`, let's say, `val x = 7+4`. We remember he started replacing `x` by `7+4`. That is, the first step was like: `(7+4)*((7+4)-1)`. Such kind of harmless

Thus, because part b) of the exercise aims at the basics of data-binding and data-flow and this task can be considered cognitively *harder* than part a), we might take that knowledge into account during the design phase and organize the interactions accordingly, maybe as two independent exercises.

Thus, from a pragmatic perspective, we believe complexity could be relevant in a practical situation where relatively many given exercises might need to be carefully adapted to fit in an e-Learning environment.

### 4.2.3 Dynamic Generation of Exercises

So far, we have seen a static perspective of exercise generation as done in ActiveMath. This view corresponds to the (pre)authoring facility when it is performed manually. However, as described in [50], ActiveMath also provides a more powerful functionality to dynamically generate exercises using a domain oriented representation instead of a dialog-oriented one. (In [50] the domain is symbolic differentiation calculus).

Because such a facility represents a challenging issue in our particular domain of interest, we come back to this issue later on in this work. At this point we just notice that specific domain reasoning rules for symbolic manipulation of small programs could be required in order to provide a similar capability as in calculus domain.

Just as a simple example, consider case b) of Example 3. In this case, and certainly depending on the *intelligence* of the exercise generation procedure, we might need rules for expressing data-flow related operations; just as differentiation rules are used in [50].

Finally, we want to mention that another similar approach to the automatic generation of exercises is described in [24]. They report deriving exercises directly from the domain model, which in that case is the Databases (DB) and SQL. They use a *random-walk* on the DB-Schema representation, exercises are in this case DB queries.

### 4.2.4 Domain Specific Reasoning Rules

As described in [50] the exercise module in ActiveMath allows dynamic the generation of exercises where interaction can be added dynamically based on the diagnosis of student answers to the exercise task. Central to this feature is a domain reasoner (DR) working on the exercise task. The domain reasoner behaves like a (rule-based) expert system that provides intermediate solution steps to the task, in order to represent the solution like a plan, actually in the form of a graph. Nodes represent derived subtasks according to the solution proposed by the DR. Using that information, the exercise module is able to indicate which specific rules the student can apply or compare his/her answers against the expected ones with respect to the solution graph. In addition, buggy-rules can be also included which serve to encode possible common user misconceptions and errors for the specific task. For instance, consider the chain rule for derivation:

$$\frac{d[f(g(x))]}{dx} \Rightarrow_{\{expert, chainrule\}} \frac{d[f(z)]}{dz}(g(x)) \times \frac{d[g(x)]}{dx} \quad (1)$$

---

'misconception' (from the programming language perspective) should be recognized early before it gets accumulated and becomes potentially harmful.

The structure of this domain specific rule lets us formulate a recursive solution plan for a task matching  $\frac{d[f(g(x))]}{dx}$  which includes solving both intermediate tasks  $\frac{d[f(z)]}{dz}(g(x))$  (outer layer) and  $\frac{d[g(x)]}{dx}$  (inner layer).

For instance,  $\frac{d[\log(\sin(x^3))]}{dx}$  is a matching task that evidently reduces using Rule (1) to  $\frac{d[\log(z)]}{dz}(\sin(x^3))$  and  $\frac{d[\sin(x^3)]}{dx}$  as recursive subtasks.

Additionally, incorrect (buggy) rules can be known and applied by the expert reasoner to identify potential wrong steps like the following when the inner derivation step  $\frac{d[g(x)]}{dx}$  is missing:

$$\frac{d[f(g(x))]}{dx} \Rightarrow_{\{\text{buggy, chainrule, inner\_layer}\}} \frac{d[f(z)]}{dz}(g(x)) \quad (2)$$

In the example below, such a buggy rule would account for explaining some wrong answer from a student, for example like  $\frac{1}{\sin(x^3)}$  as an answer for  $\frac{d[\log(\sin(x^3))]}{dx}$ .

That kind of rules allows the response module to report a missing inner layer as a possible cause of the wrong or incomplete answer.

We also notice that a DR also needs to make use a at lower-level of other tools for symbolic manipulation like a computer algebra system (CAS). For example, to compare symbolic expressions not only syntactically but also using their normal forms. In the same direction, a matching operation associated with a rule selection can be more involving, because in case some previous symbolic transformation is required to allow a syntactic matching and rule application. Think of AC-matching for instance.

Robinet (et al) reports in [40] a dynamic model to identify so-called *intermediate metal steps* when solving linear equations and inequations like for instance  $2x + 9 = 8 + 6x$  (the target e-Learning system is APLUSIX). In this case, domain rules include moving monomials from one side of equation to the other and algebraic calculations. Notice that rule selection is don't care non-deterministic. If a student would wrongly response, let us say, with  $8x = 17$  then the model attempts to find the most probable path (a combinations of movements and calculations) explaining the error. Such a path is then used for tuning a student model and diagnosis (for instance, to detect whether student has more problems with movements or with calculations or both).

## 5 On Bloom's Taxonomy

We briefly review Bloom's Taxonomy as a standard framework for analyzing educational objectives and this turns to be useful for answering questions concerning exercise complexity, in particular. We restrict the presentation to the cognitive taxonomy. We will use the revised version as presented in [2], which is two-dimensional. We believe that it is adequate in our context where procedural programming skills and knowledge are both relevant. We cover the material using similar notions as in conceptual modeling and semantic based reasoning ([7]). The reason is that the taxonomy could be conceived as a procedure for formal reasoning on semantic complexity of exercises.

### 5.1 Basic Notions

Expressed in our biased computational and simplified perspective, Bloom's Taxonomy is basically an empirical reasoning framework. It deals with the classification of statements representing objectives, goals and tasks we want someone to achieve and accomplish. Classification is two-dimensional. In principle, it can be used with statements referring to any domain. It is strongly human-oriented, statements are expressed in natural language and consequently ambiguity can be a problem, as one might expect <sup>11</sup>. Once educational statements get classified, expert based recommendations (a kind of expert rules) can be used to associate the classification with recommendations about best practices in terms of organizing learning priorities, instruction delivery, assessment and alignment, roughly speaking.

