

Universidade do Minho
Instituto de Educação

Abílio António Freitas Belo

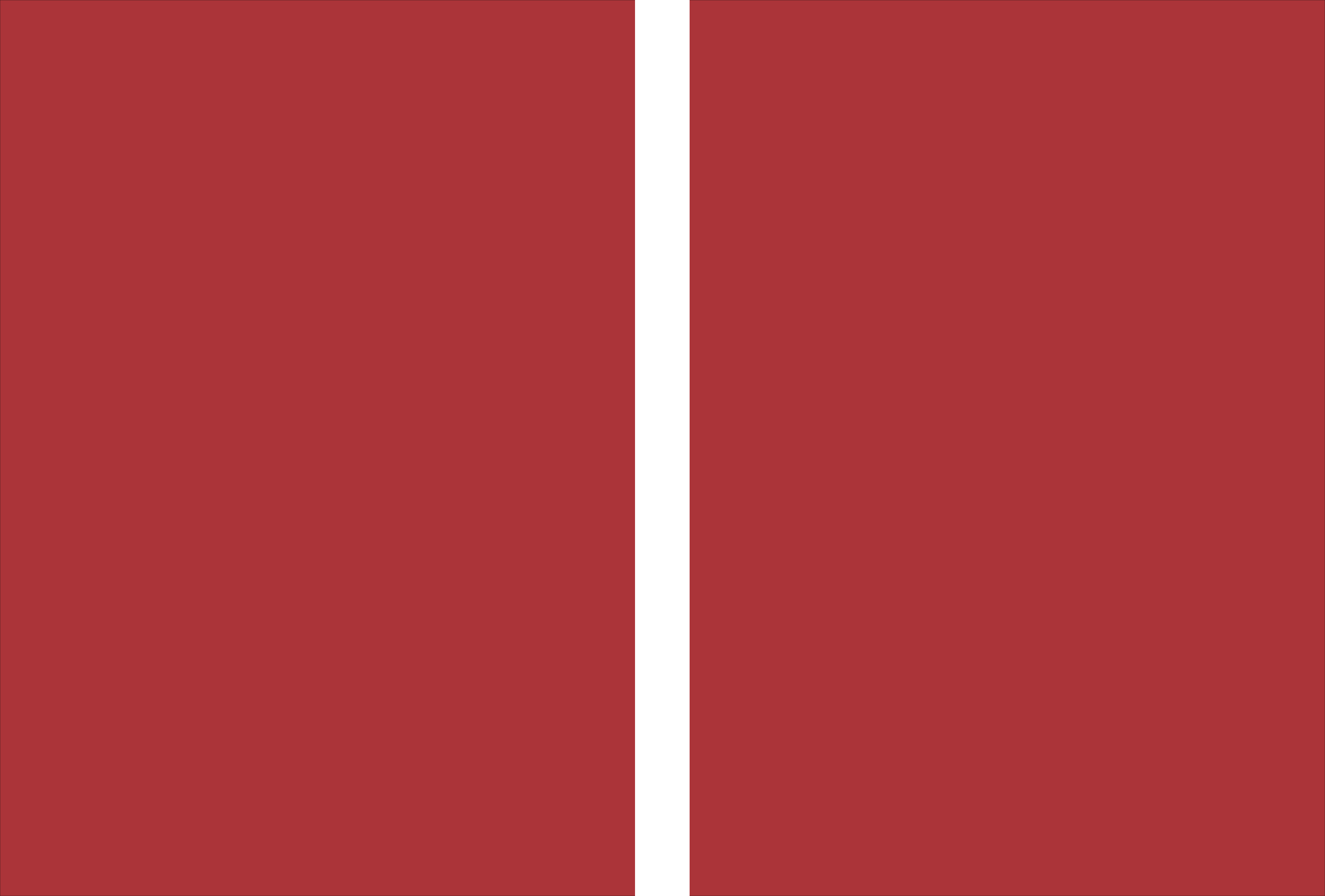
**The evaluation of teachers professional
performance at Universidade Nacional
De Timor Lorosae [UNTL]**

The evaluation of teachers professional
performance at Universidade Nacional
De Timor Lorosae [UNTL]

Abílio António Freitas Belo

UMinho | 2018

September, 2018





Universidade do Minho
Instituto de Educação

Abílio António Freitas Belo

**The evaluation of teachers professional
performance at Universidade Nacional
De Timor Lorosae [UNTL]**

Doctorate thesis in Ciências da Educação
Desenvolvimento Curricular

Supervisor:
Professor Doutor José Augusto Pacheco

September, 2018

DECLARATION

Name: Abílio António Freitas Belo

E-mail: abilbel@yahoo.com.br

Passport number: N739392

Title of the PhD Thesis: The evaluation of teachers professional performance at Universidade Nacional De Timor Lorosae [UNTL]

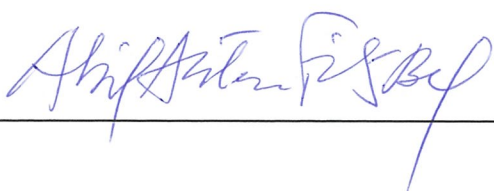
Supervisor: Professor Doutor José Augusto Pacheco

Conclusion year: 2018

Doctorate Designation: Ciências da Educação - Especialidade de Desenvolvimento Curricular

The partial reproduction of this thesis is only authorized for research effects, through a written statement by the interested party that it is committed to them.

University of Minho, 28/09/2018

Signature:  _____

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of thesis elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of University of Minho.

University of Minho, 28 de Setembro de 2018

Full name: Abilio Antonio Freitas Belo

Signature : Abilio Antonio Freitas Belo

THANKSGIVING

First of all, I want to convey thanks for the gift given by God so that I can complete this thesis well and smoothly. This thesis is the most important part of the research conducted with students and lecturers from the Universidade Nacional de Timor Lorosae in the field of doctoral research majoring in Education and specializing in curriculum development, delivered to the Institute of Education of the University of Minho.

Secondly, I also want to express my special thanks and sincere acknowledgment to:

- Professor Doctor José Augusto Pacheco, for his sustainable guidance at the time I started registering as a doctoral student until I ended this thesis well and accurately as expected by the Institute. His guidance really helped me in my independence and fostered my intellectual abilities until I reached the completion of this thesis.
- The University of Minho which creates appropriate institutional conditions and unlimited assistance for me to be able to carry out this research.
- My colleagues from the Institute of Education Department who have directly or indirectly contributed to the achievement of my thesis work.
- Thank you also the Institution of UNTL representing the Government of Timor Leste allowing me to take my doctoral degree.
- Lecturers and students of Universidade Nacional Timor Lorosae who actively participated in filling in the list of questions received and also involved in the interviews that I individually conducted with them.
- Sister Natalia Fernandes who is always ready to help me in every work and business related to the research process and thesis writing from the beginning until the completion of my thesis.
- My family, my wife Cecilia Lomuntad Belo and my son Patrick Belo for their moral support and love and many other types of support that I cannot mention one by one.
- For my sister-in-law, Judith da Costa Marçal and my niece Angela Marçal Belo and Hilaria Marçal Belo who were both, directly and indirectly, contributing ideas and moral support in finalizing this research.

For all, thank you very much.

Evaluation of teachers' professional performance at Universidade Nacional de Timor Lorosae [UNTL]

Abstract

The present study evaluated teacher performance in Universidade Nacional de Timor Lorosae. The tool measures four areas, namely, pedagogical, professional, personal and social competencies. The purpose of the study was to (1) determine how students' (gender, age, and faculty) and teachers' (age and experience, academic degree, and faculty) background influence the evaluation of teacher performance; (2) the difference between the students' evaluation and teachers' self-evaluation on the teacher performance was also determined; and (3) the implementation of the Teaching Performance Evaluation (TPE) in the educational institution. The teacher performance appraisal was administered to 342 students who rated their teachers and 192 faculty who evaluated themselves. The quantitative analysis was conducted by determining the means of the teacher performance and classified by age, gender, years of experience, and educational attainment. The chi-square was used to determine the association of the background of the students and teachers on the results of the teacher performance assessment. The chi-square was also used to determine the difference between student and teacher evaluation. The results showed that students' background such as gender and faculty where the teacher belongs influence teachers' performance. The teachers' background such as years of experience and academic degree influences their self-evaluation. There was a significant difference on students' evaluation and teachers' self-evaluation on the performance. The qualitative analysis showed that teachers recognize the importance of TPE through its contribution, regulations, and implementation. Moreover, teachers' attitude on evaluation is shaped by obstacles and recommendations. Lastly, teacher performance is improved through formative assessment and sample practices.

Keywords: East Timor; evaluation; professional development; public higher education; student's assessment; teacher performance evaluation.

Avaliação do desempenho profissional dos professores na Universidade Nacional de Timor Lorosae [UNTL]

Resumo

O presente estudo avaliou o desempenho dos professores na Universidade Nacional de Timor Lorosae. A ferramenta mede quatro áreas, nomeadamente competências pedagógicas, profissionais, pessoais e sociais. O objetivo do estudo foi (1) determinar como as características dos alunos (sexo, idade e departamento) e dos professores (idade, experiência, grau acadêmico e departamento) influencia a avaliação do desempenho do professor; (2) a diferença entre a avaliação dos alunos e a autoavaliação dos professores sobre o desempenho do professor também foi determinada; e (3) a implementação do Teaching Performance Evaluation (TPE) na instituição educacional. A avaliação de desempenho do professor foi administrada a 342 alunos que avaliaram seus professores e 192 professores que se autoavaliaram. A análise quantitativa foi realizada determinando as médias do desempenho do professor e classificadas por idade, sexo, anos de experiência e nível de escolaridade. O qui-quadrado foi utilizado para determinar a associação dos antecedentes dos alunos e professores com os resultados da avaliação de desempenho do professor. O qui-quadrado também foi usado para determinar a diferença entre avaliação do aluno e do professor. Os resultados mostraram que os antecedentes dos alunos, como sexo e departamento, em que o professor pertence, influenciam o desempenho dos professores. As características dos professores, como anos de experiência e grau acadêmico, influencia sua autoavaliação. Houve uma diferença significativa na avaliação dos alunos e na autoavaliação dos professores sobre o desempenho. A análise qualitativa mostrou que os professores reconhecem a importância do TPE por meio de sua contribuição, regulamentos e implementação. Além disso, a atitude dos professores em relação à avaliação é moldada por obstáculos e recomendações. Por último, o desempenho do professor é melhorado através de avaliação formativa e práticas de amostra.

Palavras-chave: Timor Leste; avaliação; desenvolvimento profissional; ensino superior público; avaliação do aluno; avaliação de desempenho de professores.

Index

INTRODUCTION.....	1
Contextualization	1
Delimitation of the study (the educational evaluation field)	3
Relevance, contribution and pertinence of the study	13
Research question and objectives	14
Objectives.....	15
Formulation of Hypotheses	16
The structure of the study	16
CHAPTER I – REVIEW OF THE RELATED LITERATURE.....	19
1.1. The concept of evaluation	19
1.2. Historical perspectives on the evaluation concept.....	24
1.2.1. Evaluation and objectives (Ralph W. Tyler).....	24
1.2.2. Evaluation and practice (L. J. Cronbach)	36
1.2.3. Evaluation and its logic (Michael Scriven)	43
1.2.3.1. Making evaluative judgments	48
1.2.4. Evaluation and decision (Daniel L. Stufflebeam)	51
1.2.4.1. The CIPP evaluation model’s theoretical roots and applications	52
1.2.5. Responsive evaluation (Robert E. Stake).....	54
1.3. The curriculum evaluation.....	59
1.3.1. Program evaluation.....	62
1.3.2. Bradley’s effectiveness model	67
1.3.3. Eisner’s connoisseurship model.....	69
1.3.4. Michael Patton’s utilization-focused evaluation	69
1.4. Teacher performance evaluation	71
1.4.1. The accountability model	73
1.4.2. The professional development model	74
1.4.3. The mixed model.....	76
1.5. The professional development.....	80
1.6. Empirical studies	88
1.6.1. Teacher’s evaluation by students	88
1.6.2. Teachers self-evaluation.....	91
1.6.3. Differences / Similarities between students and teachers competences.....	94
1.6.4. Teacher performance evaluation (interviews).....	95
CHAPTER II – METHODOLOGY	99
2.1. The nature of research	99
2.2. The research design	102

