

# Aggregation of a Comparative Non-Parametric Statistics to Didactic Engineering

*Péricles César de Araújo*

e-mail: [pericles@uefs.br](mailto:pericles@uefs.br)

Pontifical Catholic University of São Paulo  
São Paulo, SP, Brazil

*Sonia Barbosa Camargo Iglioni*

e-mail: [siglioni@pucsp.br](mailto:siglioni@pucsp.br)

Pontifical Catholic University of São Paulo  
São Paulo, SP, Brazil

*Celina Aparecida Almeida Pereira Abar*

e-mail: [abarcaap@pucsp.br](mailto:abarcaap@pucsp.br)

Pontifical Catholic University of São Paulo  
São Paulo, SP, Brazil

*José Miguel Bezerra Filho*

e-mail: [profjmb@msn.com](mailto:profjmb@msn.com)

Pontifical Catholic University of São Paulo  
São Paulo, SP, Brazil

## Abstract

*In this article, the assumption is made that the use of a suitable quantitative data analysis method can increase the reliability of the results presented in research. In this context, it is the goal of this article to present an alternate proposal for quantitative data analysis that aggregates information and that is compatible with the qualitative approach of Didactic Engineering. This methodology, like all methodologies of qualitative research, rejects the resources of Classical Statistics, such as parametric analysis, the control case, or experimental groups and control groups. The validation method of Didactic Engineering is internal. The analyses are based on the comparison between the a priori and the a posteriori analyses. Considering that it is worth discussing such a rejection, an alternative resource for the use of formalized tests in the theory of Non-Parametric Statistics is presented. The justification is that this resource does not require a population model, and does not require a large number of hypotheses. In addition, Non-Parametric Statistics provides the Didactic Engineering methodology with a treatment that meets the prerogative of falsifiability in Popper's scientific method. This argument is based on theories of mixed methods research and complementarity.*

## 1. Introduction

The proposal to use Non-Parametric Statistics as a resource for the analysis of information that does not require a population model and that does not require a large number of hypotheses takes into consideration that this resource also provides the Didactic Engineering methodology with a treatment that meets the prerogative of falsifiability in Popper's scientific method. Whereas Popper constitutes the epistemological basis of quantitative methods, the epistemological bases of qualitative methods are the arguments of Kuhn (1996). Using complementarity, we have added quantitative aspects to Didactic Engineering, a qualitative method.

Complementarity, as defined by Bohr (1995), claims that human nature is comprised of two images, just as waves and particles are considered complementary aspects of matter. According to Otte (2003), in the realm of Mathematical Education, complementarity refers to symbols and concepts, in a reciprocal sense, that reciprocally readjust themselves to capture the essential elements of the cognitive and epistemological development of scientific knowledge and mathematical concepts.

Thus, the authors accept Popper's propositions that the problems of demarcation and induction consist of the same thing, or rather, that there is a tight bond between them, and that the Scientific method is the critical attitude. The inductive method, supposedly legitimized by long series of observations and experimentation, cannot provide, according to Popper, a satisfactory demarcation criterion. Popper states that valid induction that is legitimized by a long series of repeated observations, cannot guarantee true generalizations or inferences. Thus, it is concluded that induction gives rise to theories that are merely probable and not certain.

For Popper, scientific validation occurs through conjectures (the result of analyses that are *a priori* to the experimental phase) and refutations (the result of analyses that are *a posteriori* to the experimental phase).

Didactic Engineering, the method analyzed in this article and for which we intend to present a complementary proposal for refining its parameters for inference and scientific validation, is not suited to the inductive model and is characterized by presenting the *a priori* and *a posteriori* analysis phases, and thus it is suited to the criteria posed by Popper. This is the justification by this article's authors for proposing the use of Non-Parametric Statistics tests as an alternative resource that adds testability to the Didactic Engineering method. Didactic Engineering, as defined by Artigue (1990), is a qualitative method, that is, a descriptive method that deals with process and in which the researcher is not concerned with evidence that proves hypotheses defined prior to the beginning of the process, or, in Creswell's words (2009, p271):

Qualitative research is a means for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions or procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data. The final written report has a flexible structure.

Mathematical Education, understood as an educational project, that is a set of approaches to formal education, also, as an object of research, moves in the direction of the ideas of Giere (1988), for whom explanation is a human activity and mathematics, as one of the many types of mental schema, is a modern means to understand science. As Creswell (2009) suggests, the researcher in Mathematical Education who seeks to explain or understand the educational process uses the methods of qualitative, quantitative, or mixed research. To the authors, quantitative and mixed methods meet the characteristics presented by Creswell (2009, p. 271, 272):

Quantitative research is a means for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures.

Mixed methods research is an approach to inquiry that combines or associates both qualitative and quantitative forms. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study.

According to Teddlie and Tashkkori (2009, p.4), these three methodological movements are associated with basic cultural differences and special interests among groups of researchers. This statement follows Kuhn's argument (1996, p.46), which considers scientists' work to be influenced by the means of education or literature to which they are exposed.

In this article, we seek to present the possibility of adding quantitative aspect to qualitative methods in the philosophical supposition of the combination of the quantitative and qualitative approaches, in the complementarist perspective, in the sense of Bohr (1995, p.51). We also consider the combination of qualitative and quantitative methods in a single study to be widely practiced and accepted in many areas of health research (Sale et al, 2008, p. 364). This is why the authors of this article propose the use of Non-Parametric Statistical Tests as an alternative resource to add testability to the Didactic Engineering method.

Didactic Engineering is characterized as being a qualitative action research method, and rejects the Classical Parametric Statistics method, the control case or experimental groups and control groups. The approach is comparative. The analyses are performed by comparing the expectations, experimentation, and results and, thus, are based on an internal validation. These characteristics allow the use of Non-Parametric Statistics, which does not require a population model, does not require rigorous or numerous hypotheses about the parameters, and can be applied to small samples. In the scope of Non-Parametric Statistics, and for example, the Wilcoxon (Before and After) test was considered, since it more closely adheres to the paradigm of Didactic Engineering. To facilitate the calculation procedures, the use of an open statistics program (R-Project) was proposed.

Figure 1 presents a flow diagram of the iterations of the action research process of Didactic Engineering.

The Wilcoxon test, like all classical statistical tests, allows for the rejection of the tested hypothesis. The expectation in this article is that this test aggregates a functionality to the Didactic Engineering methodology, making it a more efficient method for handling small samples, meeting the prerogative of falsifiability in the scientific method, that is, "*the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability*" found in page 60 of [3]. Figure 2, an adaptation of the flow diagram found in page 5 of Spagnolo [5], presents a summary of these ideas.