The classification procedure is described as a matrix (See Table 2 for an illustration). Rows represent the *knowledge dimension* with four categories: *Factual*, *Conceptual*, *Procedural* and *Metacognitive*. Columns represent the *cognitive process dimension* where six categories are considered: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate* and *Create*. In both cases, a dimension represents a continuum of complexity. Categories in a dimension identify specific ordered points over this continuum. For instance, *Apply* is more complex than *Remember* as a cognitive process. And *Conceptual* is more complex than *Factual* as knowledge. Over this last dimension, complexity usually contrasts concreteness against abstractness, with *Factual* being more concrete than the others.

Knowledge	Cognitive Process					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual	List primitive data types in a language					
Conceptual				Decompose a structured concept in its parts		
Procedural		How to implement a sort algorithm				
Metacognitive					Criticize learning programming methodology	

Table 2: Taxonomy Table (includes hypothetical examples)

In order to classify a statement, the procedure first requires us to *normalize* it by putting it in

<sup>11</sup>For instance, see [27] for a study evidencing discrepancies in categorizing questions in the specific case of CS. Similar findings are reported in [45]. In both cases, however, it is suggested that discrepancies decrease as assumed knowledge of students get clarified among those persons who are categorizing questions.

the form of a pair  $V \times NP$  for some verb  $V$  and noun-phrase  $NP$ . Using so-called *clues*, which are association rules,  $V \times NP$  can be classified as  $P \times K$  for some cognitive process  $P$  (mapping  $V$ ) and knowledge category  $K$  (mapping  $NP$ ). The expression  $P \times K$  actually denotes a matrix cell, useful teaching hints and practical strategies, among others, are associated with it. Gathering them is main the purpose of the classification, actually.

To show a simple artificial example, the statement '*To be able to distinguish between an interpreter and a compiler*' could be normalized as  $distinguish \times translator$ , where *distinguish* is the main verb and  $interpreter \sqcap compiler$  (a composed)  $NP$ . Let us assume, it is a subconcept of concept  $translator$ , with respect to some CS knowledge-based domain. According to a well known clue of the framework, *distinguish* is an action whose closest related cognitive process is *Analyze*.

By means of this association, our original sentence can be expressed in the framework's vocabulary as  $Analyze \times ConceptualKnowledge$ . In terms of complexity, we might infer that for accomplishing such an objective more than just memorizing a plain list of facts is involved: because knowledge contemplates several concepts and their relationships. In addition, analytical thinking skills are also required which are more complex than remembering. Hence, a particular technique to help learning of that kind of knowledge and skill should be used, consequently.

We have used freely a simple symbolic notation in this example just to suggest that a formalization of the procedure is, at least in principle, quite possible and technically interesting, especially when we are trying to find requirements for a tool. So, we bring to attention that the semantic based inference procedure could be automated to some extent to support the methodology. But naturally the essential question remains:

How can CS1 exercise concepts be adequately aligned with to Bloom's Taxonomy of verbs, nouns and action rules? What are the right concepts?

Such a question, and posed this way, is too abstract, indeed. Hence, we want to consider and propose some ideas that we hope, serves to refine it and decompose it into more concrete questions and related issues as well as to identify paths to cope with them.

We summarize the general principles fundamenting Bloom's Taxonomy, expressed according to our particular understanding and objectives:

1. Dimensions of learning objectives are knowledge and processes. They are domain-independent.
2. Those dimensions can be (linearly) arranged according to categories representing increasingly levels of cognitive complexity
3. In order to *reach* one particular category, the inferior category must have been *reached* first
4. Domain-specific cognitive dimensions can always be semantically related with domain-independent cognitive dimensions. Thus, domain-specific learning objectives satisfy the same principles of cognitive complexity.

### 5.1.1 The Need for Non Standard Reasoning

As we already mentioned, sometimes it turns out difficult to make non ambiguous decisions on some particular cases in absence of specific knowledge about the learner as mentioned in the example in [45]. The question makes sense in a dynamic situation, like the following question implies: how to dynamically judge the complexity of an exercise, statically assumed by its author to assess a particular classification, let us say,  $P \times K$ , if the applier (the student) does not have yet reached level  $K$  at the knowledge dimension. What can we conclude in such a case?

Such a kind of question is evidently relevant in the e-Learning context, for instance, preparing an appropriated response for an exercise interaction, as we had mentioned before. For such a case, we might hypothesize some explicit *rules* to be used and added to Bloom's framework, which we call non-standard. For instance, like the following claim:

If student exhibits knowledge on a level less than  $K$  and the exercise statically assumes  $P \times K$  then the dynamic complexity becomes  $Create \times K$ . In other words, the exercise requires students to create knowledge of level  $K$  in order to perform  $P$ .

It is interesting that such a rule can be geometrically interpreted as moving the complexity to the last cell of the row  $K$ , which is the one with the highest skill requirements.

Evidently, it might be debatable whether such a kind of reasoning is empirically valid or not. The point we want to make is that a Bloom's based methodology would need to contemplate similar forms of reasoning when used in a dynamic way as we may have modeling interactions between exercises and students. We identify this as an important open question.