2.3. The case study of UNTL.....	104
2.4. Sampling and respondents	107
2.5. Instruments of data collection	111
2.5.1 Documental analysis.....	111
2.5.2 Questionnaire	112
2.5.3 Interview	114
2.6. Data analysis procedures.....	115
2.6.1 Statistical analysis	116
2.6.2 Content analysis	119
2.7. Ethics in research.....	122
CHAPTER III – QUANTITATIVE RESULTS ABOUT TEACHERS COMPETENCES EVALUATION...	125
3.1 The students' evaluation on teachers competences.....	125
3.2 The teachers' competences self-evaluation.....	141
3.3 Comparison between the students' evaluation and teachers' self-evaluation on teachers' competences.....	158
CHAPTER IV – QUALITATIVE DATA ANALYSIS RESULTS.....	165
4.1. Category 1: Recognition and importance attributed to the Teacher Performance Evaluation (TPE)	167
4.2. Category 2: Attitudes / Positioning towards TPE.....	169
4.3. Category 3: Practices for improving teacher performance	171
CHAPTER V – DATA ANALYSIS AND INTERPRETATION	175
5.1. Research questions and objectives.....	175
5.1.1. How do students evaluate the UNTL teachers' competences?	175
5.1.2. How students' background characteristics (like gender, age and faculty) influence their evaluation?	177
5.1.3. How do UNTL teachers self-evaluate their competences?.....	179
5.1.4. How do teachers' background characteristics influence their self-evaluation?	182
5.1.5. To what extent do students' evaluation of UNTL teachers and the teachers' self-evaluation competences differ or converge?.....	185
5.1.6. How is the Teaching Performance Assessment (TPE) being implemented in the educational institution?.....	189
5.2. Hypotheses	193
5.2.1. The personal and the social competences were the highest evaluated by students, as compared to the pedagogical and professional competences.....	193
5.2.2. There is a relationship between students' evaluation on teacher's competences and students' background characteristics.....	194
5.2.2.1. The female students attribute a more positive evaluation to teachers'	

competences than male students	194
5.2.2.2. The older students give a more positive evaluation to teachers' pedagogical and professional competences than younger students	195
5.2.2.3. Students from humanities provided higher evaluation to personal competences, while students in natural and exact sciences provided higher evaluation to pedagogical competences of the teachers	195
5.2.3. Teachers self-evaluate their pedagogical and professional competences lower than their personal and social competencies	197
5.2.4. UNTL teachers' with more age and years of experience had higher self- evaluation than younger and least experienced teachers	198
5.2.5. UNTL teachers with a higher academic degree have a higher self-evaluation on their competences	199
5.2.6. UNTL teachers' from political science department highest self-evaluated their competences	200
CONCLUSIONS AND RECOMMENDATIONS.....	203
Conclusions.....	203
Recommendations.....	209
REFERENCES.....	211
APPENDICES.....	1
Appendice 1. Questionnaire on performance evaluation of teachers.....	3
Appendice 2. Questionnaire on teacher self-assessment.....	4
Appendice 3. Interview script	5
Appendice 4. Interview transcript	5

Table of Tables

Table 1. Epistemological models of evaluation	66
Table 2. Bradley's effectiveness model for curriculum development indicators	68
Table 3. Criteria for a curriculum evaluation model.....	70
Table 4. Research design framework of quantitative and qualitative study	103
Table 5. Hypothesis and the theoretical framework of quantitative study	104
Table 6. Characterization of students sample (N=342)	109
Table 7. Characterization of teachers sample (N=192)	110
Table 8. Criteria for verifying the degree of consensus of the answers.....	117
Table 9. Degree of mean agreement	118
Table 10. Descriptive statistics and correlation of students' evaluation on teacher competences ...	125
Table 11. Descriptive statistics of students' evaluation on teacher pedagogical competences	126
Table 12. Descriptive statistics of students' evaluation on teacher professional competences .	127
Table 13. Descriptive statistics of students' evaluation on teacher personal competences	127
Table 14. Descriptive statistics of students' evaluation on teacher social competences.....	127
Table 15. Frequency statistics of students' evaluation on teacher competences	128
Table 16. Statistical association between students' evaluation on teacher competences and students gender	129
Table 17. Statistical association between students' evaluation on teacher competences and students age.....	131
Table 18. Statistical association between students' evaluation on teacher competences and students faculty	132
Table 19. Proportion of agreement on students' evaluation of pedagogical competencies, by faculty	133
Table 20. Proportion of agreement on students' evaluation of professional competencies, by faculty	134
Table 21. Proportion of agreement on students' evaluation of personal competencies, by faculty	135
Table 22. Proportion of agreement on students' evaluation of social competencies, by faculty	136
Table 23. Proportion of agreement between the student's evaluation of pedagogical and professional competences.....	137
Table 24. Proportion of agreement between the student's evaluation of pedagogical and personal competences	138
Table 25. Proportion of agreement between the student's evaluation of pedagogical and social competences	139
Table 26. Proportion of agreement between the student's evaluation of professional and personal competences	139
Table 27. Proportion of agreement between the student's evaluation of professional and social competences	140

Table 28. Proportion of agreement between the student's evaluation of personal and social competences	140
Table 29. Descriptive statistics and correlation of teachers' competences self-evaluation.....	141
Table 30. Descriptive statistics of teachers' pedagogical competences self-evaluation	142
Table 31. Descriptive statistics of teachers' professional competences self-evaluation	142
Table 32. Descriptive statistics of teachers' personal competences self-evaluation	143
Table 33. Descriptive statistics of teachers' social competences self-evaluation.....	143
Table 34. Frequency statistics of teachers' competences self-evaluation (N = 192)	144
Table 35. Statistical association between teachers' competences self-evaluation and teachers' gender	145
Table 36. Statistical association between teachers' competences self-evaluation and teachers' age	146
Table 37. Statistical comparison of teachers' competences self-evaluation and teacher's time of experience	146
Table 38. Statistical comparison of teachers' competences self-evaluation and teacher's degree	148
Table 39. Statistical comparison of teachers' competences self-evaluation and teacher's faculty	149
Table 40. Proportion of teacher's self-evaluation agreement in relation to pedagogical competences, by faculty.....	151
Table 41. Proportion of teacher's self-evaluation disagreement and agreement in relation to professional competences.....	152
Table 42. Proportion of teacher's self-evaluation agreement in relation to personal competences	153
Table 43. Proportion of teacher's self-evaluation disagreement & agreement in relation to social competences	154
Table 44. Proportion of agreement between the teacher's self-evaluation of pedagogical and professional competences.....	155
Table 45. Proportion of agreement between the teacher's self-evaluation of pedagogical and personal competences	155
Table 46. Proportion of agreement between the teacher's self-evaluation of pedagogical and social competences	156
Table 47. Proportion of agreement between the teacher's self-evaluation of professional and personal competences	157
Table 48. Proportion of agreement between the teacher's self-evaluation of professional and social competences.....	157
Table 49. Proportion of agreement between the teacher's self-evaluation of personal and social competences	158
Table 50. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to pedagogical competences.....	160
Table 51. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to professional competences	161

Table 52. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to personal competences.....	161
Table 53. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to social competences	162
Table 54. Sociodemographic, professional and academic background characteristics of the interviewees.....	166
Table 55. Matrix of qualitative analysis	166
Table 56. Recognition and importance attributed to the Teacher Performance Evaluation (TPE)	167
Table 57. Attitudes towards TPE	169
Table 58. Practices for improving teacher performance	171

Table of Figures

Figure 1. Relationship in 3 C's to a quality of the evaluation.....	72
Figure 2. Layers of analysis in identifying contents and forms of teachers' professional development.....	86

Table of Graphics

Graphic 1. Student's evaluation on teacher's competences	128
Graphic 2. Proportion of agreement on students' evaluation of pedagogical competencies, by faculty	134
Graphic 3. Proportion of agreement on students' evaluation of professional competencies, by faculty	135
Graphic 4. Proportion of agreement on students' evaluation of personal competencies, by faculty	136
Graphic 5. Proportion of agreement on students' evaluation of social competencies, by faculty	137
Graphic 6. Teacher's competences self-evaluation	144
Graphic 7. Means of teachers' self-evaluation ratings by teacher's time of experience	147
Graphic 8. Means of teachers' self-evaluation ratings by academic degree	148
Graphic 9. Means of teachers' self-evaluation ratings by faculty	150
Graphic 10. Proportion of teacher's self-evaluation agreement in relation to pedagogical competences, by faculty	151
Graphic 11. Proportion of teacher's self-evaluation agreement in relation to professional competences, by faculty	152
Graphic 12. Proportion of teacher's self-evaluation agreement in relation to personal competences, by faculty	153
Graphic 13. Proportion of teacher's self-evaluation disagreement & agreement in relation to social competences, by faculty	154
Graphic 14. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to pedagogical competences	160
Graphic 15. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to professional competences	161
Graphic 16. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to personal competences	162
Graphic 17. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to social competences	163

List of Acronyms

- AERA – América Educational Research Association
- ANAAA – Agencia Nacional Acreditação Avaliação Académico
- ASCE – Applied Strategies for Curriculum Evaluation
- CAT – Committee on Ability Testing
- CEDU – Certificação do Docente Universitário
- CIPP – Context, Input, Process and Product
- CRD–TL – Constituição República Democrática Timor–Leste
- DRET – Democratic Republic of East Timor
- ECDU – Estatuto Carreira Docente Universitario
- ECTS – European Credit and Transfer System
- ESEA – Elementary and Secondary Education Act
- HOTS – Higher Order Thinking Skills
- IQ – Intelligence Quotient
- MDG – Millennium Development Goals
- METL – Ministry of Education Timor–Leste
- MO – Modus Operandi
- NAEP – National Assessment of Education Progress
- NCLB – No Child Left Behind
- NRC – National Research Council
- OECD – Organization for Economic Co-operation and Development
- PD – Professional Development
- PDN–TL – Plano Desenvolvimento Nacional Timor–Leste
- PED – Plano Estratégico Desenvolvimento
- Ph.D – Doctor of Philosophy
- PISA – International Student Assessment Program
- PTE – Program Theory Evaluation
- SAT – Scholastic Aptitude Test
- TPE – Teacher Performance Evaluation
- UK – United Kingdom
- UNESCO – The United Nations Educational, Scientific and Cultural Organization
- UNTAET – United Nations Transitional Administration in East Timor
- UNTL – Universidade Nacional de Timor Lorosae
- US – United States
- USA – United States of America
- UTOS – Units Treatments Observation Outcomes

INTRODUCTION

Contextualization

Studies on the education system in East Timor are still very recent, given the context of recent independent country and the scarce number of years of decisions aimed at consolidating education policies appropriate to Timorese reality.

After a period of political and social instability following the referendum in 1999 and the United Nations Transitional Administration in East Timor (UNTAET), the new country lost much of its skilled workforce in all sectors, including education (Millo & Barnett, 2004).

Despite the progress made (DRET, 2011), there are still indicators that are of concern to the Timorese authorities and which require reform measures capable of changing the general panorama of education, namely high rates of dropout and repetition, and for which the following factors: lack of textbooks and learning materials; the reduced number of teaching hours; the low preparation of teachers; high teacher ratios per teacher; deficient physical infrastructure; high number of students contrasting with the high rate of teacher absenteeism; the incomplete preparation of students for the language of instruction, Portuguese (Albergaria-Almeida, Martinho & Cabrita, 2014, p.666).

That is why the Timorese authorities see the reform of education through the training of national human resources as a key means to reduce inequalities, promote the social and economic integration of the population, contribute to the eradication of poverty and achieve a better quality of life of populations, thus achieving those that are the country's *Millennium Development Goals* (DRT, 2009).

The broader goals of the current reform are to contribute to socio-economic development and improve the capacity of the country's human resources (Berlie, 2007). Since teacher quality is one of East-Timor's priorities for development ("Teachers are a priority", 2013) it is essential to implement a new curriculum to ensure the training of high-quality teachers, both scientific and along with procedures and criteria that are in line with the country's educational goals (Albergaria Almeida, Martinho & Lopes, 2013).

The performance evaluation is carried out in accordance with the principles enshrined in the articles of Decree-Law no. 14/2008 of May 7, 2008, the *Regime for the Evaluation of the Performance of Public Administration Employees*.

This document constitutes an important instrument in the introduction of a new culture of management, as it can allow an assessment of the resources allocated to each of the bodies and functions of the public sector, as well as the creation of conditions for greater professional motivation, qualification and permanent training of human resources.

The purpose of the evaluation is to improve the performance of workers by helping them to achieve higher levels of performance in order to increase career opportunities according to their potential and to value individual contributions to the team. On the other hand, according to article no. 3 of the aforementioned Decree-Law, the purpose of performance appraisal is to assess, hold accountable and recognize the merit of managers, employees, agents of the Public Administration, in function of productivity and achievement of objectives, services and public bodies. It is also an instrument for the evaluation of the probationary official regarding the fulfillment of the conditions to integrate a career in the public function.