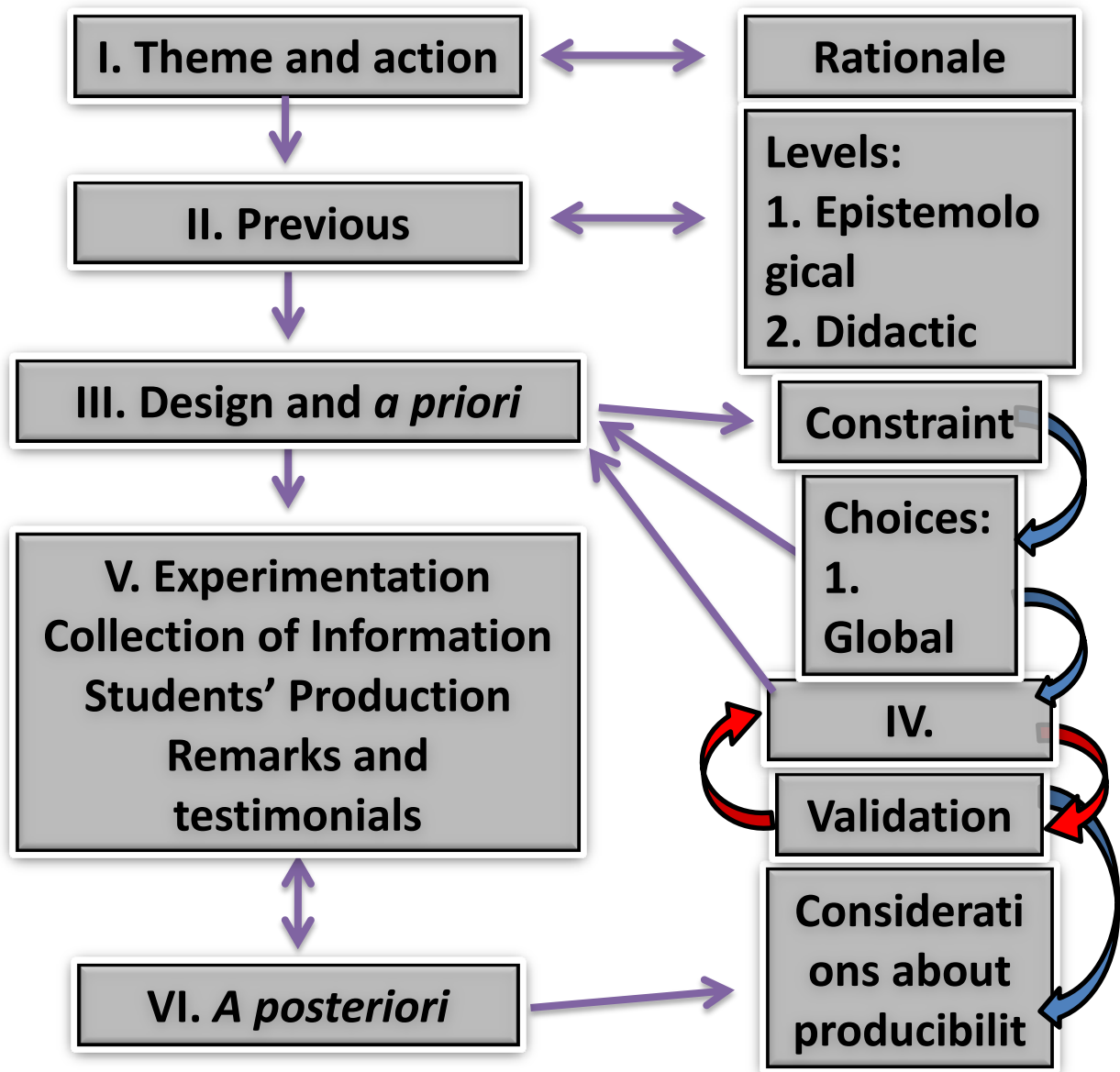


Figure 1: Flow diagram of the Iterations of the Didactic Engineering Process

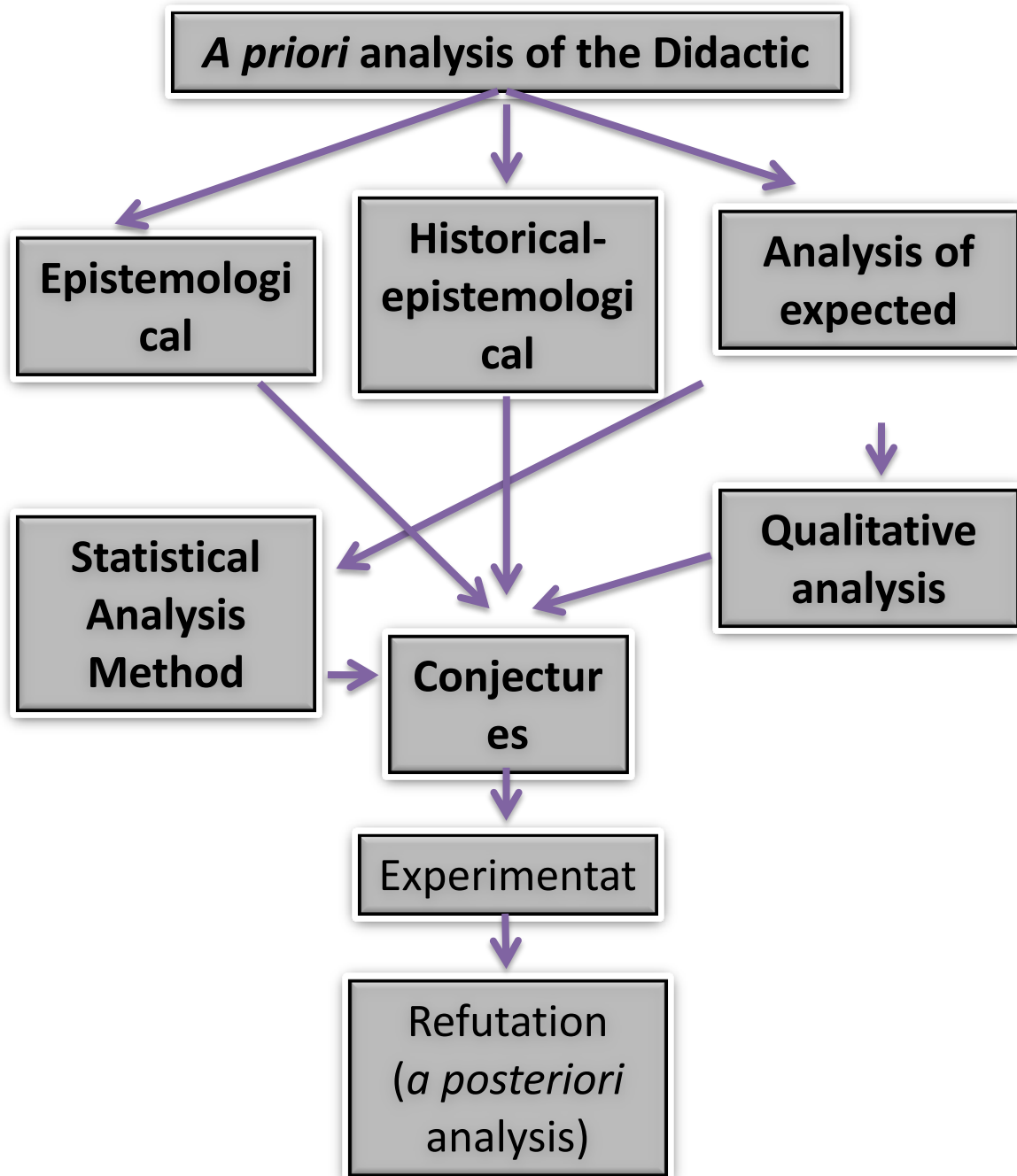


Figure 2: Aggregation of the Wilcoxon Test to Didactic Engineering

Statistical models, according to Spagnolo [5], provide research in Mathematics Education with the possibility of transferring successful experiences from other areas. Nevertheless, a deeper theoretical understanding is necessary so that the use of these models can provide benefits. A comprehensive study that considers the various statistical approaches is necessary to obtain reliable results.

Research in Didactics makes it a paradigmatic target in relation to the other research paradigms in education sciences, in which both the paradigm of this discipline, the object of the analysis, and the paradigm of experimental sciences are used. Research in Didactics can be considered a type of experimental epistemology. Didactic Engineering, one of the fundamental tools of research in Mathematics Education, is the *a priori* analysis of a didactic situation, which signifies the analysis of the epistemological representations, the epistemological history, and the behavioral expectations found in pages 2-3 of [5].

## 2. The Problem

It is a fact that, in most cases, when the Didactic Engineering methodology is applied, the sample size is small, with fewer than 30 observations, for example, and a nominal or ordinal variable or intervals are used for the measurement levels. Considering that the sample size is small, asymptotic results of Classical Parametric Statistics cannot be used.

In Didactic Engineering, the *a priori* analysis is important in identifying the relationships among the research variables. The supposed relationships are based on prior knowledge or the suspicions of the teaching and learning conditions, as affirmed by Spagnolo et al. [6]. Here, we will concentrate on the application of tests to verify the quality of the experimentation by comparing the results of a Before analysis, in the scope of the *a priori* analysis, and an After analysis, in the scope of the *a posteriori* analysis.

The Wilcoxon test in the computing environment of the open statistics program R ([www.r-project.org](http://www.r-project.org)) makes Didactic Engineering a falsifiable scientific theory from Popper's viewpoint.

## 3. Didactic Engineering and Quantitative Complementarity

The theory of Didactic Engineering was elaborated in an analogy between the actions of Mathematical Didactics and those of Engineering. That is found in page 283 of [1]:

*The notion of Didactic Engineering emerged in mathematical didactics in the early 1980s. The idea was to use this term to label a type of work: the work that is comparable to the work of the engineer who, to produce a precise design, relies on the scientific knowledge under his command, accepts being subjected to scientific control, but, at the same time, finds himself obliged to work on objects that are much more complex than the pure objects of science, and thus, must practically approach, with all the means at his disposal, problems that science does not want to or cannot yet undertake.*

