

# How to improve understanding of object constructs using a slightly modified explanation

Rudolf Pecinovsky<sup>1</sup>

*Department of Information Technologies, University of Economics, Prague, Czech Republic*

## Abstract

Most textbooks and courses explain basic object oriented (OO) constructs in a very similar way. Extensive experience with teaching different kinds of courses at various levels, from primary and secondary school through to university and requalification courses for professional programmers shows that many students have a difficulty with this traditional approach. In this paper we show that a modified approach leads to a better understanding of the basic OO constructs.

## Keywords

OOP, Design Patterns First, teaching, introductory courses

## 1. Introduction

Object Oriented Programming (OOP) is a fundamental paradigm of modern programming languages. Over the last 10 years we have been teaching OOP at computer clubs for students from primary schools as well as at high schools, grammar schools and universities. At the same time we have been teaching industry-based courses to retrain professionals from structured programming paradigm to OOP and to improve their knowledge and skills. As a result, we have experience with teaching a range of students from complete beginners to students with advanced knowledge of programming obtained from textbooks or other courses.

Both beginners and advanced programmers experience problems with mastering certain object oriented constructs. We have succeeded in modifying the explanation of OO constructs so that beginners improve their understanding and advanced programmers learn to avoid poor programming habits acquired as a result of incorrect understanding of OOP.

### *1.1 Summary of perceived problems*

Almost all textbooks explain the basic object oriented constructs<sup>2</sup> in a way that is more or less borrowed from older C++ textbooks. However, such an explanation involves many definitions that are difficult to understand for the beginners. From their point of view these constructs are often inconsistent and confusing.

In addition to these basic constructs several secondary constructs are explained as new, but using a slightly modified explanation, we can explain the secondary constructs as natural

---

<sup>1</sup> Address: ICZ, Na Hřebenech II 1718/10, 140 00 Prague 4, Czech Republic,  
E-mail address: rudolf@pecinovsky.cz

<sup>2</sup> In this article I concentrate on Java textbooks, however we encounter similar problems in textbooks dealing with other programming languages. Often we can use the modifications I suggest for teaching Java in these languages too.

extension of the basic ones. Moreover, we can refer to the constructs that students have already mastered. This slight modification can improve the understanding of both sets of constructs. Furthermore, this decreases the number of problems that the students may encounter when using these constructs in their programs.

Our experience has shown that it is useful to explain the following topics in a slightly different way than textbook authors have done so far:

- what is an object,
- the difference between objects and classes,
- the concept of the interface,
- constructors and construction of objects,
- the keyword `this`,
- class inheritance.

In the following sections we deal with each of the above mentioned topics and show the modification of explanation that proved useful in our attempts to improve understanding of the subject matter.

We have compared our method with the following textbooks [Arnold, Gosling & Holmes \(2005\)](#), [Barnes & Kölling \(2004\)](#), [Deitel & Deitel \(2007\)](#), [Eckel \(2007\)](#), [Fain \(2004\)](#), [Horstman \(2007a\)](#), [Horstman \(2007b\)](#), [Horstman & Cornell \(2007\)](#), [Horton \(2002\)](#), [Liang \(2006\)](#), [Morelli & Walde \(2006\)](#), [Schildt \(2004\)](#), [Schildt \(2005\)](#), [Sierra & Bartes \(2009\)](#), [Winder & Graham \(2006\)](#), [Zakhour, Hommel, Royal, Rabinovitch, Risser & Hoeber \(2006\)](#) and [Zukowski \(2002\)](#). We can divide these textbooks into three groups:

- [Barnes & Kölling \(2004\)](#), [Horstman \(2007a\)](#), [Horstman \(2007b\)](#) and [Sierra & Bartes \(2009\)](#) – mainly concentrate on explanation of the best programming practices. They intend to teach how to think and how to program in a *true* object oriented way.
- [Arnold, Gosling & Holmes \(2005\)](#), [Eckel \(2007\)](#), [Horstman & Cornell \(2007\)](#), [Horton \(2002\)](#), [Schildt \(2004\)](#), [Zukowski \(2002\)](#) – mainly teaching the language with its APIs. Teaching the art of programming is secondary.
- Remaining texts claim that they teach OOP, however the style of explanation and the discussed topics indicate that they belong to the second group.