The performance evaluation also aims at pursuing the following objectives: (i) motivate employees and agents; (ii) improve their professional performance; (iii) encourage communication between managers and their subordinates; (iv) improve integrated management of human resources; (v) to promote excellence in the quality of the provision of services to the public; (vi) identify training needs that can improve performance and help achieve the institution's objectives; (vii) assist in setting performance objectives for the coming year.

The Constitution of the Democratic Republic of East-Timor (CRD-TL, 2012) states that "The State shall guarantee access to the highest levels of education according to its capabilities" (CRD-TL, Part II, Section 59, p.4). The State must then ensure the quality of teaching and the improvement of the skills and competences of teachers particularly university teachers. The teachers training, the productivity of their work in the monitoring of their performance and results are the key elements of the quality of educational institutions, in particular, and of the entire Timorese education system.

(...) Basic Education Law of East Timor

Timorese Higher Education system comprised, in 2004, 17 higher education institutions in operation, serving more than 13,000 students. At the beginning of 2011, there were 11 institutions in operation, 9 of which had academic accreditation and served approximately 27,010 students. Since 2009, female enrollments in higher education have increased by 70% (PED, 2011, p.25). Timor Lorosa'e National University (UNTL), established in 2000, is the only public university in East

Timor to play a major role in public higher education service and in scientific and specialized research (PED, 2011, p. 26).

Delimitation of the study (the educational evaluation field)

The evaluation of teachers' performance has received increasing attention from the academic community and policy makers around the world, "coupled with the idea that it is one of the decisive aspects for improving the quality of education" (Flores, 2010, p. 7). The need to raise standards of education and raise the quality of student learning has led governments to introduce reforms in schools and to enforce greater accountability among teachers.

The rapid expansion of tertiary education in recent decades has allowed a large number of students to be accommodated in many countries, but it has also created a need for administrative reform to implement a more efficient, effective and economical management model. Higher education institutions were thus called upon to provide a public service, not only broader, but also with greater administrative efficiency and accountability, in response to the demands of different stakeholders (governments, business, industry, Workers' organizations, students and the community at large) (Ka-ho, 2003).

At the same time, society today is increasingly demanding the demonstration of the value of the work of Higher Education teachers, and it is also increasingly recognized that university institutions need particular attention that meets the specificities of their human resources management (Jaquith, Mindich, & Darling-Hammond, 2010) positioning themselves in order to define quality criteria and the effectiveness of the service they provide.

We will therefore address performance evaluation within the paradigm of new human resources management in Public Administration to argue that the current performance appraisal model applied to all East Timorese civil servants does not meet the demands and specificities of Timorese Public higher education.

Because it is too comprehensive, the current evaluation system used to evaluate the performance of university teachers does not contain the necessary adjustments to the objectives and challenges of higher education in the country. Therefore, this thesis, of a theoretical, empirical and critical nature, intends to outline the bases and the guiding principles for the design of a new model of teacher performance evaluation in East Timor that meets the challenges of the current public higher education system.

In the latter half of the nineteenth century the UK and the US measured teacher performance in public schools (Pollit & Bouckaert, 2011). Woodrow Wilson believed in the need to create an

administrative system based on efficiency criteria and Taylor advocated a generic approach to measuring workers' efficiency (Pollit & Bouckaert, 2011, p. 106).

The interest in measuring public sector activity has, however, intensified over the last quarter century with the increasing pressure of contemporary societies on the state and, in particular, on the public sector, in order to modernize its model and its management practices (Bouckaert & Halligan, 2008) (Boyne, Kenneth, Jr., & Walker, 2006).

The increasing sophistication of the needs of citizens and companies that the public administration must meet has led to the successive and continuous search for better efficiency, efficacy and quality of public services. In addition to economic performance indicators (Hood, 1996), service quality also emerged as a priority (Coe, 2003).

Pressures to meet performance goals or standards - such as decreased hospital waiting lists, better test scores, higher crime rates, or better university research scores - has introduced deep changes in public organizations.

The evaluation should therefore be based on rigorous principles and criteria such as rigor, transparency, requirement and objectivity, with a view to promoting the professional development of teachers on the basis of recognition of merit, professional effort and excellence. In addition to these principles, evaluation must be viewed in an integrated and contextualized way, associated with the evaluation of the school and based on a prior and clearly defined professional profile, appropriate to the functions carried out by the teacher.

Therefore, the need to allocate qualified human resources to ensure higher levels of efficiency and effectiveness was recognized. Performance evaluation has played a key role in providing background information to support decision-making in human resource management, particularly in terms of career selection and promotion (such as determining whether salary increases and provide feedback among supervisors, or evaluators, and employees) (Coutts & Schneider, 2004).

The new management of human resources in public administration has advocated the principles of decentralization, autonomy, accountability and flexibility (Perry, 2010), in order to introduce changes towards a closer approximation to the citizen and, therefore, the provision of a more effective, efficient, higher-quality and less wasteful services (Ka-ho, 2003).

The evaluation of performance then was integrated into several dimensions and became more extensive. On the one hand, in addition to the management function, namely the monitoring function, the performance evaluation started to intervene in making decisions, in control and in accountability. On the other hand, it has become more extensive because its application is no

longer merely internal, but also serves the members of the legislative bodies and even the public. Performance evaluation in public administration has thus become, over successive governments, a "change imperative" in the management of public affairs (Bouckaert & Halligan, 2008).

The impact of the changing of economic, social and knowledge contexts upon the education service as a whole has caused a move from the traditional post-war model of autonomous professional in which decisions about the curriculum, teaching, learning and assessment are the business of teachers. Now, what students learn, what they must achieve as the outcome of learning and what standards apply are the everyday business of the state.

Concerned with the need to raise standards of achievement and improve their positions in the world economic league tables, governments over the last 20 years have intervened more actively to improve the system of schooling. Higher expectations for higher quality teaching demands teachers well qualified, highly motivated, knowledgeable and skillful, not only at the point of the entry into teaching, but also throughout their careers (Day, 2002).

According to Mathison (2009) the genesis of educational evaluation field is the stipulations in the *Elementary and Secondary Education Act* (ESEA) passed in 1965 by Lyndon Johnson in his war on poverty. The ESEA provides federal assistance to schools, communities, and children in need. With Johnson's conceptualization of ESEA, educational evaluation was seen to be a public good (just like education and schooling) that should serve the common public good.

While the passage of ESEA marks the beginning of the formalization of educational evaluation, one prior event, the *Eight Year Study* (Tyler R. W., 1949) also played an important role in educational evaluation, although it is more often associated with developments in curriculum theory and design. *The Eight Year Study* involved 30 high schools dispersed throughout the USA serving diverse communities. Each school developed its own curriculum and each was released from government regulations, as well as the need for students to take college entrance examinations. With dissension early in the project about how its success should be evaluated, Ralph Tyler was brought on board to direct the evaluation, which was funded by the Rockefeller Foundation. Out of the *Eight Year Study* came what is now known as the '*Tyler Rationale*' the commonsense idea that what students were supposed to learn should determine what happened in classrooms and how evaluation should be done (Tyler, 1949).

Contemporary educational evaluation is rooted in student assessment and measurement. The distinction between measurement and evaluation, suggested by Ralph Tyler more than 50 years ago and later elaborated on by others, had an enormous influence on the development of evaluation

as an integral part of the teaching and learning process. For many years, educational evaluation focused mainly on students' achievements; it concentrated on the use of tests and was immensely influenced by psychometrics. Another major and long-standing influence on educational evaluation is to be found in a variety of programs to accredit schools and colleges. Mainly a U.S. enterprise, accreditation programs began in the late 1800s and are an established reality throughout the U.S. today (Kellaghan, Stufflebeam, & Wingate, 2003, p. 2) (Alkin & King, 2016, p. 570).

The evaluation field has experienced many decades of differences of opinions about which methodologies are best. Shadish (1998) claims that differences about methodologies are not based on arguments about methods choices, but they are reflective of the different philosophical assumptions that guide methodological choices. He wrote that most debates in the evaluation field are “about epistemology and ontology, about what assumptions we make when we construct knowledge, about the nature of many fundamental concepts that we use in our work like causation, generalization and truth” (p. 3).

Christie and Alkin (2013) identified three branches of evaluation: the methods, the use and the values branch. Educational evaluation encompasses a wide array of activities focusing on tests and measurements, including student measurement, testing, program evaluation, school personnel evaluation, school accreditation and curriculum evaluation. It occurs at all levels of education systems, from the individual student evaluations carried out by classroom teachers, to evaluations of schools and districts, to district-wide program evaluations, to national assessments, to cross-national comparisons of student achievement. As in any area of scholarship and practice, the field is constantly evolving, as a result of advances in theory, methodology, and technology; increasing globalization; emerging needs and pressures; and cross-fertilization from other disciplines (Kellaghan, Stufflebeam & Wingate, 2003).

It was only in the mid-1960s and early 1970s, with the increased demand for program evaluation made necessary by various evaluation requirements placed on educational programs and projects by governmental organizations and other agencies that educational evaluation dramatically expanded and changed in character. While earlier evaluation, as noted above, had focused on student testing and the educational inputs of interest to accrediting organizations, the new thrust began to look at a variety of outcomes, alternative program designs, and the adequacy of operations. To meet new requirements for evaluation, evaluators mainly used their expertise in measurement and psychometrics, though they also took advantage of two other resources: research methodology and administration.

Educational evaluation, in its principles, was based on a goal-centered model. Knowing if a project can achieve its objectives should be the end of an evaluation. For this, there is nothing better than using rigorous measuring instruments (such as tests, for example), according to an experimental (or rather quasi-experimental) methodology. This approach was known as "psychometric" because it relied heavily on the type of research carried out by laboratorial psychology, with very detailed statistical analysis.

In the 1960s it was the principle of "revolt" against this predominantly quantitative model, which materialized in the early 1970s with open "conflict" between the evaluators known as "traditional" and those who proposed a "new assessment." To Varela (2001) the new evaluation aimed to forget the dominance of the model by objectives and statistics and to base a humanized assessment that placed at the center of the educational process, responsible for the complex changes of the students, an evaluation that "illuminated" through interpretations reality, the same reality.

These new evaluators do not share all the same ideas, but they have in common the adoption of a new evaluation paradigm that can be called qualitative, as opposed to the previous quantitative paradigm, but itself with several approaches.

Recalling the same techniques used by ethnography, these evaluators value the evaluation study in natural environments and present the evaluator as an instrument for data collection and analysis. The evaluator aims at the rich and profound description of the reality that he studies, through the use of techniques such as observation, interview and giving more importance to the process than to the product (Schofield & Anderson, 1984). This position provoked a great debate in the 70's.

With the passage of time the antagonisms became blurred and the quantitative-qualitative quarrel seems, if not definitely outdated, at least greatly diminished in its virulence. Nowadays the evaluators tend towards certain eclecticism, using the methods that best suit their purposes, without taking care of radicalisms. The defense of mixed methods was done in time and, in general, was welcomed by the most enlightened scientific community. As a consequence, we now see greater conceptual freedom to use the methods that are most appropriate for each case.

Mertens and Wilson (2012), based on the work of Guba and Lincoln's (2005) about the concept of paradigm in research and evaluation, described four sets of philosophical assumptions that constitute a paradigm – axiology, ontology, epistemology and methodology – and added the fourth branch of Social Justice.

Research methodology, mainly quantitative but later also qualitative, provided the guidance for data collection procedures and research designs that could be applied in evaluation. Administration

theory and research helped to improve understanding of planning and decision making, which evaluations were designed to service, as well as of the politics of schools.

Kellaghan, Stufflebeam and Wingate (2003) identify three distinctive features that set educational evaluation apart from other types of evaluation. First, it has been strongly shaped by its roots in testing and student assessment, on one hand, and curriculum and program evaluation on the other. In other areas (e.g., health services or criminal justice), evaluation focuses mainly on programs and is usually considered as a type of applied research. Although it took many years for educational evaluation to come to the point where it would not be perceived only as student assessment, such assessment is still an important element of the activity.

Second, education is the predominant social service in most societies. Unlike business and industry, or other social services such as health and welfare, education affects, or aspires to affect, almost every member of society. Thus, public involvement and the concerns of evaluation audiences and stakeholders are special significance in educational evaluation, compared to evaluation in other social services, and even more so when compared to evaluation in business and industry (Kellaghan, Stufflebeam & Wingate, 2003).