These considerations by Artigue about Didactic Engineering take on the label of a process for solving practical problems in Mathematics Education through scientific methods, and bring, in their origin, a proposal for an open methodology; that is, an action research about the teaching systems in the scope of the research methodology in this area.

As Spagnolo states, found in page 1 of [6]:

*In the research on Mathematics Didactics, and more generally in Humanities, various researchers use qualitative analyses to experimentally falsify hypotheses that have been formulated a priori, i.e., in a direction contrary to the investigation. This methodological approach applied to research is shown in most cases to be insufficient for analyzing all the variables involved in the contingent phenomena of teaching/learning, although in some cases (ad hoc analysis of protocols, videos, etc.), it allows some interesting relationships to be detected. However, if the number of subjects becomes very great, the*

*qualitative analysis can no longer extract all the relationships existing between the variables involved. A quantitative analysis based on statistics will prevail and will be completed by a qualitative analysis, essential for a contextual interpretation.*

Later in this text, we will review the general concepts and characteristics related to the methodology of Didactic Engineering, based and found in pages 281-307 of Artigue [1], Spagnolo [5], and Spagnolo et.al. [6].

Didactic Engineering is characterized, in relation to other types of methodologies, by experiments in class, action research, and by the validation methods that are associated with it. Indeed, research that makes use of the experimentation in class is most often situated in a comparative approach with external validation based on the statistical comparison of procedures of experimental groups and control groups. This paradigm is not that of Didactic Engineering, which lies in another direction, in registering the case studies in which validation is essentially internal, based on the comparison of the *a priori* analysis with the *a posteriori* analysis. However, in our view, the theory of Didactic Engineering in this form does not present itself in a manner that enables it to satisfy the criterion of Popper's scientific status, since it leaves out the use of resources of statistical methods of internal validation. In this case, there is, for example, the Wilcoxon (Before and After) test method of Non-Parametric Statistics, methods that favor Popper's criterion. The proposal to use Non-Parametric Statistics also has a complementarist perspective, without loss of the qualitative properties that are well-defined in Didactic Engineering.

Popper states that the observation is always selective; it requires a determined object, a defined task, an interest, a point of view, and a problem. The objects can be classified, taken as similar or dissimilar, related according to the needs and theoretical interests of the problem under investigation, of the conjectures and anticipations and accepted theories as a backdrop, of its set of references, of its horizon of expectations.