## 5.2 Bloom's Taxonomy and CS1

As a case of study, we want to review the way [32, 33] uses Bloom's Taxonomy in CS1. In this case, it is an object-first course using `Java` and where no e-Learning is directly considered. Basically, the main purpose is to define more realistically the course objectives and from there to implement a more aligned grading system<sup>12</sup>. The author explicitly avoids performing a detailed explicit one-by-one mapping of his statements to the categories of the taxonomy. Instead, he proposes three general complexity categories which are considered adequate to his CS1 vision and empirical considerations. The proposal is shown in Table 3<sup>13</sup>. The objectives of CS1 under

Process Level	Bloom's Levels
Reading and Understanding	<i>Remember</i> and <i>Understanding</i>
Writing small code fragments	<i>Apply</i> and <i>Analyze</i>
Writing non-trivial programs	<i>Evaluate</i> and <i>Create</i>

Table 3: CS1 Cognitive Process Model according to [32]

this model are defined to address the first two categories, only. Thus, we realize that this model

<sup>12</sup>The original version of the taxonomy is used and as a consequence only the cognitive process dimension

<sup>13</sup>Identifying Bloom's original categories with the verbs used in the revised version one-by-one

pays special attention to develop reading and understanding skills. In fact, that is sufficient for passing it, according to the grading model. The second category is meant as writing code in a well-defined context.

Consequently with this decision, we want to understand it as providing well-defined specifications referring to a domain fully understood by the learner. This should include the appropriate class structure, the signatures and finally the methods should be kept small. So that focus is set as much as possible on expressing well-known concepts at the local (method) level. We want to add, following [15] among others, that *templates* should be initially used so that the process *Apply* can be incrementally evolved into *Analyze*.

Focusing on reading and understanding offers some benefits by the elaboration of exercises because questions forces the student to solve problems without being able to run code on a machine. The author strongly exploits this by using Multiple Choice Questions (MCQ) as a testing technique. For instance, example in Figure 5.2 in [31] exercises on iterations over arrays.

The exercise design intentionally contains distractors to test misconceptions, as the author explains. In contrast with the exercise shown in Figure 1 the intended function is specified, however. Interesting questions remain how to adapt such a kind of exercise in a interactive e-Learning environment according to our requirements. We will discuss this, in the second part of this work.

```
public static int maxPos(int[] y, int first, int last){
/* Returns the position of the maximun element in the
* subsection of the array "y", starting at position "first"
* ending at position "last".
*/
    int bestSoFar = first;
*** missing code goes here ***
}
Which of the following is the missing code from "maxPos"?
(a) for(int i=last;i>first;i--)
    if(y[i] < y[bestSoFar])
        bestSoFar=i;
(b) for(int i=first+1;i<=last;i--)
    if(y[i] < y[bestSoFar])
        bestSoFar=i;
(c) for(int i=last;i>first;i--)
    if(y[i] < bestSoFar)
        bestSoFar=i;
(d) for(int i=last;i>first;i--)
    if(y[i] > y[bestSoFar])
        bestSoFar=i;
(e) for(int i=first+1;i<=last;i--)
    if(y[i] > y[bestSoFar])
        bestSoFar=i;
```

Figure 4: MCQ Exercise on Reading and Understanding in [31]

It is worth mentioning that the presented CS1 model does not directly match the knowledge

dimension of the taxonomy, which poses an interesting question, indeed. Actually, we have not found any CS1 approach dealing with this dimension, directly.

As an attempt to elaborate some ideas for this issue, we briefly consider the learning model used in [3]. It serves us as a contrast because the CS1 model is different to [32, 33], and not explicitly aligned with Bloom.

Instead, the course model is an implementation of the so-called *early-bird* pattern ([4]). We may interpret it in terms of driving the course by modeling not directly by language function, which in a SE/OOP context appears to be natural. In fact, authors explicitly aim at SE as a student target.

From the point of view of the Bloom's Taxonomy, we claim, the course is driven by the knowledge dimension, the process dimension is subordinated, in such a sense. Moreover, the knowledge dimension is covered iteratively (like a spiral model), so the course consists of several cycles covering a core set of new topics added on the top or extending previously covered material. Using models and `Bluej` apparently help to deal with programming language related details, especially at the beginning. In any case, `Java` plays a central role in this course.

By taking a look at the material in [3], we have identified a categorization as shown in Table 4. We have provided some simple examples as clues as seen in the table. It seems to us, under a

Knowledge	Bloom's Level	Example (Student ...)
Behavioral	<i>Factual and Conceptual</i>	Understands a model of geometric figures with different attributes
Implementation	<i>Procedural</i>	Understands overloading of constructors in Java
Enhancement	<i>Metacognitive</i>	Recognize redundancy as a maintenance problem

Table 4: CS1 Proposed Knowledge Model derived from [3]

model-based approach the *Factual* and *Conceptual* knowledge levels naturally merge building a uniform category. Every cycle starts with a specification of a conceptual model which the student has to understand and *play* with before learn how to implement it (procedurally)<sup>14</sup>. Initially the specification is UML-like, later on can be more informal allowing learning of basic design principles as well. In terms of exercising, we find difficult to describe interactions because of the graphical nature of specifications, we are more interested in textually expressed code. However, [3] provides a lot of exercises at the *Reading* and *Understanding* level.

Considering the cognitive level of *Writing small code fragments*, we want to present an exercise which is (mini)pattern-oriented and likewise suitable for exercising refactoring. It also requires exercising some basic understanding. It is shown in Figure 5<sup>15</sup>. As we did before, we raise the question about the representation in terms of e-Learning. Although, the example is quite

<sup>14</sup>The tool `BlueJ` supports consistently such *playing* task.

<sup>15</sup>It includes minor changes.

simple it would demand symbolic generative manipulation of source code if exercises have to be automatically generated could be reused. For instance, a method like the following:

```
public int getX(){return x;}
```

But this has to be performed independently of the name `x` and its type `int`.

Add setter and getter methods for the field `x` in the following class declaration:

```
public class A{  
    private int x;  
}
```

Figure 5: Exercise on Small Code Generation based on [3]

Finally, it is important to account for the model described in [11, 12]. Such a project would pioneer the teaching of Scheme to novices and an adaption of an environment to take beginners learning needs into consideration. As course model, and from a Bloom's based perspective, we consider it shares a lot with [3] in goals, pattern (recipe) based methodology (very related to [15]) and particularly the central role a tools plays on incremental learning. A special interesting feature is the existence of different programming languages according to student level. We identify this a requirement in the e-Learning context, certainly.