Third, teachers play very important roles in educational evaluation as evaluators, as evaluation objects, and as stakeholders. They are a unique and extremely large and powerful professional group, with a high stake in evaluation and a long history as practicing evaluators assessing the achievements of their students and must be taken into account whenever evaluation is being considered (Kellaghan, Stufflebeam & Wingate, 2003).

Education is one of the main pillars of the evaluation field, and thus it is important that those who work in educational evaluation should be part of the general evaluation community, participating in its scientific meetings and publishing their work in its professional journals. There is much that they can share with and learn from evaluators in all areas of social service, industry, and business. However, educational evaluators should also be sensitive to the unique features of their own particular area of evaluation and work to develop its capabilities so that they can better serve the needs of education and its constituents (Kellaghan, Stufflebeam, & Wingate, 2003).

Evaluation is the more relevant and actionable when it is actually engaged with those who are being evaluated and is able to generate processes of development and learning related to real problems (Ivaldi, Scaratti & Nuti, 2015, p. 499). Then, evaluation involves questioning the idea that evaluation research should directly serve the decisions of policy-making and practice (Hammersley, 2013).

In this regard, as Contandriopoulos and Brousselle (2012) suggest, evaluation results need to be embedded in a real-life context in order to be relevant. In particular, they point out that the *use* and the *usability* of the evaluation depend on both the distance between actors' different readings of problems, and the balance that they attribute to the cost/benefit ratio of investing the energy and resources that evaluation requires.

The process and the results of evaluation practice, therefore, depend on how different and often divergent opinions, criteria and interests are combined in a context. It is this that determines the specific conditions of knowledge-use and exchange, including the use of knowledge produced through evaluation (Ivaldi, Scaratti & Nuti, 2015, p. 499).

Social science stream that focuses on the use of social research methodology

It was only in the mid-1960s and early 1970s, with the increased demand for program evaluation made necessary by various evaluation requirements placed on educational programs and projects by governmental organizations and other agencies, educational evaluation dramatically expanded and changed in character. While earlier evaluation had focused on student testing and the educational inputs of interest to accrediting organizations, the new thrust began to look at a variety of outcomes, alternative program designs, and the adequacy of operations. To meet new requirements for evaluation, evaluators mainly used their expertise in measurement and psychometrics, though they also took advantage of two other resources: research methodology and administration.

Although its historical roots extend many centuries ago, widespread evaluation research is a relatively modern 20th century development. The application of social research methods to program evaluation coincides with the growth and refinement of the research methods themselves as well as with ideological, political, and demographic changes.

In the evaluation literature, there is a rich and growing body of knowledge about the utility of evaluations. The approach to evaluation known as utilization-focused evaluation perceives the purpose of any evaluation to be defined by the utility of the information it provides to its users, where users can be multiple stakeholders including funders, the nonprofit board, program staff, and so on (Alkin, 2013). This school of thought points to numerous factors that contribute to the utility of evaluations, such as building organizational readiness for (useful) evaluations and informing intended users of the potential controversies and limitations of their methodological choices (Alkin, 2013; Patton, 2012).

In the public sector, evaluation plays an important part in policy-making and plan-making processes. Both processes follow a problem identification and definition phase, formulation phase, implementation phase, and evaluation phase. Evaluation is about determining how successful an intervention has been and the identification of areas for improvement (Pal, 2014). It is a structured process that aims to create and synthesize information about interventions in order to make judgments regarding resultant changes, the desirability of an intervention, and the degree of fit between intended and unintended outcomes.

So, carrying out a good evaluation is also an opportunity to improve an organization, because of the structural link between shared evaluative orientations and broader aspects of organizational life. It also makes possible the transformative capacity of taken-for-granted activities, through a process of discussion and negotiation about the meanings attributed to events and situations, roles and visions as well as to problems. In this perspective, the evaluation is conceived as contributing to organizational transformation by mediating the inevitable political, cultural and relational implications that are characteristic of all evaluations.

Evaluation is one of the critical steps in the process of performance improvement. Evaluation feeds evidence-based information back to the next cycle of performance improvement. Evaluation data can/should be used during performance analysis, intervention design, development, or selection, and during intervention implementation and maintenance. Evaluation is also the key to managing change in organizations. However, organizations often neglect to conduct comprehensive evaluations on their programs due to environmental barriers or the lack of practitioners' evaluation expertise.

Evaluation is similar to, but also different from, research. Evaluation is conducted with quantitative and/or qualitative data obtained from research activities such as surveys, interviews, observations, tests, or extant data reviews. But, unlike research, evaluation is often client-based and context-specific. To conduct a program evaluation, evaluators need to clearly understand not only the characteristics of the program itself, but also the composition of stakeholders of the program, including upstream stakeholders and downstream stakeholders.

While doing so, evaluators will review or develop a logic model of the program, which clearly identifies the program's means (inputs/resources and activities) and end results (outputs, outcomes and impacts). A program logic model serves as a road map for successful planning and execution of the program. A program logic model helps evaluators as well as the client organization understand the roles that upstream and downstream stakeholders play and recognize dimensions

to be investigated depending on whether the evaluation has a formative or summative purpose. To improve the quality of the program (formative), it is necessary to include process-based dimensions as well as outcome-based dimensions. To assign the overall value of the program (summative), the outcome-based dimensions are often the focus, although it is a good practice to investigate process-based dimensions as well. Selection of dimensions or wording of the dimensional sub-evaluation questions may also differ, depending on whether the evaluation focuses on the program's merit or worth (Chyung, 2015).

The evaluation research methodology, mainly quantitative but later also qualitative, provide the guidance for data collection procedures and research designs that could be applied in evaluation. Administration theory and research helped to improve understanding of *planning* and *decision making*, which evaluations were designed to service, as well as of the politics of schools (Alkin & King, 2016).

Looking at the evaluation as a social practice (Saunders, 2012), starting from the distinction between the concept of use (the organizational capacity to use evaluation outputs and usability) and evaluative design can facilitate the potential use of findings. The evaluation as a social practice, argues Saunders (2012), is considered as the kind of routine rule-governed behavior evaluators engage in as they carry out their work.

Ledermann (2012) argues the multifaceted nature of evaluation, especially when applied to complex situations and objects.

Evaluation is considered an empirically oriented discipline that generates information about programs in order to improve the program or guide future decisions (Pal, 2014).

Second, there is an emphasis on both program operation and outcomes. That is, evaluation is not only concerned with program effectiveness but also the process of delivering programs such as the organizational methods used to deliver the program, program inputs (e.g., resources), program outputs (e.g., tangible measures of a program), and cost effectiveness (Howlett, Ramesh & Perl, 2009). Finally, program evaluation is used to help make programs work both efficiently and effectively (Weiss, 1998), and as a means to ensure accountability and quality assurance (Cousins, Goh, Elliott & Aubry, 2013; Pal, 2014).

Evaluation can be used to enhance the quality and implementation of plans, improve the planning process, and demonstrate the effectiveness of plans. Through an evaluation, we can empirically document the deficiencies, or strengths, in plans and identify specific weaknesses that undermine implementation and plan effectiveness (Berke, Smith & Lyles, 2012). Evaluation can provide an

objective and systematic approach to study plans, improve the plan preparation process, and assess whether plans achieved their stated goals and objectives.

In order for plans to be effective and evaluable, there needs to be a clear relationship between the main components of a plan (i.e., its goals, objectives, and policies) and implementation mechanisms (Oliveira & Pinho, 2011).

Plan goals and objectives should be clear, policies measurable, and concepts clearly operationalized. This would allow planners to identify indicators needed to support successful monitoring and evaluation of plans. Poorly defined goals and a lack of connection between policies and plan implementation gives rise to the possibility that plans will be unable to achieve their stated goals and, more importantly, make monitoring and evaluation difficult (Stevens, 2013; Baynham & Stevens, 2014).

According to Christensen (2015), exemplary planning practice should focus on both process and outcome. Evaluation can be used to demonstrate the value of the planning process and of plans. Plans and planning activities are regularly criticized by the public, politicians, and other professions who claim that planning is costly, imposes undue controls and burdens on landowners, and fails to make a difference (Laurian et al. 2010).

Evaluation helps planners respond to these criticisms by demonstrating the outcomes and impacts associated with plans. Perhaps the greatest benefit of evaluation is that it holds planners and those involved in plan preparation accountable; this can be a way to legitimize the field of planning (Laurian, et al. 2010) (Oliveira & Pinho, 2011).

Plan evaluation provides the critical final link between plan preparation, implementation and outcomes. It is through evaluation that planners are able to discern whether a plan is being implemented as intended, and to identify the effects of plans. Evaluation is also used to determine whether plans should be reviewed in order to realign goals and policies so that a preferred outcome is achieved (Stevens, 2013).

Evaluation constitutes a complex and articulated practice that Patton (2012) considers both individual and organizational dimensions and involves different stakeholders. The quality and effectiveness of the delivered services depends on organizational functioning and the way in which staff and other organizational actors interpret their work. Evaluation can provide context-driven knowledge that becomes a basis for a reflexive process of organizational change, development and innovation. Thus, the assessment of organizational functioning and the outputs or services it produces (based on knowledge sharing, reflection and learning) can lead to the identification of

problems and criticalities, enhancing adaptation and potential solutions. Evaluation is not a neutral procedure capable of producing standardized results, but rather an inductive process that makes it possible to analyze the system of activities, the bundle of practices and the relationships in which practitioners are involved (Ivaldi, Scaratti & Nuti, 2015, p. 498).

Evaluation can be used to enhance the quality and implementation of plans, improve the planning process and demonstrate the effectiveness of plans. Evaluation can provide an objective and systematic approach to study plans, improve the plan preparation process, and assess whether plans achieved their stated goals and objectives. Through an evaluation, we can empirically document the deficiencies, or strengths, in plans and identify specific weaknesses that undermine implementation and plan effectiveness (Berke, Smith & Lyles, 2012).

The evaluation of social policies is conceived as a process oriented to judge the value and merits of governmental political initiatives, whose results are capable of expressing both conflicts and social conciliations developed in different contexts and institutions. Based on power decisions and governmental institutions, this type of evaluation tends to be recognized as a strategy of administrative, governmental and social interests, with two objectives: a promotion of the construction of self-consciousness of the institution, or agency, about limits and possibilities of their actions; and formulation of information and analysis on the policy for the use of evaluated institution, government and society (Pacheco, 1998).

Relevance, contribution and pertinence of the study

The relevance and pertinence of this study is the contribution to the understanding of a culture of professional development in East Timor and its impact on models of performance evaluation of UNTL teachers. It is hoped that the study will raise understanding of professional development culture and that it will be useful to improve the formulation and analysis of TPE to the country in this area.

The results to be achieved fall within the scope of the evaluation of professional performance on the understanding that the study to be carried out will be one of the first contributions to the exploitation of this subject in the education system of East Timor, responding equally to the prosperity of the UNTL, that is, the design and implementation of a model of professional teacher evaluation.

As a UNTL professor for several years, with management responsibility at the level of the department director of one of the department of philosophy, I have the indispensable knowledge of the context of the object of study, assuming an increased motivation not only institutional

responsibility, but also the pertinence of the subject. The evaluation of the professional performance of teachers, as part of an integrated evaluation, is a key part of an education system that states, according to the Education Law, that higher education is oriented towards the development and improvement of the country.

The main purpose of the self-assessment is to ensure higher education institutions established standards of quality and that they continually evaluate the extent to which they meet educational quality goals. From this self-assessment will come recommendations for improvements or enhancements to policies, processes, programs, services, facilities, and human resources.

Effective self-assessment serves both internal and external purposes. It is concerned with quality assurance and encourages institutional improvement through rigorous self-analysis which is the heart of the accreditation process.

The self-assessment should not be viewed as an isolated and a standard process in which an institution periodically engages. Rather, self-assessment should be an integral part of the institution's ongoing planning and evaluation efforts involving teachers, students, political agents and community.

Students' views and perceptions of good teaching are important areas to consider in designing effective systems of teaching evaluation. In the field of the professional performance assessment and the improving quality of education in UNTL, we address the issue of the UNTL's lectures competences: divergences and convergences between the students' evaluation and the teachers' self-assessment.

Higher education institutions engage in advancing good teaching for two main reasons. First, they are interested in demonstrating that they are reliable providers of good quality education, while serving multiple stakeholders with different expectations. Second, they are required to respond to the increasing demand for meaningful and relevant teaching. Students as well as employers and policy makers want to assure that education will prepare students well to have a rewarding employment, to develop a professional growth career, and to contribute to the country's social and economic progress.