Artigue suggests that in Didactic Engineering, the conception phase occurs based on a general didactic theoretical framework, on the didactic knowledge acquired in the domain being studied, but also on a certain number of preliminary analyses: 1) the epistemological analysis of the content covered by teaching; 2) the analysis of the usual teaching methods and its effects; 3) analysis of the students' understanding and of the difficulties and obstacles that mark its progress; 4) the field analysis of restrictions in which the didactic realization will be situated; and 5) consideration of the specific objectives of the research. In the preliminary analyses, the researcher is concerned with explanations. Although it is not visible at the level of publications, there is an interactive and selective process performed by the researcher that serves as a basis for the conception of the engineering, it is revised and deepened throughout the different phases of the work, as a result of the needs that arise, and they are, therefore, preliminary only at the first level of elaboration.

At the *a priori* analysis phase, Artigue states, found in page 291 of [1], that the researcher makes the decision to act on a certain number of system variables that are not determined by the constraints: command variables that he/she assumes are pertinent variables in relation to the problem being studied. In order to facilitate the engineering analysis, it seems useful to us to distinguish the two types of command variables: the macro-didactic, or global, variables related to the global organization of the engineering, and the micro-didactic, or local, variables related to the local organization of the engineering; i.e., the organization of a section or a phase, with a few of them possibly being general order variables or variables dependent on the didactic content that is being taught. One original feature of the Didactic Engineering method, as highlighted earlier, resides in its validation method, essentially internal. This validation process occurs from the

conception phase, via an *a priori* analysis of the didactic situations of engineering, strictly linked to its local conception.

The *a priori* analysis is necessary to design an analysis of the control of the meaning: very schematically, if the constructivist theory places the principle of engaging the student in constructing his/her knowledge through the intermediary of the interaction with a certain *milieu*, the Theory of Didactic Situations, which serves as a reference to the engineering methodology, has had, since its beginnings, the goal of making itself a theory for the control of the relationships between meaning and situations. The objective of the *a priori* analysis is, therefore, to determine to what degree the choices made allow the control of the students' behaviors and their meanings. Thus, it will be based on hypotheses, and it is these hypotheses for which the validation will be, in principle, indirectly in play, in the comparison performed in the fourth phase between the *a priori* analysis and the *a posteriori* analysis.

Another phase of Didactic Engineering is the experimentation, which, according to Artigue, is classical. For this researcher, this phase is followed by the so-called *a posteriori* analysis phase, which is based on the set of data collected in the experimentation: observations made during teaching sessions, but also the students' production, both in class and out. The information is often completed by other data obtained by using other research tools: questionnaires, individual interviews, or in small groups realized at various times while teaching or afterwards. In addition, as we have indicated, the comparison of the two analyses, the *a priori* analysis and the *a posteriori* analysis, is essentially the basis for the validation of the research hypotheses. According to Artigue, the process of internal validation that is in play here does not fall into the usual trap of statistical validations, associated with experimentation in class, which consists of being implicitly based on the principle that the measurable differences observed are associated with the command variables upon which we build to differentiate experimental classes and control classes. In the previous paragraph, we did not raise certain questions associated with internal validation, but rather those associated with the cited criterion for Popper's scientific status.

Popper states that there is a tendency in the dogmatic attitude to verify and confirm our laws, and in this way, to neglect the refutations. In the critical attitude, there is a distinction in relation to the dogmatic attitude; the critical attitude is translated by the modification of the laws and schemas, and, when possible, to test and refute. Thus, the critical attitude is identified by the scientific attitude, while the dogmatic attitude is a characteristic of a pseudo-scientific attitude, a primitive attitude. This primitive attitude has a logical function since the critical attitude needs to have revision as its starting point. For Science, it should begin by criticizing the dogmatic attitude and the myths, and not by the observations nor by experiments. The tradition of the scientific attitude is necessarily critical because when its theories are transmitted, the critical attitude towards them is also transmitted. The goal of having an attitude that is free from the discussion of the theories is to discover their weaknesses, in the sense of improving them, and these weak points can only be found in the most remote logical consequences from which they can be derived. The trial and error method or the conjecture and refutation method is a rational procedure of the task of testing theories.

As Spagnolo [5] states, in a semiotic perspective, the analysis of knowledge of a discipline allows the content to be managed in relation to communication difficulties and misunderstandings of such content. This position is not very original regarding the Human Sciences, but it represents a true innovation for the technical and scientific disciplines. In any case, a didactic situation constitutes a problem for the student to resolve, such as a traditional problem (that is, in the



scientific or mathematics framework), or a strategy for better organizing the knowledge to be adapted to a situation.

A good didactic word problem is one that, in terms of building knowledge, permits the best formulation in ergonomic terms. The didactic situation variables are all the possible variables that are involved in the pedagogical process. The didactic variables are those that allow a change in the students' behaviors. The didactic variables are, therefore, a subset of the didactic situation variables.

We can conjecture and test our hypotheses by means of confidence models and the development of experimental research, as Spagnolo [5] defends. The author explains that by confidence models we mean the models that allow the possibility of making predictions about the didactic phenomena. A good model should permit the communication of hypothetical results (reliable, but refutable, conjectures) of the research for the community of teachers, by means of a solid argument, such as an *a priori* analysis and statistical tools. It should be capable of foreseeing conclusions.

As Spagnolo argues, the problems related to communication in a specific discipline may include:

- Preparation of the appropriate didactic situations;
- The analysis of the errors and obstacles derived from the communicative processes;
- Tools for a deeper and better understanding of the communicative processes;
- Tools for the preparation of didactic situations;
- A falsifiable hypothesis that can be subjected to an experimental test and subjected to proof by systematic attempts to discover its faults.

#### **4. Didactic Engineering as a Non-Parametric Statistic**

As we have seen, according to Artigue, the internal validation process does not use the statistical validations associated with experimentations in class. However, the internal validation process of Didactic Engineering has its equivalence in the scope of Statistics. Dependent samples is the equivalent concept, since with dependent samples, or paired samples, or linked samples, we obtain two values, a Before value and an After value, for individuals who present the same characteristic.

According to Spagnolo [5], didactic research leads us to collect elementary information, which, in general reveals a student's behavior in a situation. In this way, the statistical data are made up of student, situation, and behavior. Spagnolo [5] refers to the differences in the use of statistics by the teacher and by the researcher. For him, the teacher must make many decisions and must make them quickly, so as to be able to correct them if they are not adequate. He also considers that the teacher cannot wait for the result of the statistical treatment for all of his/her questions. The researcher, according to Spagnolo, follows an inverse process, since he/she should seek to understand which hypotheses correspond to questions of interest, which data should be collected, which treatments should be used, and what the conclusions are.