We note that the objective of this article is not to review the above mentioned textbooks, but to use them as examples of the traditional way of explanation of object oriented constructs and compared it with the proposed approach. We are not going to enumerate the explanations of the topics in the above mentioned textbooks, but we focus on summarizing these approaches.

## 2. Everything is an object

The textbooks explain the term *object* in two ways:

- most of them explain it using *real* world examples,
- others do not explain it and assume that it is a generally known term which does not need a special explanation.

In both cases students meet similar problems. Their general understanding is that *object* is something tangible, they have not come across the idea that objects are used also for

representing abstract ideas (e.g. beauty, size, direction, connection, interruption, calculation, etc.). If they meet such objects in a program for the first time, it will take some time until they accept the fact that abstract ideas can be also represented by objects.

### 2.1 Recommendation

We have discovered that it is useful to explain to students at the very beginning that in object oriented programming we treat as an object everything that can be expressed by a noun, including the abstract terms mentioned above. Some students may be confused by it for a while, and they find it difficult to describe an abstract term by means of an object. Therefore, we explain here that in programs each object is represented with a set of data items (attributes) that describes the object. From the program's point of view the object is just this set of attributes and it does not matter whether the set represents a *physical* object or an abstract idea.

Most students quickly understand that besides the attributes that characterize cars, chairs, animals or other physical objects they can equally define attributes that characterize colors, directions, beauty, connections and other abstract terms. To facilitate this understanding, we have to use objects of this kind often from the very beginning of the course. Among suitable candidates for these *abstract* objects are, for example characteristics of graphical objects such as colors or directions.

## 3. Classes versus objects

In most textbooks the class is explained as an abstraction describing some properties of a group of objects, which we call instances of their parent class. Authors often explain that a class serves as a blueprint or a template for the objects that the program uses. Some authors note that we can look at a class as a factory capable of creating objects on demand.

Students sometimes struggle with understanding the difference between classes and objects, especially when we introduce static attributes and methods.

### 3.1 Recommendation

Our experience shows that students understand this topic better, when we explain that classes are also objects (we treat everything that can be named by a noun as an object, therefore classes should also be treated as objects). However, classes are special objects with special properties:

- They are the only objects that can create new objects called instances of their parent class. When advanced students complain that other objects can also create new objects, we explain, that these “other objects” can only return objects that are originally created by a class. A new object can be created only by its parent class.
- Classes are the only objects that are not instances of any class (this is true in Java, in other languages – e.g. in Smalltalk this may not be true). We should address classes in programs directly by their names (we cannot assign a class to a variable). In some languages (e.g. in Java) classes are represented by a special class-objects – in Java by instances of the class `Class`. However, a class-object is just a representative of the class, the class-object is not the class itself.
- Classes define two types of attributes:

- First type of attributes is marked with the modifier `static`. We can interpret static attributes as attributes that do not move and stay “statically” in their class.
- The attributes that are not marked with the modifier `static`, declare attributes for instances of the class. Each created instance takes its own copy of these attributes.
- Similarly, classes define two types of methods:
  - Methods marked with the modifier `static` belong to the class and can use directly only members (attributes and methods) of their class. Members of instances and other classes should be qualified by their owner.
  - Other methods belong to instances. They have (similarly to constructors) a hidden constant parameter `this`, which is initialized by a reference to their instance. Although the parameter `this` is not included in the list of parameters, it can be used in the body of a method.
 

These instance methods can use the members of their instance and its class directly. (We observe that the class behaves as if it were a proper parent and allows all of its instances to use its (i.e. static) attributes and methods.) Other members should be qualified by their owner.

It is quite astonishing how this small difference in explanation helps students to understand the term class and how it helps them to solve some more complex problems.

This explanation is (unintentionally) endorsed also by IDE *BlueJ*, which we use in our introductory programming courses. In *BlueJ* we work with classes and objects in a very similar way. Classes as well as their instances are represented by rectangles, whose context menus display all messages that can be sent to the corresponding class/object. The only difference is that classes are shown in the class diagram while instances are shown in the object bench. Thus students find this explanation consistent with their experience.