Research question and objectives

The evaluation performance of higher education teachers in East Timor is our research object. The question of this research will address the issue of teacher performance appraisal and its professional development at the National University of East Timor (UNTL). The area to explore is

the evaluation, taking into account their paradigms, models and theories in the context of curriculum practices of higher education within the education system of East Timor.

Students' views and perceptions of good teaching are important areas to consider when designing effective systems of teaching evaluation, as well as the teachers' self-assessment and self-perception. In the field of the professional performance assessment and the improving quality of education in UNTL, we point to four main research questions as shown below:

1. How do students evaluate the UNTL teachers' competences and how do students' background characteristics (like gender, age and faculty) influence their evaluation?
2. How do UNTL teachers self-evaluate their competences and how does their professional background (gender, age, faculty and degree) influence their self-evaluation?
3. To what extent do students' evaluation of UNTL teachers and the UNTL teachers' self-evaluation competences differ or converge?
4. How is the Teaching Performance Evaluation (TPE) being implemented in the educational institution to which the respondents belong?

Objectives

The aim of this study is to address the issue of the implementation of teaching performance assessment in Timorese higher education system, in the specific case of UNTL. By this way, four general objectives and the specific objectives were defined respectively:

1. To describe the student's evaluation of UNTL teachers' competences.
 - 1.1. To analyze if the student's gender, age and faculty influence the evaluation they made on UNTL teachers' competences.
 - 1.2. To identify the highest evaluated UNTL teacher's competences by students.
 - 1.3. To identify the lowest evaluated teachers' competences and those need improvements.
2. To describe the self-evaluation of UNTL's teachers' competences.
 - 2.1. To analyze if the age, time of service, academic degree and department of UNTL teachers influence the self-evaluation that they made of their competences.
 - 2.2. To identify the highest self-evaluated UNTL teacher's competences.
 - 2.3. To identify the competences with lower self-evaluation that needs improvements.
3. To compare the student's evaluation and the teachers' self-evaluation, in order to identify divergences and / or convergences.
 - 3.1. To identify what UNTL teachers' competences have a similar perception and evaluation made by students and teachers.
 - 3.2. To identify what UNTL teachers' competences have a different perception and evaluation made by students and teachers.
4. To know how the TPE (Teacher Performance Evaluation) is being implemented in the educational institution to which the respondents belong.

- 4.1. To identify the attitudes towards the teacher performance evaluation in Timorese higher education system: straightforward and/ or the obstacles.
- 4.2. To know the teacher's experiences in the implementation of teacher performance evaluation in Timorese higher education system.
- 4.3. To identify the teaching and learning competences that should be developed by teachers in a professional development program.
- 4.4. To point out suggestions and recommendations for UNTL teachers in relation to effective implementation of teacher performance evaluation in higher education system.

Formulation of Hypotheses

H1: The personal and the social competences were the highest valued by students, rather than the pedagogical and professional competences.

H2: There is a relationship between students' evaluation on teacher's competences and their background characteristics:

H2a: The female students attribute a more positive evaluation to teachers' competences than male students;

H2b: The older students give a more positive evaluation to teachers' pedagogical and professional competences than younger;

H2c: The students in the humanities attribute a greater evaluation to personal and social competences, while students in natural and exact sciences attribute a greater evaluation to pedagogical and professional competences.

H3: Teachers self-evaluate their pedagogical and professional competences lowest than their personal and social competences;

H4: Teachers with more age and time of experience are those higher self-evaluated their competences, than younger and least experience teachers;

H5: The teachers with a higher academic degree have the highest self-evaluated their competences;

H6: The teachers from political science faculty have the highest self-evaluated their competences.

The structure of the study

The structure of this study is composed of Introduction and five important chapters. The introduction part covers the contextualization, delimitation of the study, relevance, contribution and justification of the study, the research questions and the objectives of the study.

Chapter I contains the theoretical framework namely review of the related literature which underlines the concept of evaluation, historical perspective on the evaluation concept, the concept

and models of professional development, as well as the empirical studies on students' evaluation and teacher self-evaluation in order to define the theoretical framework and formulate the hypothesis of investigation.

Chapter II is composed of the research methodology which covers the nature of the research, the identification of sampling and respondents, instruments of data collection, corpus of documental analysis, data analysis techniques and ethics in research. It reviews the definitions of evaluation concept, keeping in mind that it is considered both an art and a science. We aim to provide a portrait of the current state of the theory and practice of educational evaluation. It is the purpose of this second chapter to attempt to do this, to sketch the international landscape of educational evaluation: its conceptualizations, practice, methodology, background roots and the functions it serves.

Chapter III presents the quantitative data analysis results about the student's evaluation on teacher's competences, the UNTL teacher's competences self-evaluation and then the comparison of both.

Chapter IV presents the qualitative data analysis results of in-depth interviews to 16 UNTL teachers who are at the time holding management positions at UNTL.

Chapter V presents the data analysis and interpretation and ends with the conclusion that reviews the main research contributions and pinpoints the policy implications of the study.

CHAPTER I – REVIEW OF THE RELATED LITERATURE

This chapter provides an overview of the research focused on the evaluation concept, historical perspectives of evaluation in educational field, the concept of professional development and their models. The concepts of evaluation and professional development were mainly searched and synthesized for this study. The literature reviewed primarily includes peer-reviewed articles and studies centered in the field of historical perspectives of the concept of evaluation and professional development. A review of the literature was conducted to gain an understanding of relevant topics directly related to educational evaluation and professional development. Additional references such as published reports and online sources were identified.

1.1. The concept of evaluation

The historical roots of the evaluation concept assume different formal and informal uses by humans for thousands of years. Shadish and Luellen (2005, p. 183) have commented that the history of evaluation is as old as the history of human activity:

humans (a) identify a problem, (b) generate and implement alternatives to reduce its symptoms, (c) evaluate those alternatives, and then (d) adopt those that results suggest will reduce the problem satisfactorily.

Namely, people engage in activities they refer to as evaluation in order to foster “use”.

The term “evaluation” refers to processes that are an integral part of the daily lives of individuals in all societies. Whatever may be its particular form or purpose, the act of evaluation implies (a) an initial experience or “finding-out” that is (b) interpreted by means of standards, rules or principles in order to (c) arrive at a judgment of goodness or desirability.

When described in these general terms, evaluation is readily seen as fundamental regulating mechanism in the lives of men and women in all societies (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956). It is a means by which individuals and groups constantly interpret their own experience, which is not surprising that the generalized skill or capacity of evaluation was assigned by Bloom and his collaborators (1956) to the highest level of hierarchy of cognitive functions.

Shadish and Luellen (2005) provide evidence of evaluation dating back thousands of years, including the discussion of the evaluation of the Hebrew diet regimen in Chapter 1 of the *Biblical Book of Daniel* and personnel evaluation in China more than 4,000 years ago (Shadish & Luellen, 2005). As Guba and Lincoln (1981) pointed out, a Chinese emperor in 2200 b.C. required that his public officials demonstrate their proficiency in formal competency tests.

In the United States, the concern for evaluating schools can be traced at least as far back as the recommendations of the Committee of Ten, which at the end of the 19th century set perhaps the first example of “evaluative standards” for the nation’s secondary schools (National Education Association, 1969). Scriven (1996, p. 395) noted that “evaluation is a very young discipline - although it is a very old practice”.

Synthesizing what the dictionaries tell us, we can define evaluation as “the process of determining the merit, worth, or significance of things” (Scriven, 2012, p. 15) (near-synonyms are quality/value/importance). Reports on the results of this process are called *evaluations* if complex, *evaluative claims* if simple sentences, and we here use the term *evaluand* for whatever it is that is evaluated (optionally, we use *evaluee* to indicate that an evaluand is a person) (*Ibidem*).

Evaluation could be also defined as a study designed and conducted to assist some audience to assess an object’s merit and worth. This definition should have a widely acceptable since it agrees with common dictionary definitions of evaluation and it is also consistent with the definition that underlies published sets of professional standards for evaluations (Joint Committee on Standards for Educational Evaluation, 1981, 1988, 1994). (1981) (1988) (1994)

Boulmettis and Dutwin (2011, p. 28) identified two different definitions of evaluation: evaluation as

the systematic process of collecting and analyzing data in order to determine whether and to what degree objectives have been or are being achieved” and evaluation as the systematic process of collecting and analyzing data in order to make a decision.

The first definition reflects the philosophy that evaluator is interested in knowing only if something worked, that is whether it was effective in doing what it was supposed to do. The second statement reflects the philosophy that evaluation makes claims on the value of something in relation to overall operation of a program, project or event.

According to Mertens and Wilson (2012, p. 1):

evaluation is situated in a broad landscape in terms of its diverse meanings in different disciplines, sectors, nations, and venues. The hallmarks of the evaluation field are their interdisciplinary roots and the ways in which the resultant conversations around the meaning of evaluation have benefited from this diversity of perspectives.

Nowadays, the evaluation has been differentiated, organized, formalized and professionalized in several areas, from literary and critic of art to the economic-financial evaluation of companies (Rodrigues, 1999, p. 18) which makes it an unavoidable activity in all human activities (Joint Committee on Standards for Educational Evaluation, 1981). This fact led some authors to defend and transpose the evaluation practices and concepts produced in other fields to the curricular

evaluation field, enabling the expansion and development of methodology and knowledge regarding educational evaluation (Smith, 1981, 1986).

According to Jacob and Boisvert (2010), a key obstacle to broad based agreement about evaluation professionalization issues lies in the uncertain identity of the discipline. To be sure, different evaluation scholars use different definitions of evaluation. But as for other expert occupations, the lack of a monolithic definition does not undermine the legitimacy of a genuine knowledge-based practice.

Each definition displayed in evaluation textbooks fits within a distinctive perspective about the major intent of the function. Some authors concentrate on the experimental tradition of the discipline and define evaluation as *the systematic, rigorous and meticulous application of scientific methods to assess the design implementation, improvement or outcomes of an activity or a program* (Rossi, Lipsey & Freeman, 2004) Others hold the view that the evaluator is merely “an educator (whose success is to be judged by what others learn)” (Cronbach & Associates, Towards Reform of Program Evaluation, 1980).

Chelimsky and Shadish (1997) stress the normative, proactive and socially useful characteristics of the evaluation discipline and its three-fold mandate: accountability (to measure results or value for funds expended, to determine costs and to assess efficiency); knowledge creation and dissemination (to generate insights about public problems, policies, programs and processes, to develop new methods and to critique old ones; and developmental (to strengthen institutions, to build agency or organizational capability).

To Picciotto (2011, p. 166), the elasticity of the evaluation term reflects the eclectic, diverse and adaptable character of the evaluator's craft, at the same time diverse definitions of the term articulate alternative answers to the 'how' and the 'why' questions.

In the evaluation literature authors vary in their emphasis of each of these three basic components: *valuing, systematic inquiry* and *use for decision making*. At first in all definitions the goal attainment is placed central and the value aspect is prominent, since whether or not program goals are attained provides the basis for judging it as either successful or unsuccessful. Thus Tyler (1950) defines evaluation as “the process of determining to what extent educational objectives are actually being realized”. Also Provus’ “*Discrepancy Evaluation Model*” (Provus, 1971) depends heavily on pre-established goals which serve as a basis for judging the success of a program.

Michael Scriven (1967) defends the idea of “*Goal Free Evaluation*” and also emphasizes the valuing aspect, although he denounces program goals as providing the basic orientation for making

judgements. Instead of goals and objectives the demands and needs of clients or relevant audiences of the program that is to be evaluated are seen as the basis for choosing evaluation standards (i.e. the norms used to determine “success” or “failure” of a program).

Guba and Lincoln (1986) warn of the need to integrate the human factor and the study of the context for a complexity of educational situations.

To Scriven (1991), evaluation is context driven, so if the study is in a different place, it is a different context. From the brief overview of views on the evaluation phenomena in the relevant literature it has also become clear that there are some important “contextual conditions” at stake when we deal with educational evaluation. The most important dimension on which these conditions manifest themselves is the variation in positions and interests in the evaluation process and outcomes of relevant parties. This realization gives cause to paying considerable attention to the political and organizational contexts.

The systematic inquiry, judgement, and use in decision-making settings are three elements present in Weiss’ definition of educational evaluation as “judging the value of educational objects, on the basis of systematic information gathering, in order to support decision making and learning” (1993, p. 94).

According to Weiss (1993, p. 93) evaluation is a rational enterprise that examines the effects of policies and programs on their targets – whether individuals, groups, institutions, or communities – in terms of the goals they are meant to achieve.