#### **5. Construction of a Hypothesis Test Appropriate for Didactic Engineering**

Gillies [2] states that Popper's approach has been strongly justified by the practice of Classical Statistics. In the Statistical Implicative Analysis algorithm, for Spagnolo et. al. [6], a test for hypotheses is constructed that follows the theory of Classical Statistics, in a parametric approach. In the construction of this test, the Binomial probability distribution is mainly used with parameters  $n$  and  $p$ ; the Poisson probability distribution with a single parameter is also used, as well

as a distribution convergence model, i.e., a Gaussian asymptotic model. We consider that the parametric tests and the Statistical Implicative Analysis are better suited to *macro-didactic engineering*, cases in which we have a large volume of data.

As we observe, the parametric tests require suppositions about the nature or form of the distributions involved and a large number of data; the non-parametric methods do not depend on such requirements. Thus, the non-parametric hypothesis tests are better suited to *micro-didactic engineering* in which the number of data is small. In addition, as Artigue [1] affirms, the investigations in *micro-didactic engineering* are easier to initiate in the classroom.

For testing a hypothesis, Classical Statistics, and in particular Non-Parametric Statistics, adopts a specification of a rejection region and considers the hypothesis being tested as refuted if the value observed from the statistical test is included in the rejection region. There always exists a confidence level, i.e., a probability that the tested hypothesis is true. For example, the Wilcoxon (Before and After) test of Classical Non-Parametric Statistics can at the same time be considered a methodology that is equivalent to the falsification methodology defended by Popper.

## 6. Didactic Engineering and the Wilcoxon Test

Since Mathematics Education has aspects of a behavioral science and since the methodology of Non-Parametric Statistics and its techniques of hypothesis testing, in particular the Wilcoxon test, are applied to the data of Social Sciences, Psychology, and Behavioral Sciences [4], we therefore also consider that the Wilcoxon (Before and After) test of paired samples, i.e., two samples that are dependent, in the sense that the data values correspond to the pairs, is a suitable test for aggregating the internal validation of Didactic Engineering, of the comparison of the two analyses, *a priori* and *a posteriori*, the *falsifiability*, or *refutability*, or *testability* defended by Popper. Figure 3 synthetically illustrates our proposal.