Introduction of classes as a special kind of objects helps also in the explanation of other topics:

- Students have no problems with understanding the difference between class and instance attributes and methods, and they can use both almost from the beginning of the course.
- Students understand more easily the rules for loading a class by a `ClassLoader` and it helps them later to understand better the principles of inheritance.

#### 4. Constructor and parameter `this`

Another topic whose understanding and use sometimes causes problems are constructors and the keyword `this`. Almost all textbooks follow the original description published in [Strustrup \(1991\)](#), which says: “*Constructor is identified by having the same name as its class.*” The text does not differentiate if the constructor is a method.

The above mentioned textbooks differ in the explanation of what is a constructor. [Arnold, Gosling & Holmes \(2005\)](#) and [Horstman \(2007\)](#) explain that a constructor is not a method and therefore it may not return anything. Accepting this explanation leads us to assign the responsibility for returning the new object to the new “operator”<sup>3</sup>.

Most authors define a constructor as a special kind of a method having the same name as its class. However, most of them do not explain why the reflection, exceptions and almost all debuggers use for constructor the name `<init>`. They either ignore it (e.g. [Deitel & Deitel](#)

---

<sup>3</sup> De iure the new is not an operator, however, many teachers and programmers understand it so.

(2007), on page 927) or they do not discuss the difference between this name and the name introduced at the beginning of the course (e.g. Horstman & Cornell (2007)).

Taking a constructor as a method with the name of its class and not declaring the type of its return value, we should introduce a new construct `this(...)` serving for invoking another constructor and transfer the responsibility for the initialization of the created object to it.

#### 4.1 Recommendation

When we look at the constructor syntax we can interpret it in two ways:

- A method identified by its class name and declaring no return type.
- A method identified by the empty string and declaring its class as a return type.

The latter explanation is closer to the actual implementation. It appears that it is more efficient to explain constructors in this way and explain them as methods with special properties:

- In Java the internal name of constructors is `<init>`. However, this name violates the rules for identifiers and therefore Java authors decided to declare constructors in source code as “empty-string” methods, or more precisely as methods, whose names are empty strings.
- A constructor must return a reference to the newly created object. This reference is obtained from the hidden parameter `this`, which is initialized by the caller. Because the returned value is known *a priori*, the language syntax theoretically does not need the statement  
`return this;`  
In fact, we do not write it, it is inserted by the compiler on our behalf to prevent mistakes.
- A constructor can be used only for initialization of the newly allocated memory and therefore it can be invoked only by the new “operator” or by another constructor of the same class. In the latter case the invocation statement must be the very first statement of the calling constructor body to ensure that the memory is not initialized.

After the above explanation students understand better the explanation of the following syntactic rules. We explain:

- Construction of a new object proceeds in two steps:
  - First, the new “operator” is called with a parameter defining the name of the class, whose instance we want to create (the parent class). This parameter determines the size of the memory allocated for the created object and it also specifies other information needed for the creation of the object (e.g. the address of the VMT). Additionally, the allocated memory is filled with zeroes and/or compile-time constants.
  - Second, the “empty-string” method (constructor) is invoked with argument `this` pointing to the allocated memory and possibly also with other arguments. The constructors’ task is initializing this memory so that it correctly represents the object.

We can outline the described behavior by writing the statement in two lines:

```
new ClassName    //Invoking the new "operator", memory allocation
(/*parameters*/); //Invoking the constructor
```

- As we have noted, the constructor can be used only for initialization of the newly allocated memory. If it is invoked by another constructor, this invocation must be the very first statement in its body. Nothing may precede it, not even an opening brace.

- If a constructor delegates its responsibility for initializing the object to another constructor, it should qualify this invocation by `this` as we are used to doing with normal methods. However, in this case we do not write the *dot*. So instead writing

```
this.(*parameters*);
```

we write only

```
this.(*parameters*);
```

When we explain constructors in this way, students more easily understand the `this()` statement as a means for delegating the responsibility for initializing the object and they also understand what the `<init>` appearing in exception messages or debugger windows means.

This explanation establishes a good basis for the following explanation of static initializers and invocation of super constructor. Everything fits logically together.