By objective and systematic methods, evaluation research assesses the extent to which goals are realized and looks at the factors that are associated with successful or unsuccessful outcomes. The assumption is that by providing “the facts”, evaluation assists decision-makers to make the choices among future courses of action. But evaluation is a rational enterprise that takes place in a political context, so political considerations intrude in three major ways that the evaluator should recognize and manage. First, the policies and programs are proposed, defined, debated, enacted and funded through political processes, and in implementation stage they remain subject to pressures that arise out of the play of politics. Second, because evaluation is undertaken in order to feed into decision-making, its reports enter the political arena. There evaluative evidence of program outcomes has to compete for attention with other factors that carry weight in the political process. Third, and perhaps least recognized evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as well: the problematic nature of some programs and the unchallenged ability of others; the legitimacy of program goals and program strategies; the utility of strategies of incremental reform, and even the appropriate role of

the social scientist in policy and program formation (Weiss,1993, p. 94). The pertinence of evaluation is the contribution on program operations' improvement, as recognized by Weiss (1972, p. 318):

The basic rationale for evaluation is that it provides information for action. Its primary justification is that it contributes to the rationalization of decision-making . . . Unless it gains serious hearing when program decisions are made, it fails in its major purpose.

All forms of evaluation consist of systematic information gathering and making some kind of judgment on the basis of this information. A further expectation is that this "valued information" is used for decisions on the day-to-day running of education systems or for more involving decisions on the revision and change of the system (Kellaghan, Stufflebeam & Wingate, 2003).

Evaluation results are expected to be used by relevant audiences and their results are expected to shape or, at least, have a certain impact on policy decisions (Scheerens, Glas, & Thomas, 2005, p. 730). There are authors who seem to altogether leave out the judgmental component from their definitions of evaluation and define evaluation in terms of providing information for *decision making*. Stufflebeam's earlier CIPP-model (2007) is an example of this as are authors who speak of "utilization focused evaluation" (Alkin & King, 2016). It could be argued that in these approaches the judgmental component is merely left implicit, since valuing is always there whenever information is interpreted as favoring or disfavoring a particular decision alternative.

Both elements of "*valuing*" and "*systematic inquiry*" are present in the definition presented by the Joint Committee on Standards for Evaluation, led by Daniel Stufflebeam: "*evaluation is the systematic investigation of the worth or merit of some object*" (Joint Committee on Standards for Educational Evaluation, 1981).

Program evaluation for the purposes of Stufflebeam's (2001, p. 11) monograph is characterized as "assessment of any set of coordinated activities directed at achieving goals". The purpose of the assessment is captured in his definition of evaluation as "a study designed and conducted to assist some audience to assess an object's merit and worth" (*Ibidem*). In other words, the evaluation is a program defined as coordinated activities to achieve goals. This understanding of program evaluation is common in the literature.

Evaluation is minimally defined as careful retrospective assessment of public-sector interventions, their organization, content, implementation and outputs or outcomes, which is intended to play role in future practical situations. Evaluation is primarily concerned with interventions, i.e. actions of some kind that are taken to influence the world, and not with persons, commodities or states of the world. Evaluation is equivalent to *ex post* evaluation, i.e. assessment of adopted, ongoing or

finished interventions, but excludes evaluation *ex ante*: calculated appraisals of the consequences of proposed and considered interventions that are performed *before* interventions are adopted and put into practice. It should be noted that assessments performed on interventions in empirical pilot trials are included in the evaluation category (Vedung, 2010, p. 264).

According to Alkin and King (2016) evaluation's contribution consists in improve programs, so it is intended to be a practical process. They defend that the broad history of evaluation as we know it today follows two streams: there is one stream, most heavily identified with the education field, focuses on tests and measurements; there is a second stream, a social science stream that focuses on the use of social research methodology (p. 569).

1.2. Historical perspectives on the evaluation concept

The purpose of this topic is to review the historical perspectives of the concept of evaluation and to proceed with the discussion of some models of educational evaluation that, from our point of view, constitute the substratum of the evaluation of human performance in educational contexts, and suggesting different ways of looking at the evaluation of educational processes. Theoretical concepts related to educational assessment are described through the theories of Ralph W. Tyler, L. J. Cronbach, Michael Scriven, Daniel L. Stufflebeam and Robert E. Stake.

1.2.1. Evaluation and objectives (Ralph W. Tyler)

Ralph W. Tyler (1902-1994) belongs to the short list of individuals dominating the educational scene during the second half of the 20th century. His scholarly contributions to the fields of evaluation and curriculum development prior to 1950 were at the core of discourse in both for the decades afterward. To prove it were the genesis and development of the National Assessment of Educational Progress, the Center for the Advanced Study in the Behavioural Sciences, the "Eight Year Study" and a host of federal, state, institutional and organizational educational improvement initiatives from the 1950s through the 1980s.

He has made substantial contributions to the fields of curriculum, testing, evaluation which earned him the title "the father of evaluation as a field of study" (Goddard, 1995, p. 78). Directly through his work, and indirectly through his many famous students, he has deeply influenced many noteworthy developments in education, including objective-referenced testing, objectives-based program evaluation, mastery learning, achievement test construction, item banking, the taxonomic classification of educational outcomes, and cooperative test development. He has been instrumental in the development of several national testing programs including the General

Educational Development Program, the Cooperative Testing Program, and the National Assessment of Educational Progress (Madaus & Stufflebeam, 1989).

Ralph W. Tyler was born in Chicago, in 1902. Tyler received his bachelor's degree in 1921 from Doane College in Crete, Nebraska. Immediately following his college graduation, Tyler began a teaching position in a high school in Pierre, South Dakota. He then proceeded to earn his master's degree from the University of Nebraska in 1923 and his Ph.D. from the University of Chicago in 1927. In 1927, Tyler joined the faculty at Ohio State University.

Tyler's work has to be analyzed in the context of the Great Depression of 1929, shortly after Tyler arrived at the Ohio State University. People were worried about their material losses and blamed much of it on the banks, the government and the schools. The newspapers were reporting how bad the schools were, and a big conference was held in 1933 on "*The Crisis in Education: Will the Schools Survive?*". The economic recession led to an environment of deep social and political crisis that led to the attribution of education and schooling to become effective systems for the development and modernization of American society at that time (Stufflebeam & Shinkfield, 2007). Between 1929 and 1938 Ralph Tyler served in the Bureau of Educational Research and Service at The Ohio State University. One of his major activities involved assisting professors throughout the University to evaluate their courses with the aim of improving instruction, the teaching skills and ultimately student learning. The faculty was also concerned about the large numbers of students who failed to return after the first semester each fall.

Ralph W. Tyler proposed an evaluation model based on the model of curriculum development that started from the previous definition of educational objectives and taking them as the main references of the assessment and the measurement of the school performances. One of Tyler's many contributions had been to specify learning objectives and to evaluate whether existing tests assessed those objectives adequately (Shepard, 2016).

To Castro (2015) the educational evaluation in Ralph W. Tyler refers to the first ideas of evaluation in the thirties, when he proposes that to evaluate it is necessary to identify objectives and characteristics of the object that will be evaluated.

According to Tyler (1982) the evaluation should not be limited only to school performance, but also consider curriculum. Rather than assessing only the students' abilities, Tyler's assessment model was based on the degree to which the program objectives were achieved. And considering that the evaluation has a decision making component, Tyler proposed a systemic approach to evaluation, since the results observed would allow us to conclude on the effectiveness of the curriculum and,

at the same time, to identify those aspects that needed to be adjusted. Tyler's proposal, although simple, modified the evaluation that was focused on the abilities of the individuals, to consider now the curriculum (Castro, 2015).

During his work at Ohio State University, he coined the term "educational evaluation" to encompass his idea of aligning measurement and testing with educational objectives. His concept of evaluation consisted of gathering comprehensive evidence of learning rather than from written testing. Tyler was the first educator to propose "evaluation" in the sense of determining and then attempting to assess the values inherent in the ongoing educational activity (Goddard, 1995, p. 77). It was during his long tenure as the first director of the *Center for Advanced Study in the Behavioural Sciences* (beginning in 1953) that Ralph Tyler had his greatest influence on educational research from virtually every perspective: economic, political, social, historical, philosophical, comparative, anthropological, methodological and aesthetic (Goddard, 1995, p. 80).

There were also the social, cultural and political constraints of the time, notably in the context of the "cold war" and the "space race", played by the great world powers of that time (United States of America and Union of Soviet Socialist Republics). This (also known as) "Star Wars" came to surface and accelerate the tendency to adopt a certain instrumentality of school education in favor of the scientific, technical, technological, social and labor development of the contemporary world. This trend was reinforced by the launch of the Sputnik I in 1957 by the Union of Soviet Socialist Republics. Faced with this alleged scientific backwardness, the government of the United States of America reacted with a large investment in national education, reporting education to the ideals of progress and development of the USA. To change the course of the nation's development, new educational objectives and contents were defined (M. Fonseca, F. Gonçalves, A. Mouraz and H. Ramalho, 2004, p. 79).

New procedures and modalities of evaluation were introduced, standardizing the evaluation instruments, transforming a school evaluation into a national system of evaluation that was deeply standardized with technical and bureaucratic rationality. What forced the curriculum and educators to develop new goals, contents and learning experiences that should be promoted by the school (W. Doll Jr., 1997, p. 70).

These social, cultural and political circumstances have led to the emergence of new evaluation standards within schools, framed in the general assumption of "school performance measure" (Stufflebeam & Shinkfield, 1989, Wolf, 1992), based on the comparison between the actual

performance and the desired performance, and having as reference the "standard criterion" or "standards of acceptability" (Popham, 1977, pp. 13-15).

Early presidents of the *American Educational Research Association* were leaders in the testing movement. Their intentions were to improve education by means of testing, which included both IQ and achievement tests. Early measurement experts acknowledged in scholarly articles that IQ tests could not measure inherited ability of groups with vastly different opportunities to learn, and yet ability testing was promoted as a beneficial means for matching instruction to individual differences until the insights of the civil rights era in the 1960s.

Shepard (2016) describes examples of the expansion of testing during the middle part of the twenty century are seen in papers by *American Educational Research Association* (AERA) presidents Alvin C. Eurich (1945–1946) and Arthur E. Traxler (1950–1951). Eurich's (1944) review article described the development of psychological tests to classify officers and enlisted men for the U.S. Navy during World War II. Traxler (1952) reviewed tests for selection into graduate schools, most particularly the Graduate Record Examinations and the Miller Analogies Test. Walter W. Cook (1958) wrote a paper for NCME about "What Teachers Should Know About Measurement." Cook's list included knowledge of percentile ranks and item-discrimination calculations as well as consideration of the mental processes involved in answering test questions. Julian C. Stanley's presidential address was entitled "Reliability Revisited". Stanley's contributions in measurement were most famously associated with his use of the *Scholastic Aptitude Test* (SAT) to identify precocious youth. Chester Harris wrote an overview for the 1962 *Review of Educational Research* issue on "Educational and Psychological Testing." His is perhaps the one slightly critical paper among all the papers written in this period. In summarizing the progress that had been made since Binet, Simon, and Terman, he noted that "we have a considerable amount of machinery" but nevertheless "seem to be studying many of the same problems in the same way" (p. 103). Harris, an expert in factor analysis, acknowledged that it was a good thing that the field had learned not to "announce the existence of an aptitude, mental ability, or personality trait on the basis of naming a factor derived from some conveniently available set of test responses" (p. 104).

In 1969, Ralph Tyler led the effort to establish the *National Assessment of Education Progress* (NAEP). Tyler (1969) distinguished objective-referenced tests from norm-referenced tests designed for sorting and argued that NAEP was needed to provide census-like data to monitor achievement trends over time. Beginning with the SAT test score decline from the mid-1960s onward, tests became both the messengers of crisis and the means to institute reform, through successive waves

of test-based accountability: minimum-competence test in the 1970s, back-to-basics tests in the 1980s, standards-based reforms in the 1990s, and the *No Child Left Behind* (NCLB) Act in the 2000s. Each new wave brought with it a ratcheting up of both standards and stakes attached to test results (Shepard, 2016, p. 117).

According to Popham (1977) Tyler's evaluation model of simplicity is based on the assumption that evaluation should occur with a procedure of limited scope and simultaneity with a outlined scope to avoid the use of subjective criteria and indicators of measurement. The purpose of the Tyler evaluation model is to verify only the degree of achievement of the defined objectives (B. Bloom, J. Hastings & G. Madaus, 1983). These objectives are therefore designed before the scholar and educational experience and in a totally technocratic way.

Differently, Ben Wood (one of the most prominent figures in testing) argued that it was sufficient to limit testing to measures of recall and recognition of facts and knowledge. If students had the necessary knowledge, they could apply it to solve problems. Tyler argued and then demonstrated that measures of higher order objectives such as application did not correlate highly with measures of mastery of factual knowledge. He concluded that if higher order cognitive skills were the objectives of instruction they had to be measured directly (Madaus & Stufflebeam, 1989, p.xii).