## 6 A Conceptual Model for Exercises

In the first part of this work we have elaborated what we may call the big picture with respect to requirements on programming exercises. We did it on the basis of several different perspectives. Cognitive elements, case studies, course models and an e-Learning environment were considered. In this second part we want to conceptualize and refine those requirements bringing them to a more concrete level. By this means, we want to figure out how they can be further developed pointing to a design and potential implementation, in a continuation of this work.

By now, our goal is to develop a proposal containing general answers to the questions we have raised during the analysis. For such a purpose we present a general framework for reasoning on exercises. Although, our proposal certainly stays at the conceptual and architectural level we are convinced that the main ideas are reasonably comprehensible and straightforward such that the following steps of a design are easy to appreciate. However, an understanding of the invented requirements is surely necessary for that purpose.

### 6.1 General Scenarios and Assumptions

We review the requirements and issues we have considered so far. We mostly restrict ourselves to programming exercises and related cognitive elements. Further, we present a strong simplification of scenarios we are modeling, in such a way that we can focus on more specific issues we would like to study and validate. By doing so, we will rephrase the situation in a vocabulary that might not correspond to the correct one in a strictly technical sense. However, we believe the simplification is fair enough with respect to the sources in this context.

1. We situated the problem of exercises as a part of two scenarios where several actors (we call them agents) interact. Scenarios are agent interaction systems. We have:
  - The exercise authoring scenario
  - The exercise playing scenario
  - The learning environment.
2. As the main kind of agents we have
  - An exercise
  - An author
  - A tutor
  - A student
3. Agents operate on environments that may contain other agents responsible for achieving subsidiary tasks like communication and knowledge gathering.
4. Initially, exercises can be statically created by an author agent in an authoring environment. The also can be evolve dynamically or be generated automatically as a result of interactions in environments.

5. An exercise contains a question on a (knowledge based) subject aiming at a prototypical student. It also contains a plan to answer the question and a knowledge basis for supporting the question and its answering process. Plans can be composed of alternatives subplans, so we can assume that just one plan is necessary.
6. Student and exercise interact in a playing scenario. The student is committed to solve the exercise. That means, at the end, to accomplish the goals of the plan presented by the exercise. A student also has knowledge (beliefs) and behaves accordingly to its skills and preferences during the playing. The question contained in an exercise can be considered a query to the student beliefs.
7. The question represents a static level of cognitive complexity according to the author. The exercise should be designed to match the complexity and respond appropriately in case that the matching is not occurring as expected.
8. As a match could fail, the exercise can contain in its plan a course of actions to help the student to accomplish the plan goals. It can choose to recommend alternative paths or submit the student and a diagnosis of its performance to a tutor in a learning environment where new or remedial interactions can take place.

## 6.2 Questions and Plans

The scenarios previously described are indeed quite general, programming is not playing a particular role, yet. They are also complex, we are not able to further refine them in this work. Therefore, we will concentrate on the specifics of questions, plans and complexity. In the following, we may sometimes abuse the terminology and identify exercise with questions, even though we are assuming that questions are just components of exercises, in our conceptualization.

Our next step is to cover issues on exercises and plans which essentially means to show a proposal for their representation. Complexity will be handled as a result of modeling decisions concerning the two formers.

Let us now summarize some additional important and more fine requirements:

1. Questions in exercises can be expressed in different languages. Some of them are programming languages (PL) some are natural languages (NL) or formal languages (FL) or of other kinds. Combinations of them are possible. We call them specification languages (SL).
2. Some versions of the same kind of exercises could be used in different CS1 course models and PLs. That means, some exercises can be reused and be reexpressed in different languages.
3. Some questions might require executing program fragments, so compilers and interpreters need to be integrated (as CASs are in ActiveMath). We assume learning is related to small programs written in a programming language but without specifying it explicitly. We want the model can be consider the language as a parameter, consequently.
4. Plans are actually expressed as (executable) metaprograms they may refer to expressions occurring in and operate on questions, as well. Therefore, plans require also a corresponding (metaprogramming)language. And once again, reuse can be required at the planning level.

5. An engine is required for *executing* plans. We consider it actually as a generalization of the concept of domain reasoning, we will name the domain reasoner machine, or simply domain reasoner (DR). Because the PL is a parameter we will denote it by DR<L>, where L is the specification language SL (which usually contains a programming language PL).

### **6.3 Assumptions on Complexity**

Our approach to deal with complexity will be based on the following assumption, which appears to be plausible: The complexity of a question corresponds to the complexity of the associated plan to answer it in an exercise. Question and plan are the design product of an author. Such a complexity is a function of the cognitive complexity of the operations used in the plan, which are the operations of the machine DR<L> used to execute plans. By mapping those operations to the Bloom's Taxonomy we would be able to assign a complexity to the question. Our goal in the following will be to suggest and illustrate which kind of reasoning could be involved behind those operations that we have in mind for DR<L>.

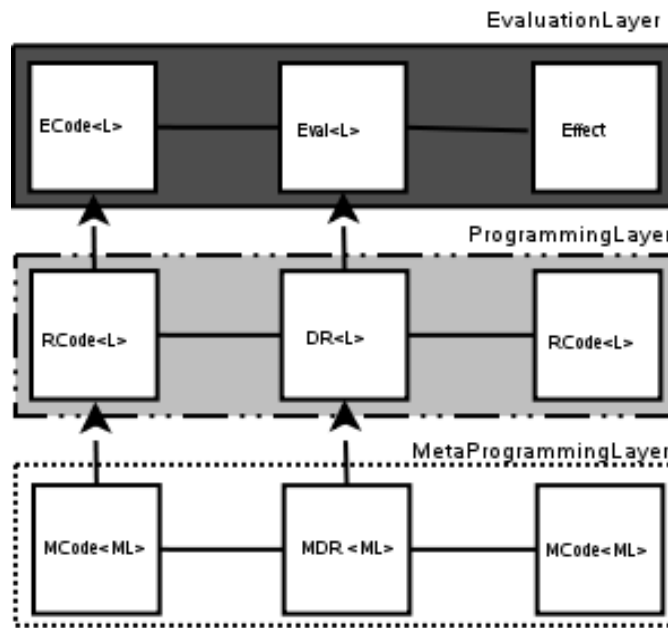


Figure 6: Reasoning Architecture

## 7 A Layered Reasoning Model

We are now able to introduce our proposal for a representation of our conceptual model that we will describe from an architectural point of view. The main idea is shown in Figure 6. Our representation model is three-layered. We use a metamodeling approach to cope with the requirements of genericity and multi-language capabilities, as facilities we already have identified and indicated, previously.