When the above explanation is used students sometimes complain that the object is not created by the constructor but by the new “operator”. Here we can use the following analogy: “Who makes cups?” They answer: “A potter.” Then we explain that the allocated memory serves similarly as potter's clay and that constructor processes this memory similarly as the potter processes the clay. Using this analogy, we regard the constructor as the author of the created object.

## 5. Inheritance

The most common problem with teaching inheritance is that it is taught too early. Some textbooks deal with it immediately after the first introduction to objects and classes.

For now we ignore that if we want students to acquire the knowledge of the OO paradigm well, we should not explain the concept of inheritance until we explain the concept of interface (a general interface as well as the Java construct interface). In addition, we should not only teach how to implement an interface, but also how to incorporate it in our design. These problems are discussed in [Pecinovský, Pavlíčková & Pavlíček \(2006\)](#), [Pecinovský & Pavlíčková \(2007\)](#) and [Pecinovský \(2009a\)](#) during explanation of the *Design Patterns First* methodology. From the textbooks mentioned at the beginning of this paper only [Horstman \(2007\)](#), [Horstman & Cornell \(2007\)](#) and [Zakhour, Hommel, Royal, Rabinovitch, Risser & Hoerber \(2006\)](#) first explain the interface and later the inheritance.

When explaining inheritance all the mentioned textbooks explain that a child should represent a special kind of its parent. However, they do not put the same emphasis on it. Mostly, they mention this rule only at the beginning of the explanation of inheritance and then they show only how we could use the inheritance to avoid writing additional code. Unfortunately, the majority of programming textbooks do not present bad examples of inheritance usage at all. This would warn the reader against a bad design early.

After such an explanation the students often remember only that inheritance serves primarily for reusing code and they also use it only for this purpose.

### 5.1 Three kinds of inheritance

At the beginning of explaining class inheritance we should introduce the three kinds of inheritance ([Lalond & Pugh \(1991\)](#)):

- **Inheritance of interface** (in [Lalond & Pugh \(1991\)](#) *subtyping*) occurs when a child inherits the entire interface from its parent, i.e. the signature as well as the contract. As a consequence, an instance of a subtype can stand in for an instance of its supertype. However, a compiler ensures the inheritance of the signature only. Maintaining the

contract is the programmer's job. Subtype implementation details are totally irrelevant; all that matters is that it has the right behavior so that it can be substituted.

- **Inheritance of implementation** (in [Lalond & Pugh \(1991\)](#) *subclassing*) – it is an implementation mechanism for sharing code and representation. The subclass inherits all the implementation from its superclass (it is the compiler's job). The subclass can change the behavior that does not fit its requirements, and it can also add new members. Here, the danger is that the overridden code and/or new members violate the parent's contract.
- **Natively understood inheritance** (in [Lalond & Pugh \(1991\)](#) the “*is-a*” *relationship*) talks about our assumption that one kind of object is a special case of another. Here an inconsistency may appear when the implementation differs from our inherent assumption. E.g. mathematicians tell us, that a map is a special kind of a set – it is a set of ordered pairs (key, value). However, in the java standard library the set is implemented as a special kind of a map.

## 5.2 Recommendation

There are two recommendations:

### 5.2.1 Postpone the explanation of class inheritance as late as possible

The reason for postponing this explanation is to offer enough time for exercising usage of interfaces. Students should learn not only how to implement a given interface, but they also should master how to recognize situations, where incorporating an interface in their design is useful.

At this point it is useful to introduce the *Decorator* design pattern and prepare at least one project, where using this pattern is more useful than the frequently (and improperly) used inheritance of classes. There are two reasons why to introduce this pattern:

- Advanced students who mastered inheritance in a previous course (or from a textbook) are provided with situations where class inheritance is not the best solution. It also helps us to improve students' attention to the ongoing explanation.
- We prepare the background for the following explanation of class inheritance.

If our lessons follow the *Design Patterns First* methodology ([Pecinovský, Pavlíčková & Pavlíček \(2006\)](#), [Pecinovský & Pavlíčková \(2007\)](#) and [Pecinovský \(2009a\)](#)), an introduction of the *Decorator* design pattern does not present a problem since the students already know several design patterns and they understand their importance.