Tyler also distinguished between the educational objectives and the objectives of the curriculum. Educational objectives can be broadly based and written in broad terms, or even vague - for example, education should prepare the person for life, or education should stimulate critical thinking skills. For the objectives of the curriculum, "Tyler resorts to Bobby's notion of structuring these goals in terms of the practical and professional work needs of contemporary society [...]. Although Tyler does not use contemporary society as the only source of needs - the student's interests and the nature of the subject are also studied - the needs of contemporary life dominate." (Doll Jr., 1997, p.70).

Another example of Tyler's past work that still influences the fields of curriculum development and program evaluation is his work in the *Eight-Year Study* (1933–1941). The Eight-Year Study was a longitudinal assessment of students from thirty progressive and traditional high schools through four years of secondary school and four years of college. The study involved 30 high schools and over 300 colleges and universities that addressed overcoming inflexibility and restriction in high school curriculum, with the purpose to evaluate schools' efficiency. Ralph Tyler from of the Ohio State University wanted to demonstrate that the public school was efficiency in relation to progressive secondary school (Castro, 2015).

The very existence of private and public progressive schools was threatened in the mid-thirties when many prestigious colleges changed their entrance requirements. They began to require courses in specified subjects like American History and Physics. Since the progressive schools were not organized around courses but around learner outcomes, this new requirement posed a real difficulty for them (Madaus & Stufflebeam, 1989, p. xii).

At the close of the study, almost 1,500 students had attended college from across the study group, with little difference in academic performance based on grade point average and other factors, with the students in experimental schools slightly edging out those in traditional schools (Kliebard, 2004).

This comparative study, in which education researchers employed an experimental design, provided a good example of a large-scale investigation beyond a school survey. Although the use of education statistics was primarily centered on how to better use resources and reduce waste in K–12 school districts, researchers began to see how statistics could be used to address unequal conditions in education more broadly (Hedges et al. 2016, p. 149).

While both Terman and Thorndike believed that intelligence was inherited and fixed, other educators like Harold Rugg and the Progressive Education Association educators questioned that notion, they believed teachers could achieve societal change through education and that students could learn and grow through curriculum. In 1917, Rugg published a textbook on statistics for teachers with the hope that they would learn to use statistics as a tool of social science. The Progressive Education Association was one of the more prominent users of statistics during the World War II era, and the *Eight-Year Study* (1932–1940) is a good example of this approach (Kliebard, 2004).

This research was a decisive test to promote the diffusion of Tyler's conception of curriculum and educational evaluation, thus creating the necessary conditions for its model to influence educational systems in the following decades (Fonseca, Gonçalves, Mouraz & Ramalho, 2004).

Currently there is intense interest in teaching, and in evaluating students' higher order thinking skills (euphemistically referred to as HOTS). The work of Tyler and his colleagues in the Eight-Year Study in defining and developing indicators of what they called critical thinking skills should provide a rich source of ideas about HOTS and their measurement.

The volume IV of the Eight-Year Study (1942) *Appraising and Recording Student Progress*, set Tyler's original purposes of evaluation, namely: to provide a periodic check on the effectiveness of the educational institution and to indicate the points at which improvements in the program are

necessary; to validate the hypotheses upon which an educational institution's operation is based; to access information basic to effective guidance of individual students; to enhance the psychological well-being of the school community; and to provide a sound basis for public relations.

Behind these five purposes is Tyler's eight assumptions about evaluation:

- i. Education seeks to change behavior.
- ii. These behaviors are the schools' objectives.
- iii. Programs are appraised by finding out if and how objectives are being achieved.
- iv. Behavior is too complex to be measured by a single dimension.
- v. In whatever forms behavior appears, it should be appraised.
- vi. Any device that provides evidence of student progress is appropriate.
- vii. The kind of evaluation done directly influences learning and teaching.
- viii. Evaluation is the responsibility of the school staff and its clientele.

So, Tyler's concept of evaluation was based on a set of principles and assumptions about the nature of learning and teaching. The objectives-based model of evaluation is clearly delineated. It would have been helpful if the editors had commented on the importance and popularity of the Tyler approach to program appraisal and its influence on the development of other evaluation models (Stone, s/d, p. 103). According Tyler (1981, p. 105):

you can't just use the objectives as the basis for comprehensive evaluation. But certainly it was very important for people starting a program ... [to] ... find out whether they were accomplishing their purposes. But it is also important to find out many other things in order to understand what is going on in a program and to guide it.

The evaluation model recommended by Tyler (1949) focuses deeply on measuring and assessing the subjects' behaviors and performances throughout the teaching-learning process, in order to verify whether or not behavior changes, according the nature of the act of educate. Therefore, the objectives should be able to evaluate if there are changes in the subjects' behavior, as the author states:

We are not able to evaluate a teaching program when students are tested only at the end of the program. Without knowing what were the students' conditions in the beginning of the teaching program, it is not possible to determine to what extent there were modifications. In some cases, it's possible they have made considerable progress towards the objectives, before they started the teaching program. (Tyler, 1949, p. 99).

The evaluative practices should be based on an unambiguous definition of educational objectives, which are the "central unit" in Tylerian model. So, the process of selection and structuring of learning experiences, as well the selection of didactic and pedagogical resources depend on these objectives. And the organization of curriculum content serves as the main reference point through

which the students' performance tests are developed and applied (M.^a Fonseca, M.^a Gonçalves, A. Mouraz & H. Ramalho, 2004).

The *Constructing Achievement Tests* (Tyler, 1934) grew out from Tyler's experience with the Service Studies. It presents a generalized approach to the construction of achievement tests. This approach of test construction involves clearly defining important learning outcomes, collecting the most direct indicators of these outcomes, recording the outcomes, improving the reliability of the indicators, and validating indirect measures of these outcomes against direct indicators of them. The essential validity question of the correctness of the inferences made from a test is answered through a demonstration of the test's correspondence with direct indicators of the trait in question. *Constructing Achievement Tests* (1934) also synthesizes the results of the Wood/Tyler debate over the necessity of measuring all outcomes of learning. It includes reference to the results of a number of applied research studies from the *Service Studies in Higher Education* (1932) book. Finally, these selections offer an excellent description of the process of constructing of "criterion-referenced" or "objective-referenced" tests. The examinations Tyler describes are not norm referenced, standardized achievement tests. Contrarily, the construction of examinations should be a cooperative process involving five key steps. For example, (1) the first step in the task of constructing examinations in botany and zoology was to define the objectives which students were expected to attain as a result of instruction in these subjects. For purposes of examination, the objectives were expressed in terms of the behavior expected of students as a result of instruction in the course. For example, one objective expected of students is an ability to recall important facts, principles, and technical terms. (2) A second is an ability to formulate reasonable generalizations from the specific data of an experiment. (3) A third is the ability to plan an experiment which might be used to test a given hypothesis in botany or in zoology. (4) A fourth, the ability to apply general principles to new situations. (5) A fifth, skill in the use of the microscope. The nature of these objectives determines the variety of achievement tests to be constructed. Measurements of the information which a student recalls may be made with a paper-and-pencil examination, while a test of skill in the use of the microscope, for example, would require a different set-up.

So, there are many types of examinations required in college achievement tests as indicated by the projects in examination improvement which are in progress in several departments of the University. Thus far these departments have formulated at least ten major types of objectives each of which will probably require different types of examinations. It has been necessary to develop new types of examinations for many of the objectives.

According to *The Taxonomy of Educational Objectives*, there are ten types of educational objectives instructors should consider. Requirements of sound evaluation by instructors includes: working from unambiguous definitions of student behaviors that specify what a student who has attained the objective can do or produce; specifying the situations where, or ways in which, students can demonstrate the behaviors of interest; determining appropriate standards; using multiple approaches to measurement; assessing all types of behaviors that are significant in the educational development of students; keeping records of student progress; and developing scales and scoring schemes that convey useful information.

Instructors should strive to assess all the important objectives of the course, not just those involving mastery of the knowledge and facts associated with the course.

All methods of evaluating human behavior involve four technical problems: (1) defining the behavior to be evaluated; (2) determining situations where the behavior can be observed; (3) recording the behavior; and (4) evaluating the recorded behavior.

Tyler formalized his thoughts on viewing, analyzing and interpreting the curriculum and instructional program of an educational institution in *The Basic Principles of Curriculum and Instruction* (1949), which was based upon findings gathered from the Eight-Year Study (Ornstein and Hunkins, 1998). Tyler describes learning as taking place through the action of the student: "It is what he does that he learns, not what the teacher does" (Tyler, 1949, p. 63). The book laid out a deceptively simple structure for delivering and evaluating instruction consisting of four parts that became known as the Tyler Rationale:

1. *What educational purposes should the school seek to attain?* – Defining appropriate learning objectives.
2. *What educational experiences can be provided that are likely to attain these purposes?* – Introducing useful learning experiences.
3. *How can these educational experiences be effectively organized?* – Organizing experiences to maximize their effect.
4. *How can we determine whether these purposes are being attained?* – Evaluating the process and revising the areas that were not effective.

Through the development of these four principles (Tyler, 1949, p. 51), Tyler introduced educational ideas leading to new considerations regarding the measurement of outcomes. This measurement of outcomes occurs by developing a list of program or curriculum objectives that indicates both the

kind of behavior to be developed in the student and the area of content or life in which the behavior is to be applied (Keating, 2006).

These educational objectives become the criteria by which materials are selected, content is outlined, instructional procedures are developed, and tests and examinations are prepared. All aspects of the educational program are really means to accomplish basic educational purposes. Hence, if we are to study an educational program systematically and intelligently we must first be sure as to the educational objectives aimed at. (Ibid, 1949, p. 52).

Tyler recognized “objectives are matters of choice, and they must therefore be considered value judgments of those responsible for the school” (Tyler, 1949, p. 52). So he defends a comprehensive philosophy of education to guide in making these judgements, as well certain kinds of information and knowledge provide a more intelligent basis for applying the philosophy in making decisions about objectives. “*If these facts are available to those making decisions, the probably is increased that judgements about objectives will be wise and that the school goals will have greater significance and greater validity*” (Ibid). For this reason, a large part of scientific studies on the curriculum field from decade of 1920 to 1950 has concerned with investigations that might provide a more adequate basis for selected objectives wisely. The technical literature of the curriculum fields includes hundreds of studies that collected information useful to curriculum groups in selecting objectives (Ibid).

The further question is then raised what sources can be used for getting information that will be helpful in this way. “A good deal of controversy goes on between essentialists and progressives, between subject specialists and child psychologists, between this group and the school group over the question of the basic source from which objectives can be derived.” (Ibid) and each of these groups defends different sources of information in objectives’ definition.

The progressive emphasizes the importance of studying the child to find out what kinds of interests he has, what problems he encounters, what purposes he has in mind. The progressive sees this information as providing the basic source for selecting the objectives. The essentialist, on the other hand, is impressed by the large body of knowledge collected over many thousands of years, the so called cultural heritage, and emphasizes this as the primary source for deriving objectives. The essentialist views objectives as essentially the basic learnings selected from the vast cultural heritage of the past (Ibid).

On one hand, the sociologists, concerned with the pressing problems of contemporary society, see in the analysis of this society the information from which objectives can be derived. Since they view the school as one agency in helping young people to deal effectively with the critical problems of

contemporary life, there are the objectives of the school to provide those knowledges, skills, attitudes and abilities that will help people to deal with these problems. On the other hand, the educational philosophers recognized that basic values in life, transmitted from one generation to another, are the basic source from which objectives can be derived. They see the school as aiming essentially at the transmission of the basic values derived by comprehensive philosophical study. (*Ibid.*, pp.52-53).

In turn, Tyler defends that no single source of information is adequate to provide a basis for wise comprehensive decisions about the objectives of the school. Each of these sources has certain values to commend it. Each source should be given some consideration in planning any comprehensive curriculum program (*Ibid.*, p. 53).

The Tyler rationale also highlighted an important set of factors to be weighed against the questions. Tyler believed that the structure of the school curriculum also had to be responsive to three central factors that represent the main elements of an educative experience: (1) the nature of the learner (developmental factors, learner interests and needs, life experiences, etc.); (2) the values and aims of society (democratizing principles, values and attitudes); and (3) knowledge of subject matter (what is believed to be worthy and usable knowledge). In answering the four questions and in designing school experience for children, curriculum developers had to screen their judgments through these three factors.