At the upper level, we have the evaluation layer, where the framework allows adding tools like compilers and interpreters or proxys for them, such that code fragments of the programming language (L) can be evaluated (as CASs systems do in ActiveMath). Expressions in L are reduced to some form of `effect` object that can be observed by the client of the layer.

Exercises are expressed at the middle layer (called programming layer) where we may have operations working on them which are, intuitively speaking, the plan playing operations. We also will have operations for generating code targeting the evaluation level.

Finally, at the lower level, we have the metaprogramming layer. Here, we similarly have operations to create metaplans and by this mean we are able to support plan reuse. A meta-language ML is used to express plans, code, concepts, which together are called reasoning code (denoted `RCode<L>`) and metaplans (denoted by `MetaCode<ML>`), too. We call evaluation code (`RCode<L>`) the representation of the programming language L.

Based on a metaplan we may generate plan instances that correspond to different exercises of the same kind. Some of them may require evaluation at the evaluation layer, naturally. We also write `Eval<L>`, `DR<L>` and `MDR<L>` to denote the evaluation/reasoning engines at the

corresponding layer. We implicitly intend that they get implemented as rule-based engines, at least at the meta and programming level.

This model offers us an additional advantage beyond modularity and genericity features. It give us an interesting metaphor in terms of cognitive models which are compatible with the Bloom's Taxonomy, as we have elaborated in the first part of this work. Thus, conceptually speaking, we may figure out that some student playing an exercise behaves like a kind of DR<L> instance. He/She has to *reason* on code and concepts and relate them with similar/equivalent code and concepts. Maybe code needs to be written and new concepts are introduced during this reasoning process by the student. The tasks involved are exactly the operations the plan associated with the exercise is demanding from him/her to perform.

By following the intended plan, the student can be *forced* during the playing to *go up*, in case the exercise demands executing a program code in L. Or to *go down* and perform an abstraction in order to match, instantiate or create a programming pattern, which require higher-order cognitive capabilities, as we know from Bloom's model. Form this perspective, we believe our abstraction will be reflecting such dynamics in a plausible way.

```

What output is produced by the following Java code:
for(int i=0;i<=3;i+=2)
System.out.print(i+" ");
a) 0 1 2
b) 0 1 2 3
c) 0 2
d) 0 2 3
e) 0 2 4

```

Figure 7: Exercise on Reading Code Operational level [31]

## 7.1 An Example on Layering

An example may illustrate some of the ideas and approaches we are aiming at. Consider the exercise shown in Figure 7. It comes from [31]. It tests the semantics of `for`-loops, the termination criteria and the operator `+=`.

Let us get situated in the authoring role. Evidently, the exercise can be generalized in different ways by using parameterized templates such that the values of the `for` statement or the operators can be changed. In such a case, we would need a metaplan. In order to keep the example simple we just write parts of it. We use an abstract syntax for more clarity. We consider the structure of the exercise and also the planning function in this example.

Aiming at the application of genericity over the exercise structure, we may want to have a parameterized code fragment by defining a code generating rule like the following:

```

#genBody(init, test, limit, assign, step) =>
  for(int i=$init;i $test $limit;i $assign $step)
    System.out.print(i+" ");
#end

```

Thus, case c), for instance, can be generated as follows, where a distraction rule (`buggy_limit`) is used.

```
#caseC(init, test, limit, assign, step) =>
  c) {genBody(init, test, buggy_limit(limit), assign, step)}
#end
```

Such kind of rules are evaluated at the MDR level, using generators targeting RCode. We intend more than textual generation occurs at the MDR level, but we just wanted to present a simple impression of the concept.

To complete the example, it would remain the matter of plan representation and manipulation. Because our main concern is actually complexity, we will develop the ideas in an independent section, for more clarity.

## 7.2 On Plans and Complexity

We briefly return to example 7 to motivate our analysis. With respect to the intended plan in that exercise let us suppose a simple one is expected by the author. Namely, the student will check each MCQ case, sequentially, until he/she finds the correct one. In each step the student runs the body loop and check it against the number in the proposed sequence. In our conceptual model, it would mean that the student will be first processing at the DR conceptual level and then going to the Eval level to evaluate statements at the operational and returning again to the DR level.

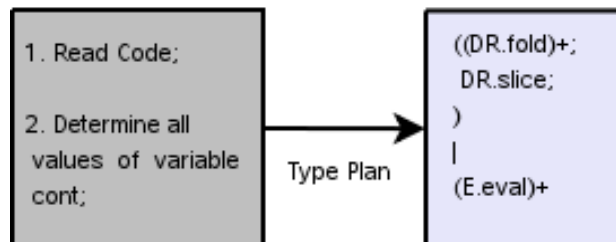


Figure 8: Plan Translation and Typing

We could certainly try to formalize such a kind of plan in a particular formal language. However, we realize that would require an independent part of the project we are trying to define; we are not in the position of going to that level of detail, yet. Nevertheless, we think it is possible.