### 5.2.2 Explain class inheritance as an automated implementation of Decorator

As the next step we inform students that in addition to the inheritance of interfaces (a language construct) there is also class inheritance. This inheritance combines the inheritance of the parent class interface with the inheritance of the parent implementation. We explain that the inheritance of implementation is internally handled as if the subclass were designed according to the *Decorator* design pattern. In other words, the inheritance of implementation is *de facto* an application of the *Decorator* design pattern in which the decorator (child) acquires both the implementation and the interface from the decorated object (parent). The compiler prepares a hidden constant attribute named `super`, in which a reference to the *decorated object* is held. Additionally, the compiler also ensures the automatic delegation of all inherited methods to `super`.

For the decorated “super” object we introduce the term *parent subobject*. In contrast to the standard decorator a constructor of a child does not take its parent (super) as a parameter, but it creates the parent subobject by calling a parent’s “empty-string” method (a constructor):

```
super (/* parameters */);
```

where, similarly to the statement `this()`, we omit the *dot*.

We explain that the parent subobject must be created before the rest of the child object is initialized to allow the rest to use the inherited members. So the child constructor has two options:

- to delegate its responsibility for initialization to one of its peers by the statement `this()` or
- to start with creating the parent object, i.e. calling its constructor by the statement `super()`.

The only exception is the situation, when we want to call the parameter-less parent constructor – then the compiler is able to insert its call for us behind the scene.

So far we did not to create the parent object in our classes explicitly, because the compiler implicitly used the parameter-less parent constructor. We may immediately show, that identical behavior can be obtained by adding the `super()`; statement into our original classes.

Our experience shows that the explanation following these rules is much more comprehensible for the students than using the traditional approach. Especially, the concept of overriding, which was difficult to understand for many students, is now clear and intelligible for most. Several programmers attending our retraining courses have commented that thanks to this explanation they finally fully understand the class inheritance.

We should not forget to remind students that the three kinds of inheritance must not interfere. They should fit together. In case of class inheritance, the compiler is able to ensure only the inheritance of the implementation and the signature. The inheritance of the contract is the responsibility of the programmer.

## 6. Summary

This paper was written in response to the problems that many students experience with understanding object oriented concepts. We show that by changing the way of explaining these OO specific constructs we can improve the comprehensibility of these concepts.

We recommend the use of objects that represent abstract concepts from the very beginning of explanation. Subsequently, class should be explained as a special kind of object with special properties – e.g. that it is the only object that can create new objects – its instances.

Further, we recommend explaining the constructor as a method whose name is an empty string and which can be used only for initializing a newly allocated memory. We show how this change makes some other constructs more logical.

Finally, we concentrate on inheritance. We suggest that the explanation of class inheritance be postponed after the explanation of interface, and that its explanation should be preceded by the explanation of the *Decorator* design pattern. *Decorator* design pattern facilitates understanding of the concept of class inheritance. In addition, the paper recommends that we explain to the students the three kinds of inheritance and emphasize that the compiler ensures only the inheritance of signature, but that ensuring correct inheritance of the contract is the programmer’s responsibility.

## Acknowledgements

This work was funded by the Faculty of Informatics and Statistics at University of Economics, Prague under the program Long term conceptual development of science and research.