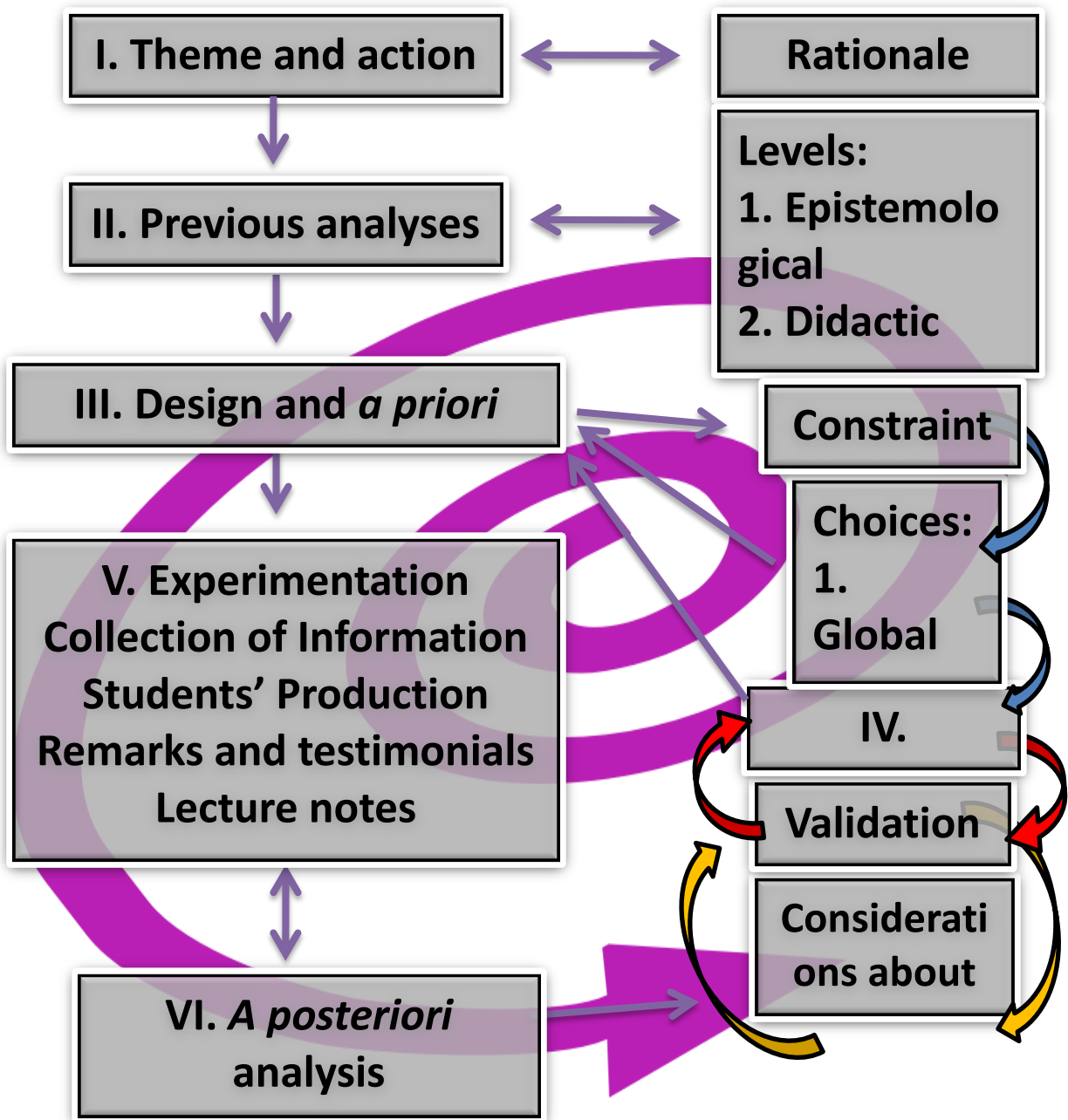


Figure 3: General View of our Proposal

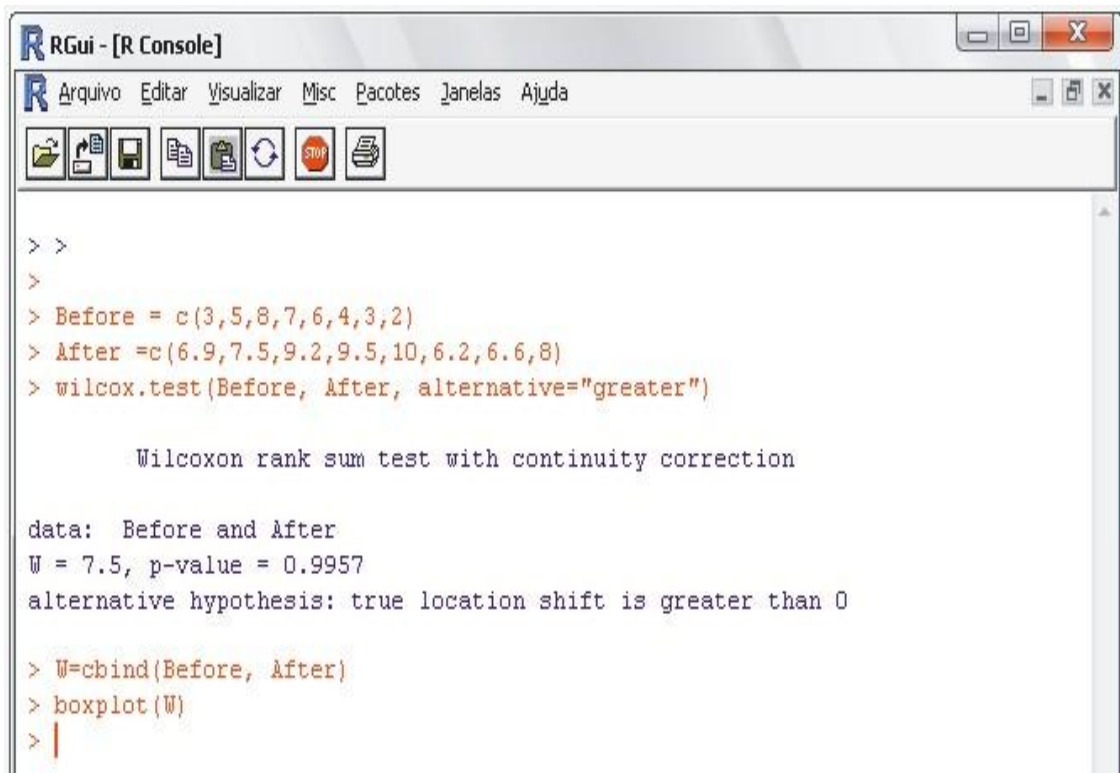
As Siegel defined, “The Wilcoxon test is extremely useful for behavioral scientists. With data about behavior, it is not rare to find cases in which the researcher can: a) say which member of a pair is ‘greater than’ the other; and b) display the differences based on the order of its absolute value. That is, the researcher can perform a ‘greater than’ judgment between the results  $d$  – the difference between Before and After – and any pair, as well as perform this judgment in relation to the differences relative to any two pairs. With such information, the researcher can apply the Wilcoxon test” found in page 84 of [4]. Our interest in the Wilcoxon test is not in its theoretical

aspect developed in Siegel, but rather in its application in Didactic Engineering. To achieve this objective, we use an algorithm represented by the function “`wilcox.test`” present in the open program R ([www.r-project.org](http://www.r-project.org)).

We present below the mechanics of the “`wilcox.test`” function by Verzani [7], by means of a fictitious example in which Before and After are represented by the line vectors defined below in the data entry format for the program R.

```
Before = c(3,5,8,7,6,4,3,2)
After = c(6.9,7.5,9.2,9.5,10,6.2,6.6,8)
```

To illustrate, we present a simulation that involved two possibilities: less than or greater than. The Before and After vectors can represent indices numbers, that is, a summary measure of various values observed in time and space, for example the Human Development Index (HDI). We are interested in observing the indices numbers that result from *a priori* analysis (Before) and *a posteriori* results (After) and comparing them to know if there is a difference, or rather, if the After situation is significantly greater than the Before situation. To make such a determination, we will use the `wilcox.test` function in the program R. We present below, in Figures 4 and 5, the data entry and the outputs of R:



```
RGui - [R Console]
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda

> >
>
> Before = c(3,5,8,7,6,4,3,2)
> After = c(6.9,7.5,9.2,9.5,10,6.2,6.6,8)
> wilcox.test(Before, After, alternative="greater")

      Wilcoxon rank sum test with continuity correction

data:  Before and After
W = 7.5, p-value = 0.9957
alternative hypothesis: true location shift is greater than 0

> W=cbind(Before, After)
> boxplot(W)
> |
```