However, we actually want to accomplish that task in a way that allows us to translate plan statements into a simpler logic language, for reasoning about complexity and eventual exercise discrepancies. Mapping the complexity of a question of an exercise to the complexity of the corresponding plan is consistent with our assumption that the complexity of an exercise is the one is expressed in its plan by its author. The idea is depicted in Figure 8 where some hypothetical operations on layers are assumed (named `fold`, `slice` and `eval` with the intended meanings).

We figure the translation as being a kind of *typing process* where the atomic types of the corresponding target formal language are `verbs` associated with cognitive layers (Eval, DR, or MDR) in our conceptual model. Such verbs will have assigned a cognitive complexity, as in Bloom, and we will need to define two things:

- How combinations of patterns denote more or less complexity, logically speaking.

- How complexity might affect student decisions during exercise playing.

We will illustrate and explain in more detail the idea of complexity pattern below (the concept is depicted in Figure 9).

Because of the inherent procedural nature of plans the logic behind the typing system should be accordingly capable of handling notions like composition, iteration and choice and the like. Hence, a dynamic-like logic seems to be a reasonable choice([22, 44]).

In such a sense, we might expect that an expression denoting a plan can be of the form  $P_1|P_2$  for choice,  $P_1;P_2$  for composition or  $(P)^*$  for iteration, among others. As already indicated the atomic propositions would denote primitive operations on the cognitive layers. Plans can also be denoted by rules; thus  $P \Rightarrow P_1|P_2$  would be a plan that decomposes into two subplans  $P_1$  and  $P_2$ , any of which can be followed to solve  $P$ , independently.

Just as a matter of illustration of the intended concept, we may want to express the possibility that a student could miss a part  $P_2$  of a composed plan of the form  $P_1;P_2$  if  $P_2$  is too much smaller than the other one  $P_1$ , in terms of complexity. As a result,  $P_2$  could wrongly be ignored. Even though the student is able to follow each part of the plan he could fail to accomplish as a whole. We express this as follows:

$$\frac{\Box_{Exercise}(\Box_{Student}(P \Rightarrow P_1;P_2)) \quad P_1 \gg P_2}{\Diamond_{Student}(P \Rightarrow P_1)} \text{MISSINGPATH}(P_2)$$

In such a case, subplan  $P_2$  is practically functioning like a strong *distractor* eventually leading to a mistake. Once again, the question is whether such an *eventuality* is casual or it is so designed at purpose. In the former, case, it is important to detect it. In the later case, how can be the exercise *animated* in an e-Learning environment such that the student can learn to recognize such distractors more carefully during ulterior testing stages.

### 7.3 Pattern Complexity and Reasoning Operations

As an exercise, we want to offer a possible *understanding* to events like distractors in terms of our plan complexity idea. For such a purpose, and just as an hypothesis, we assume that a *rational* student will follow a plan always trying to stay at those cognitive layers which are best known to him given his knowledge and skills.

However, as the plan runs, some derived subtasks can force the student to change the preferred layer; in our model, for instance, if a student is at DR layer, he might get forced going through a DR;E transition or a DR;MDR, depending on the student *cognitive preferences*. The expression DR;E means to apply the operational semantics of the language while DR;MDR would mean to understand, in terms of some conceptual model of reference, what the code fragment intentionally denotes.

We illustrate some of this kind of pattern in Figure 9. We claim, for instance, that each selection of the MCQ in example in Figure 7 corresponds to case 1). Pattern 2) might be the solution approach an expert student would prefer to follow for the same MCQ example. Because experts (just as an assumption) could rather prefer to solve (reading-and-understanding) programming questions at more abstract levels than just mechanically computing.

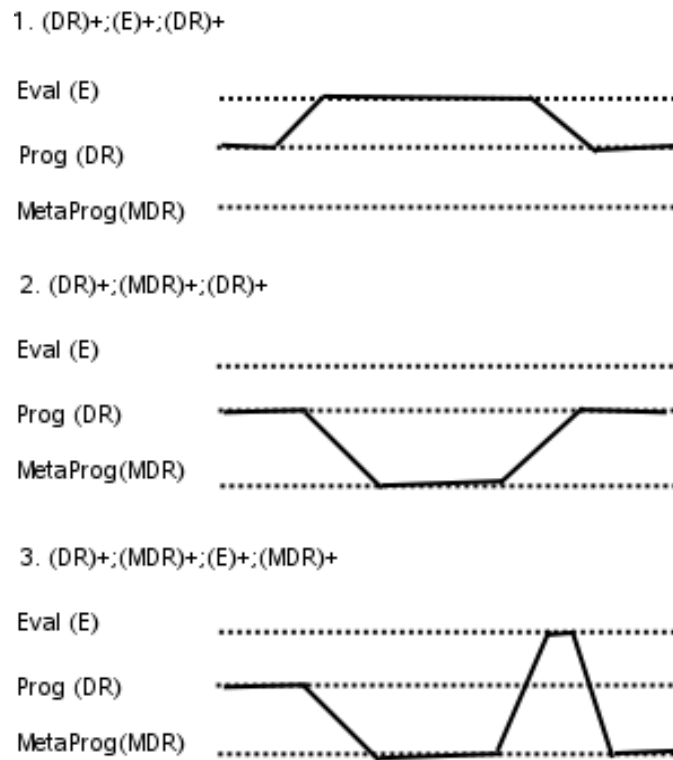


Figure 9: Patterns of Complexity

Moreover, we also claim that example in Figure 1 (and accompanying Figure 2) presented at the introductory section would match pattern 3) in Figure 9 for an expert student. We see that a transition fragment of the form  $MDR; DR; E$  is involved. This could mean a strong inflection; hence, it could eventually be ignored from the current plan, leading to a mistake.

At the contrary, a less experienced programmer could more naturally be taking a pattern like 1), which can result in a more smooth plan, if the student understands very well the operational semantics of the language.