## References

- Arnold K., Gosling J. & Holmes D. (2005). *The Java™ Programming Language, Fourth Edition*. Addison Wesley Professional. ISBN 0-321-34980-6.
- Barnes D. & Kölling M. (2004). *Objects First With Java: A Practical Introduction Using BlueJ (2<sup>nd</sup> Edition)*. Prentice Hall. ISBN 978-0-131-24933-2.
- Buchalcevová A. (2008) Buchalcevová, Alena. Where in the curriculum is the right place for teaching agile methods? Proceedings 6th ACIS International Conference on Software Engineering Research, Management & Applications (SERA 2008). Prague : Copyright, 2008, p. 205–209. ISBN 978-0-7695-3302-5.
- Deitel H. M. & Deitel P. J. (2007). *Java How to Program, 7<sup>th</sup> Edition*. Prentice Hall, ISBN 978-0-132-22220-4.
- Eckel B. (2007). *Thinking in Java (3<sup>rd</sup> Edition)*. Prentice Hall, ISBN 978-0-131-00287-2-6.
- Fain Y. (2004). *Java Programming for Kids, Parents and Grandparents*. Smart Data Processing. ISBN 0-9718439-5-3. Available from: [www.csd.abdn.ac.uk/~tnorman/teaching/CS1014/information/JavaKid8x11.pdf](http://www.csd.abdn.ac.uk/~tnorman/teaching/CS1014/information/JavaKid8x11.pdf)
- Gosling J., Joy B., Steele G. & Bracha G. (2005). *Java™ Language Specification, Third Edition*. Addison Wesley. ISBN 978-0-321-24678-3. Available from: [java.sun.com/docs/books/jls/download/langspec-3.0.pdf](http://java.sun.com/docs/books/jls/download/langspec-3.0.pdf)
- Horstman C. S. & Cornell G. (2007). *Core Java™, Volume I – Fundamentals (8<sup>th</sup> Edition)*. Prentice Hall PTR. ISBN 978-0-132-35476-9
- Horstman C. S. (2007a). *Big Java (3<sup>rd</sup> Edition)*. John Wiley and Sons. ISBN: 978-0-470-10554-2.
- Horstman C. S. (2007b). *Java Concepts for Java 5 and 6*. John Wiley and Sons. ISBN 978-0-470-10555-9.
- Horton I. (2002). *Beginning Java 2*. Wrox. ISBN 978-0-76454-365-4.
- Lalond W. & Pugh J. (1991). *Subclassing ≠ Subtyping ≠ IsA*. Journal of Object-Oriented Programming. Vol. 3, No. 5.
- Liang Y. D. (2006). *Introduction to Java Programming: Comprehensive Version (6th Edition)*. Prentice Hall. ISBN 978-0-132-22158-0.
- Morelli R. & Walde R. (2006). *Java, Java, Java, Object-Oriented Problem Solving (3<sup>rd</sup> Edition)*. Prentice Hall, ISBN 978-0-131-47434-5.
- Pecinovský R., Pavlíčková J. & Pavlíček L. (2006). *Let's Modify the Objects-First Approach into Design-Patterns-First*. Proceedings of 11th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE'2006). Bologna, ACM Press, ISBN 1-59593-346-8.
- Pecinovský R. & Pavlíčková J. (2007) *Order of explanation should be Interface – Abstract classes – Overriding*. Proceedings of 12th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE'2007). Dundee, ACM Press. Available from: [vyuka.pecinovsky.cz](http://vyuka.pecinovsky.cz)
- Pecinovský R. (2009a). *Early Introduction of Inheritance Considered Harmful*. Objekty, Hradec Králové.
- Pecinovský R. (2009b). *Using the methodology Design Patterns First by prototype testing with a user*. Proceedings of IMEM, Spišská Kapitula.
- Schildt H. (2004). *Java: The Complete Reference, J2SE 5 Edition*. McGraw-Hill Osborne Media. ISBN 978-0-07-223073-4.
- Schildt H. (2005). *Java: A Beginner's Guide, Third Edition*. McGraw-Hill Osborne Media. ISBN 0-07-223189-0
- Sierra K. & Bartes B. (2009). *Head First Java, 2nd Edition*. O'Reilly Media. ISBN 978-0-596-00920-5.
- Strustrup B. (1991). *The C++ Programming Language, 2<sup>nd</sup> Edition*. Addison-Wesley Publishing Company. ISBN 0-201-53992-6.
- Winder R. & Graham R. (2006). *Developing Java Software, 3<sup>rd</sup> edition*. John Willey & Sons, Ltd. ISBN 0-470-09025-1.
- Zakhour S., Hommel S., Royal J., Rabinovitch I., Risser T. & Hoeber M. (2006). *The Java Tutorial: A Short Course on the Basics, 4<sup>th</sup> Edition*. Prentice Hall PTR. ISBN 978-0-321-33420-6.
- Zukowski J. (2002). *Mastering Java 2*. Sybex. ISBN 978-0-7821-4022-4.