Figure 4: *Handling of the R System*

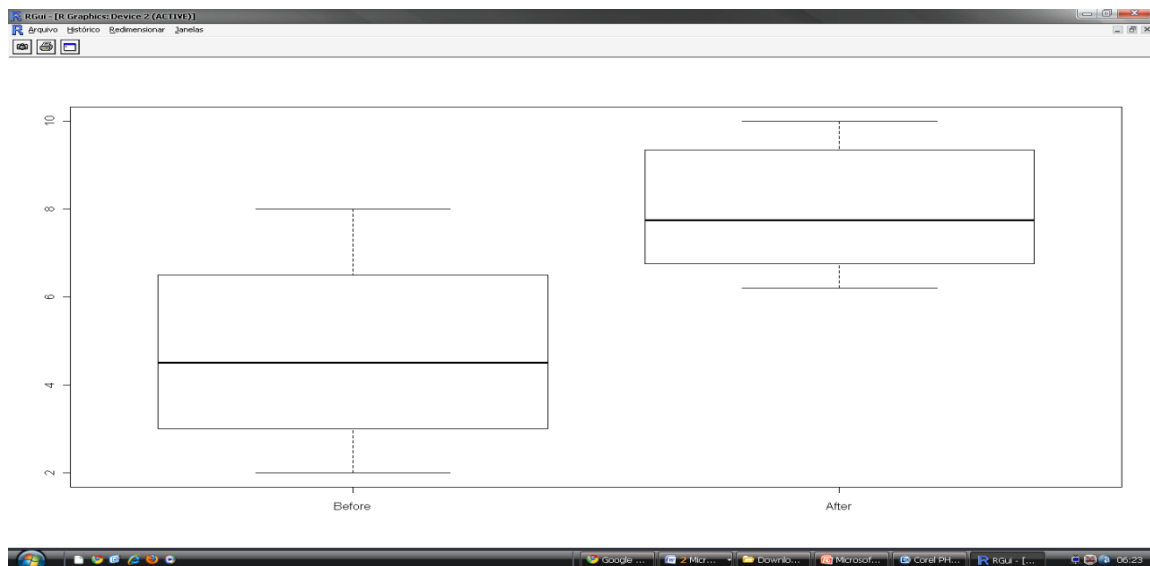


Figure 5: Box Plot Generated by R

The function `wilcox.test` from the program R tested the following hypothesis: the difference between the Before and After situation is equal to zero, i.e., the experiment applied did not produce a positive improvement on the Before situation, as opposed to the alternative hypothesis that the difference between Before and After is greater than zero. Since the significance level of the test is  $p\text{-value} = 0.9957$ , this result indicates that the experimentation produced a significant effect in the After situation. Likewise, when we analyze the boxplot we see that there is also a clear difference between the situations. To improve the analysis of the data, we can also use dendrograms. Figure 6 shows how we can make use of R to plot dendrograms.

```

R Console
> After=c(6.9,7.5,9.2,9.5,10,6.2,6.6,8)
> PerformanceMatrix<-matrix(c(Before,After),nrow=8,ncol=2)
> PerformanceMatrix
      [,1] [,2]
[1,]    3  6.9
[2,]    5  7.5
[3,]    8  9.2
[4,]    7  9.5
[5,]    6 10.0
[6,]    4  6.2
[7,]    3  6.6
[8,]    2  8.0
> InitialStatus<-hclust(dist(Before),"comp")
> EndStatus<-hclust(dist(After),"comp")
> Performance<-hclust(dist(PerformanceMatrix),"comp")
> plot(InitialStatus,ylab="Range",main="Initial Status")
> plot(EndStatus,ylab="Range",main="End Status")
> plot(Performance,ylab="Range",main="Performance")
>
    
```

Figure 6: Data Input Screen in R to Generate Dendrograms

The dendrograms above show a relationship of similarity between the fictitious students. By comparing Figures 7 and 8, we can observe that after undergoing the engendered didactic situation, there was a greater homogenization of the class performance. A rearrangement in the results also occurred after the realization of the didactic activities.

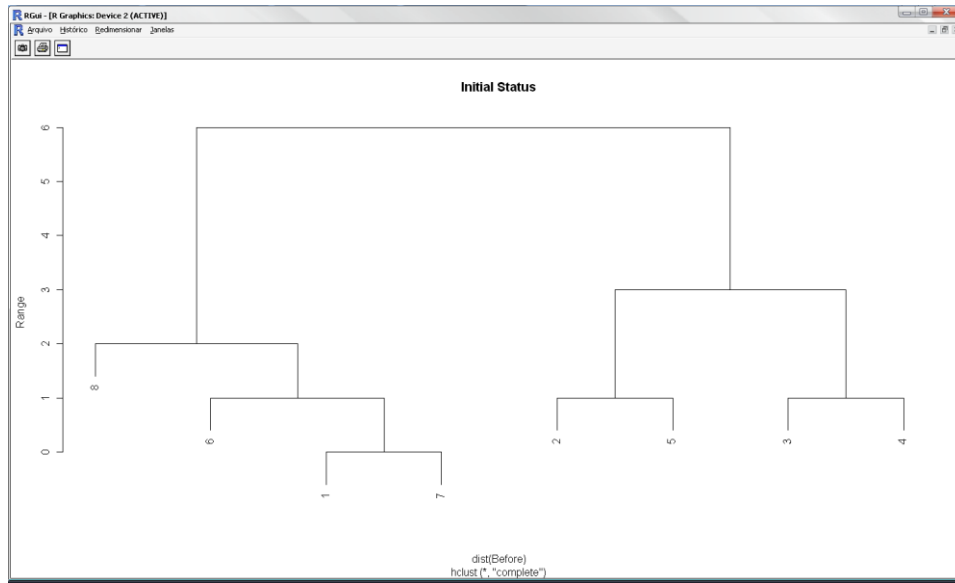


Figure 7: *Initial Status*

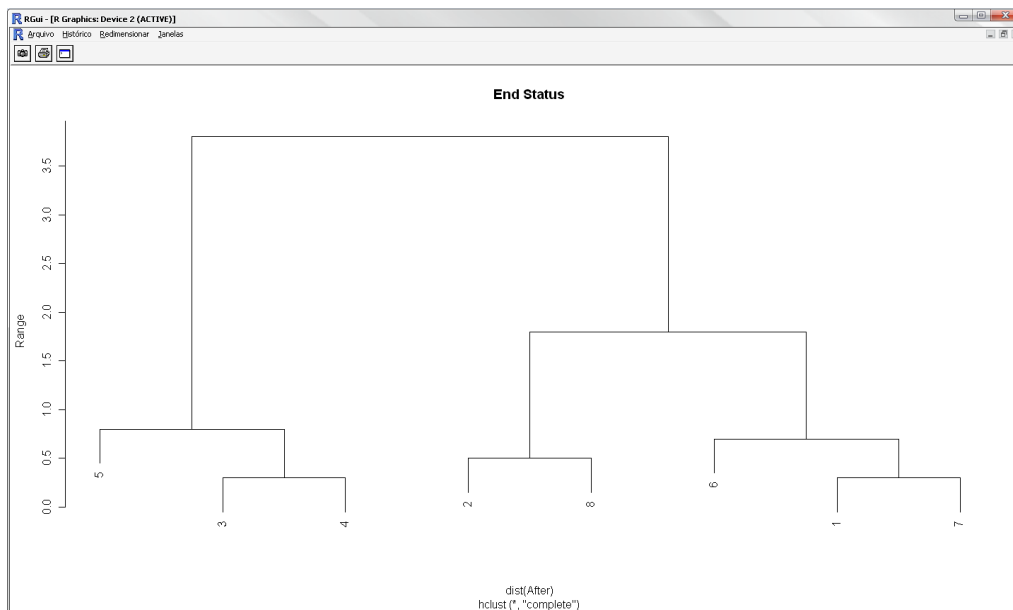


Figure 8: *End Status*

In relation to the performance of the students during the pedagogical process, we observe that students 3 and 4, who correspond to the pairs (8;9.2) and (7;9.5), respectively, show the same behavior; students 1 and 7, who correspond to the pairs (3;6.9) and (3;6.6), respectively, also show, with different measures, similar behaviors. The most important result captured by the dendrogram is student 5, who corresponds to the pair (6;10); this student, who was subjected to the Didactic Engineering methodology, obtained the maximum grade in the second evaluation after the experiment.