## 8 Conclusions and Future Work

Our main objective in this work was to identify requirements and research areas for planned research project. In retrospective, we think we have accomplished our goals in different forms and at different levels. We have developed an architectural proposal that can be useful from different perspectives, both practical and theoretically perspectives. From a particular point of view, we have gained a much better understanding of the basic elements which are the important building blocks for programming learning material into an e-Learning environment like ActiveMath (we specifically mean programming exercises). We identified the need for a formal reasoning framework as a fundamental issue. We justified the ideas on standard cognitive models and frameworks. We think our proposal in this direction is compatible with the design principles of ActiveMath. We have identified the role that cognitive reasoning on programming can play in this context.

Summarizing, we believe that the following specific points are important to guide future work:

- A review of non-standard reasoning forms not covered by Bloom's Taxonomy in terms of our proposed model.
- A formal language for expressing programming plans and exercises taking into consideration generative capabilities.
- A typing system for translating plans into adequate forms of dynamic logic-like languages.
- The precise identification of the primitive verbs and corresponding Bloom-complexity mappings (i.e. Bloom's semantics). This has to be accomplished in a cognitively sound way.
- From a more practical perspective, the design of an extension of the ActiveMath OmDoc representation in order to be able to represent code at all the three layers (Eval, DR, MDR).

## Acknowledgments

We gratefully thank the DAAD for the financial support that made possible this visit. My very special thanks go to Jörg Siekmann for having invited me, for the contacts and the material support. Very special thanks to ArtinSoft for the support and space to make this visit possible. To George Gogvadze for the introduction and pointers to ActiveMath. And last but not least, I am particularly grateful to Serge Autexier, Dominik Dietrich, Marvin Schiller, Ewaryst Schulz and Marc Wagner for all the great help and the nice and friendly company.

## References

- [1] ACM-IEEE. *Computing Curricula 2001. Computer Science Volumen*. ACM and IEEE, <http://www.sigcse.org/cc2001/>, Dec. 2001.
- [2] L. W. Anderson and D. R. Krathwohl. *A Taxonomy for Learning, Teaching and Assesing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, 2001.
- [3] D. Barnes and M. Kölling. *Objects First with Java. A Practical Introduction using BlueJ*. Prentice-Hall, 3rd edition, 2006.
- [4] J. Bergin. A Pattern Language for Initial Course Design. *SIGCSE Bull.*, 33(1):282–286, 2001.
- [5] T. J. Biggerstaff, B. G. Mitbender, and D. E. Webster. The Concept Assignment Problem in Program Understanding. In *ICSE*, pages 482–498, 1993.
- [6] R. H. Bordini and J. F. Hübner. BDI Agent Programming in AgentSpeak using Jason (Tutorial Paper). In Toni and Torroni [46], pages 143–164.
- [7] R. J. Brachman and H. J. Levesque. *Knowledge Representation and Reasoning*. Elsevier, 2004.
- [8] D. Buck and D. J. Stucki. Design Early Considered Harmful: graduated exposure to complexity and structure based on levels of cognitive development. *SIGCSE Bull.*, 32(1):75–79, 2000.
- [9] A. Di Stefano and C. Santoro. Designing Collaborative Agents with eXAT. In *Second International Workshop on Agent-based Computing for Enterprise Collaboration*, June 2004.
- [10] B. Emir, M. Odersky, and J. Williams. Matching Objects with Patterns. In *ECOOP 2007*, volume 4609 of *LNCS*. Springer, 2007.
- [11] M. Felleisen, R. B. Findler, M. Flatt, and S. Krishnamurthi. The TeachScheme! Project: Computing and Programming for Every Student.
- [12] M. Felleisen, R. B. Findler, M. Flatt, and S. Kristnamurthi. The DrScheme Project: An Overview. *SIGPLAN Notices*, 33(6):17–23, 1998.
- [13] J. E. L. Gayo, J. M. M. Gil, A. M. F. Álvarez, and H. S. Chigne. A Generic E-learning Multi-paradigm Programming Language System: IDEFIX Project. *SIGCSE Bull.*, 35(1):391–395, 2003.
- [14] G. Germain. Concurrency Oriented Programming in Termite Scheme. In *2006 ACM SIGPLAN Workshop on Erlang*, pages 20–20. ACM, 2006.
- [15] H. Glaser, P. H. Hartel, and P. W. Garratt. Programming by Numbers: A Programming Method for Novices. *Comput. J.*, 43(4):252–265, 2000.
- [16] G. Gogvadze, A. G. Palomo, and E. Melis. ActiveMath. In *International Conference on Computers in Education (ICCE 2005)*, Singapore, 2005.

- [17] G. Gogvadze and J. Tsigler. Authoring Interactive Exercises in ActiveMath. In *MathUI Workshop at Mathematical Knowledge Management*, <http://www.examat.org/>, June 2007.
- [18] G. Gonzalez. A Systematic Approach to Active and Cooperative Learning in CS1 and its Effects on CS2. *SIGCSE Bull.*, pages 133–137, 2006.
- [19] T. Greening and R. Lister, editors. *Fifth Australasian Computing Education Conference (ACE 2003), Adelaide, Australia, 4-7 February 2003*, volume 20 of *CRPIT*. Australian Computer Society, 2003.
- [20] S. Grissom, D. Knox, D. Joyce, and W. Dann, editors. *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education, 2003, Reno, Nevada, USA, February 19-23, 2003*. ACM, 2003.
- [21] M. Guzdial and E. Soloway. Teaching the Nintendo Generation to Program. *Commun. ACM*, 45(4):17–21, 2002.
- [22] D. Harel. Dynamic Logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic Volume II — Extensions of Classical Logic*, pages 497–604. D. Reidel Publishing Company: Dordrecht, The Netherlands, 1984.
- [23] J.-M. Hoc, G. T. R. G., D. Gilmore, and R. SamurĀğay, editors. *The Psychology of Programming*. Academic Press, London, 1991.
- [24] E. Holohan, M. Melia, D. McMullen, and C. Pahl. The Generation of E-Learning Exercise Problems from Subject Ontologies. In *Sixth Conference International Conference on Advanced Learning Technologies*, 2006.
- [25] C. D. Hundhausen, J. L. Brown, S. Farley, and D. Skarpas. A Methodology for Analyzing the Temporal Evolution of Novice Programs Based on Semantic Components. In *ICER '06: Proceedings of the 2006 international workshop on Computing education research*, pages 59–71, New York, NY, USA, 2006. ACM.
- [26] T. Jenkins. The Motivation of Students of Programming. *SIGCSE Bull.*, 33(3):53–56, 2001.
- [27] C. G. Johnson and U. Fuller. Is Bloom’s Taxonomy Appropriate for Computer science? In A. Berglund and M. Wiggberg, editors, *Proceedings of the Sixth Baltic Sea Conference on Computing Education Research*, volume 2007-006 of *Uppsala University Department of Information Technology Technical Reports*, pages 120–123. Uppsala University, February 2007.
- [28] M. Kohlhase. *CodeML: An Open Markup Format the Content and Presentation of Program Code; the CodeML Specification*. OmDoc.org, <https://svn.omdoc.org/repos/codeml/doc/spec/codeml.pdf>, 2008.
- [29] M. Kohlhase. *An Open Markup Format for Mathematical Documents, Version 1.2*. Number 4180 in *LNAI*. Springer, <http://www.omdoc.org/pubs/omdoc1.2.pdf>, April 2008.