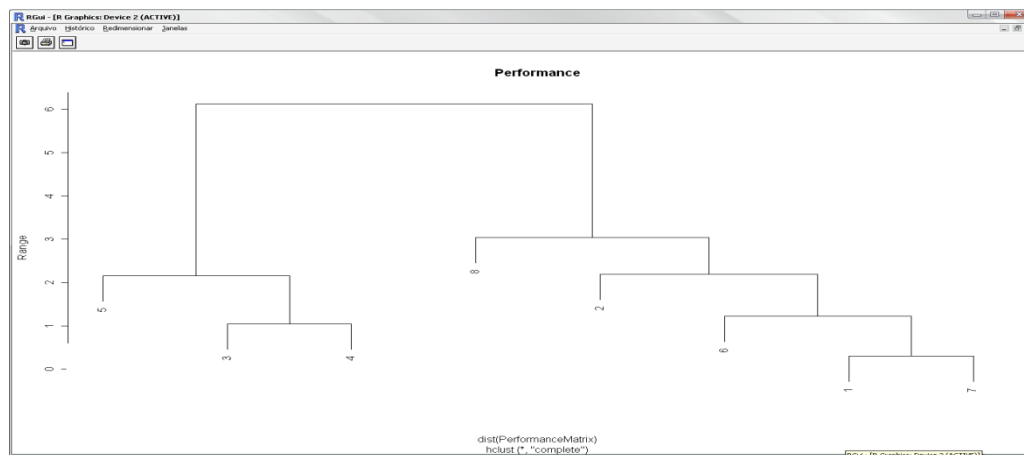


Figure 9: *Performance*

We have presented above a simulation to show how simple the use of the Wilcoxon test is in the open program R, and a descriptive analysis using a dendrogram to identify some implicative relationships. We can aggregate the Wilcoxon tests and the descriptive analysis using the dendrogram to the methodology of Didactic Engineering. We recall that the following phases are developed in the Didactic Engineering methodology: 1) Preliminary analyses; 2) Design of an *a priori* analysis of the didactic engineering situations; 3) Experimentation; and 4) A *a posteriori* analysis and evaluation.

Our proposal is to introduce, in phase 2, an evaluation that we will call Before, and in phase 4, another evaluation that we will call After. In Didactic Engineering, there is a comparison of the two analyses (*a priori* and *a posteriori*) that is based essentially on the validation of the hypotheses involved in the investigation. With the application of the Wilcoxon test, we add the property of falsifiability, or refutability, or testability to Didactic Engineering, according to the criterion of Popper's scientific status.

## 7. Conclusions and Final Considerations

This theoretical proposal, quantitative complementarity through Non-parametric Statistics, does not signify failing to recognize its great value and the power of explanation that it has in the scope of Mathematics Education. We also consider that quantitative arguments should be used in Mathematical Education research to go beyond descriptive analysis of observed facts. A great power of explanation of Didactic Engineering is in the prior analyses, as Artigue indicated: epistemological analysis of the contents aimed at teaching; the analysis of the usual teaching and of its effects; the analysis of the students' conceptions, of the difficulties and obstacles that mark their evolution; and the field analysis of constraints in which the effective didactic realization and the investigation will be found. Furthermore, the *a priori* analysis, in which the hypotheses are formulated that will be verified in the *a posteriori* analysis, is just as important as the prior analyses.

What Popper proposes is that we need to test our theories so that we can learn from our mistakes and better know the objectives of our study, and we consider, that as scientists, we are not seeking highly probable theories, but rather explanations.

The descriptive analysis of the dendrogram explains the behavior of the data in the experimentation process and the Wilcoxon Test added to the Didactic Engineering methodology

brings to this methodology an aspect of testability of internal validation without loss of its great power of explanation. The use of the free open program R makes the application of the test automatic, and democratizes and disseminates the information. The use of an open program, such as R, makes its use possible for a broader public, and thus it is a diffuser of technologies, thereby aggregating this addition value.

## 8. References

- [1] Artigue, M. (1990). Ingénierie didactique. *Recherches en Didactique des Mathématiques*, 9 (3), 281-307.
- [2] Creswell, J. W. *Projetos de pesquisa. Métodos qualitativo, quantitativo e misto*. Porto Alegre: Artmed, 2009.
- [3] Bohr, Niels. *Física atômica e conhecimento humano: ensaios 1932-1957*, Rio de Janeiro, Contraponto Editora LTDA, 1995.
- [4] Giere, R. N. *Explaining Science: A cognitive approach*. Chicago: University Chicago Press, 1988.
- [5] Gillies, D. (2009). *Philosophical Theories of Probability*. New York: Routledge.
- [6] Kuhn, T.S. *The Structure of Scientific Revolutions*, 3<sup>rd</sup> ed., University of Chicago Press, Chicago 1996.
- [7] Otte, M. *Complementarity, Set and Nombres. Educational Studies in Mathematics*. 53: 203-228, Printed in the Netherlands: Kluwer Academic Publisher, 2003
- [8] Popper, K. A. (2003). *Conjecturas e Refutações*. Translation by Benedita Bettencourt. Coimbra: Livraria Almedina.
- [9] Sale, J.E.M., Lohfeld, L.H., Brazil, K.. *Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research*, p. 363-374, In: The mixed methods reader/ Vicki L. Plano Clark, John W. Creswell, SAGE Publications, Inc., California, 2008
- [10] Siegel, S. (1977). *Estatística não-paramétrica (para as ciências do comportamento)*. Rio de Janeiro: Editora McGraw-Hill do Brasil.
- [11] Spagnolo, F. (2005). L'Analisi Statistica Implicativa: uno dei metodi di analisi dei dati nella ricerca in didatticadelle Matematiche. *Troisième Rencontre Internationale A.S.I. (Analyse Statistique Implicative)*. Palermo, Italy, 35-51.
- [12] Spagnolo, F., Gras, R., & Régnier, J. C. (2007). Une mesure comparative en didactique de mathématiques entre une analyse a priori et la contingence. Author's manuscript published in *IV Rencontres internationales d'Analyse statistique implicative*. Castellon: Espagne.
- [13] Teddlie, C. & Tashakkori, A. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. SAGE Publications, Inc. California, 2009
- [14] Verzani, John. (2010). Simple R [WWW page]. URL <http://www.r-project.org>, accessed on 02/15/2010.