- [30] M. Kuittinen and J. Sajaniemi. Teaching Roles of Variables in Elementary Programming Courses. In *ITiCSE '04: Proceedings of the 9th annual SIGCSE conference on Innovation and technology in computer science education*, pages 57–61, New York, NY, USA, 2004. ACM.
- [31] R. Lister. Objectives and Objective Assessment in CS1. In *SIGCSE*, pages 292–296, 2001.
- [32] R. Lister and J. Leaney. First Year Programming: Let All the Flowers Bloom. In Greening and Lister [19], pages 221–230.
- [33] R. Lister and J. Leaney. Introductory Programming, Criterion-referencing, and Bloom. In Grissom et al. [20], pages 143–147.
- [34] R. Lister, B. Simon, E. Thompson, J. L. Whalley, and C. Prasad. Not seeing the forest for the trees: novice programmers and the solo taxonomy. *SIGCSE Bull.*, 38(3):118–122, 2006.
- [35] M. McCracken, V. Almstrum, D. Diaz, M. Guzdial, D. Hagan, Y. B.-D. Kolikant, C. Laxer, L. Thomas, I. Utting, and T. Wilusz. A Multi-national, Multi-institutional Study of Assessment of Programming Skills of First-year CS Students. In *ITiCSE-WGR '01: Working group reports from ITiCSE on Innovation and technology in computer science education*, pages 125–180, New York, NY, USA, 2001. ACM.
- [36] E. Melis, E. Andrés, J. Büdenbender, A. Frischauf, G. Gogvadze, P. Libbrecht, M. Pollet, and C. Ullrich. Activemath: A Generic and Adaptive Web-Based Learning Environment. *International Journal of Artificial Intelligence in Education*, 12(4):385–407, 2001.
- [37] B. Meyer. Testable, Reusable Units of Cognition. *IEEE Computer*, 39(4):20–24, 2006.
- [38] M. Pedroni and B. Meyer. The Inverted Curriculum in Practice. *SIGCSE Bull.*, 38(1):481–485, 2006.
- [39] K. Powers, P. Gross, S. Cooper, M. McNally, K. J. Goldman, V. Proulx, and M. Carlisle. Tools for Teaching Introductory Programming: What works? *SIGCSE Bull.*, 38(1):560–561, 2006.
- [40] V. Robinet, G. Bisson, M. Gordon, and B. Lemaire. Searching for Student Intermediate Steps. In *UM2007: Workshop On Data Mining For User Modeling*, June 2007.
- [41] C. K. Roy, T. Noll, B. Roy, and J. R. Cordy. Towards Automatic Verification of Erlang Programs by  $\pi$ -Calculus Translation. In *2006 ACM SIGPLAN Workshop on Erlang*, pages 38–50. ACM, 2006.
- [42] G. Smolka. *Programmierung: Eine Einführung in die Informatik mit Standard ML*. Odenbourg.de, 2008.
- [43] E. Soloway. Learning to Program = Learning to Construct Mechanisms and Explanations. *Commun. ACM*, 29(9):850–858, 1986.
- [44] L. Spalazzi and P. Traversi. A Dynamic Logic for Acting, Sensing and Planning. *Journal of Logic Computation*, pages 1–36, 1999.

- [45] E. Thompson, A. Luxton-Reilly, J. L. Whalley, M. Hu, and P. Robins. Blomm's Taxonomy for CSAssesment. In *Tenth Australasian Computing Education Conference ACE2008*, 2008.
- [46] F. Toni and P. Torroni, editors. *Computational Logic in Multi-Agent Systems, 6th International Workshop, CLIMA VI, London, UK, June 27-29, 2005, Revised Selected and Invited Papers*, volume 3900 of *Lecture Notes in Computer Science*. Springer, 2006.
- [47] <http://en.wikipedia.org/wiki/Elearning>. Electronic Learning.
- [48] S. Wiedenbeck. Organization of Programming Knowledge of Novices and Experts. *Journal of the American Society for Information Science*, 37(5):294–299, 1999.
- [49] L. E. Winslow. Programming Pedagogy –A Psychological Overview. *SIGCSE Bull.*, 28(3):17–22, 1996.
- [50] C. Zinn. Supporting Tutorial Feedback to Student Help Requests and Errors in Symbolic Differentiation. In M. Ikeda and K. Ashley, editors, *Proceedings of Intelligent Tutoring Systems 8th. International Conference ITS-2006*, volume LNCS 4053 of *Lecture Notes in Computer Science*, pages 349–359. Springer-Verlag, June 2006.