Tyler (1950, p. 44) argued that

Since the real purpose of education is not to have the instructor perform certain activities but to bring about significant changes in the students' pattern of behavior, it becomes important to recognize that any statements of objectives of the school should be a statement of changes to take place in the students.

Tyler's rationale has been criticized for being overtly managerial and linear in its position on the school curriculum. Some critics have characterized it as outdated and a theoretical, suitable only to administrators keen on controlling the school curriculum in ways that are unresponsive to teachers and learners. The most well-known criticism of the rationale makes the argument that the rationale is historically wedded to social efficiency traditions. Tyler offered no substantive response to these criticisms, believing that criticism of his curriculum development work required some discussion of an alternative, which none of the critics provided.

According to Fernandes (2010), the Tyler's conception of the curriculum as a product based on previously established objectives is seen as

a previously planned product, which translates into a linear process that corresponds to an organized and structured study plan based on objectives, contents, activities and procedures of evaluation". The author defends "a broad, procedural and dynamic character of the curriculum, which assumes an embracing, open and flexible purpose, which determines the process of its development. (pp. 33-34).

The dynamic character of the curriculum is also highlighted by Pacheco (2001, p.17) who argues that "the curriculum should not be seen as a plan, fully planned, but as an organized whole according to previously planned issues, the context in which curriculum is implemented and the knowledge, attitudes, values, beliefs that the participants bring with them, with the appreciation of experiences and learning processes."

Young (2010) identifies two characteristics of the curriculum: on the one hand, a neo-conservative traditionalist component that works the curriculum as a cluster of knowledge transmitted by the school institution; on the other, the techno-instrumentalism that arises at a time when schools are governed by *rankings* and in which, while having freedom to manage their own curriculum, the most economically fragile institutions are influenced. Here, curriculum and education are not an end, but a way to respond to the market's needs, combining academic knowledge with practical/professional knowledge. This last position approaches the curriculum's vision as something in permanent construction, delineated according to the actors and the social capital they possess, and must be built in a consistent logic with the individual's beliefs, values and experiences.

A curriculum indicates what is intended should happen in a program of learning and the circumstances in which these activities can take place. The activities referred to here are learning activities; a curriculum is a collection of exercises and tasks that culminate in learning of one type or another. There are three fundamental types of learning: cognitive, skill-based and dispositional, and they have different forms and operate in different ways. Cognition comprises the manipulation of those symbolic resources (words, numbers, pictures etc.) that point to something outside itself. Skill-based knowledge is different from cognition because it is procedural and not declarative; and dispositional knowledge refers to relatively stable habits of mind and body, sensitivities to occasion and participation repertoires. More importantly, these three types of learning are knowledge-oriented, so an argument can be made that learning is a knowledge-development activity. And what follows from this is that how we construe knowledge will determine how we construct productive learning environments and ultimately how learners then learn in and from them (Scott, 2016, p. 2).

The Tyler model has several advantages: It is relatively easy to understand and apply. It is rational and systematic. It focuses attention on curricular strengths and weaknesses, rather than being concerned solely with the performance of individual students. It also emphasizes the importance of a continuing cycle of assessment, analysis, and improvement (Chapter 12, p. 360).

As Guba and Lincoln (1981) pointed out, however, it suffers from several deficiencies. It does not suggest how the objectives themselves should be evaluated. It does not provide standards or suggest how standards should be developed. Its emphasis on the prior statement of objectives may restrict creativity in curriculum development, and it seems to place undue emphasis on the pre-assessment and post-assessment, ignoring completely the need for formative assessment.

Similarly, Baron and Boschee (1995, p. 1), in their book *Authentic Assessment: The Key to Unlocking Student Success*, stress that “we are encountering fundamental changes in the way we view and conduct assessment in American schools” (...) and “sixty years have passed since we experienced such a deep-seated and thoughtful reevaluation of our assessment methods”.

1.2.2. Evaluation and practice (L. J. Cronbach)

Lee J. Cronbach (1916-2001) made major contributions in the fields of educational psychology, psychological testing and program evaluation, throughout a career that spanned more than five decades. He focused not on the technical aspects of measurement in evaluation, but on the policy-oriented nature of evaluation. This was an idea that led to a radical reconstruction of internal and external validity, which included separating the two conceptually and conceptualizing external validity in relation to usability and plausibility of conclusions, not as technical feature of research or evaluation design (Cronbach L. J., 1982).

He was considered "one of the most important and least appreciated" theorists of the first generation of evaluation, argued the linkage of evaluation to the processes of knowledge construction, policy interaction and managerial decision-making (Shaw, 2003).

Cronbach, who advocated a pluralistic methodology with an emphasis on the balance of precision (reliability), used to criticize the submission of evaluators to managers, who controlled evaluative procedures. He defends that, in a context of political agreements, conservation and change, the evaluator's role should be critical and scholarly, even if it serves partisan interest (Prestes, 2012).

In the educational evaluation field, Cronbach was one of the methodological giants. His contributions include Cronbach's coefficient alpha, generalizability theory and notions about construct validity. Thus, his strong evaluation roots in methodology and social science research led us to place him in the group of method theorists (Christie & Alkin, 2013, p. 30).

According to Chen, Donaldson and Mark (2011) Cronbach had also a crucial influence on concept of validity which, most broadly, refers to whether one measures what one intends to measure, given the assumption of an objective reality.

Lee Cronbach, observing the inoperative evaluation results against the programmed objectives, severely criticized its procedures and proposed revisions of its research design, data analysis and the usual theoretical framework (Shaw, 2003).

Despite the criticism and some rejections, the evaluation of social policies gradually opened up spaces for their credibility, acceptance and prestige in social and academic circles, and established itself as a political and social necessity. This became evident in the early 1960s under the administration of President Kennedy when, in the face of the explosion of social policies, it was necessary to evaluate political responses on its efficiency, effectiveness employed resources. Meanwhile, the evaluation of social programs had become a specialized profession, with its own organizations, journals and directed studies, which contributed to the fact that the evaluation practices of social policies gained new adherents and became more visible and legitimized in the academic circles, political and business.

Between 1965 and 1975, through programs such as the *War on Poverty*, which aimed at improving the quality of life of the American population, and at the same time controlling the applied public money, the evaluation of social policies became a "great industry" (Rossi & Freeman, 1985, p.26). As their importance increased, specialists that worked in the areas of urban development, housing, technological and cultural education created their own organizations, magazines and studies. These initiatives led to the development of an evaluation literature in the form of manuals, texts and journals, such as *Evaluations Review*, *Evaluation News* and *Education Evaluation and Policy Analysis*, among others, and led to the emergence of the great professional societies – the *Evaluation Research Society* (of a more academic nature) and the *Evaluation Network* (Pacheco, 1998, p.11).

Cronbach wrote a key paper on social program evaluation entitled *Designing Evaluations of Educational and Social Programs* (1982) which have the purpose to guide evaluators. Cronbach identified two phases in the research of evaluation questions: the divergent phase and the convergent phase. The divergent phase consists in writing down on a sheet of paper, without any kind of self-criticism and filters, all the questions that arise – no question should be excluded. In the convergent phase it is a matter of selecting these questions and arranging them with certain logic. This arrangement will begin to show what is important and what is not. On the other hand,

the task of selecting evaluation questions is not unique to the evaluator. In the interviews with those responsible and even later in the course of the evaluation, questions will surely arise, which should be given due attention. The evaluator should have some flexibility to accept integrating them into their schemes. This should not, however, undermine the decisive role of the evaluator in the final organization of the evaluation question list. The formulation of the questions can be in several formats, and it is best to start them with the following words: "To what extent ..." – thus, there is an indication that the criteria must be discriminatory, which is not the case in questions with a 'yes' or 'no' answer.

Cronbach was a Ralph Tyler's student, lending his work a kind of field orientation not present in the work of many social science researchers. Furthermore, his association with more policy-research-oriented Stanford University colleagues, notably in his book *Toward Reform of Program Evaluation* (Cronbach & Associates, 1980, pp. 65-66), helped to establish his concern for evaluation's use in decision making (Christie & Alkin, 2013, p. 30).

Following on the work of Weiss, Cronbach rejects the simplistic model that assumes a single decision maker and "go/no-go" decisions. Instead, he views evaluation as an integral part of policy research focused on policy shaping communities and necessitating potential political accommodation. Thus, he is more oriented toward evaluation utilization for enlightenment purposes than toward a concern for more immediate instrumental use. He does, however, affirm the evaluator's active role is providing data to local decision makers for instrumental use (in accordance with contractual obligations) (*ibid*).

To define the manner in which evaluators might most productively enhance enlightenment use, Cronbach coins a set of symbols to define the domains of evaluation (Cronbach, 1982). These domains consist of units (populations), treatments, observations (outcomes) and settings (UTOS). The sample of these domains examined by the evaluator is referred to by the acronym "UTOS" and it represents the larger population from which sampling took place, which is referred to as "UTOS." Cronbach maintains that the most important area of concern for evaluation is external validity, which he refers to as the plausibility of the conclusions to UTOS that are manifestly different from the population under study. He refers to this manifestly different population as "*UTOS" (pronounced "star-UTOS"). In the concern for evaluations contributing to enlightened discussion, Cronbach focuses on what he refers to as "bandwidth," by which he means that it is more important for the evaluation to focus on a broad range of relevant issues than to achieve absolute fidelity (accuracy) on a small number of issues.

Cronbach's concern about generalizing to *UTOS leads him to reject Campbell and Stanley's emphasis on experimental design and Scriven's focus on comparison programs (Christie and Alkin, 2013, p. 31). He proposes that generalization to *UTOS can be attained by extrapolating through causal explanation, either using causal modeling or the "thick description" of qualitative methods. Furthermore, it is sometimes beneficial to examine subpopulations (sub-UTOS). Thus, focusing on the subset of data for a particular group might enable generalization to other domains. Furthermore, he seeks to capitalize on naturally occurring variability within the sample as well as consequences of different degrees of exposure to treatments. This work is an extension of earlier research on aptitude-treatment interactions conducted jointly with Snow (Cronbach & Snow, 1977). Cronbach displays sensitivity to the values of the policy-shaping community and seeks to incorporate their views on the evaluation questions most worth asking. This is done systematically with an eye to what will contribute most to generalization: issues receiving attention from the policy-shaping community; issues relevant in swaying important (or uncommitted) groups; issues having the greatest uncertainty; and issues that would best clarify why a program works. Shadish et al. (1991, p.340) make the following keen distinction between Cronbach and several other major theorists:

[Cronbach views] evaluators [as] educators rather than [as] the philosopherkings of Scriven, the guardians of truth of Campbell or the servants of management of Wholey.

As such, Cronbach does not aspire to the instrumental use of evaluation as do Wholey and others on the decision-making/use branch. And he does not adhere to the strict experimental controls that Campbell advocates. Likewise, he does not call upon the evaluator to impose his own value system on the program or to reach summary judgments about programs. Rather, he sees the evaluator's role as providing "readers" with information that they may take into account when forming their own judgments.

We have placed Cronbach on the methods branch of the evaluation tree, but branching off in the direction of the use branch. This placement in part reflects the influence of Weiss on his ideas with respect to enlightenment use directed toward policy-shaping communities.

For example, there is a longstanding debate between Donald Campbell and Lee Cronbach regarding the context and generalizability of different evaluation designs (Patton, 2012). This debate centers on the importance of experiments and quasi-experiments *versus* the more contextualized approaches when explaining the causal relationships of programs and their outcomes (Cronbach, 1991). While our discussion does not focus on these debates, it is important

to recognize that there might be disagreement on the purpose of the various approaches to program evaluation.

The first, and historically dominant, evaluation model is science-driven and highly technical with a strong emphasis on quantitative methods. Scientific research methods and techniques such as systematic randomized research designs and experiments are considered integral to the evaluation process (Alkin, 2013). This approach to evaluation focuses on measuring effectiveness and efficiency (Greene, 1994). For example, a common measure for evaluation might include the extent to which programs promote or not the realization of goals or objectives. Such an approach requires that evaluators clearly identify goals and objectives and be able to measure them through quantitative processes.

An evaluative study of a social program is justified to the extent that it facilitates the work of the polity. It therefore is to be judged primarily by its contribution to public thinking and to the quality of service provided subsequent to the evaluation. An evaluation pays off to the extent that it offers ideas pertinent to pending actions and people think more clearly as a result. To enlighten, it must do more than amass good data. Timely communications—generally not “final” ones—should distribute information to the persons rightfully concerned, and those hearers should take the information into their thinking. To speak broadly, an evaluation ought *to inform and improve the operations of the social system*.

Cronbach and Suppes (1969) provided further basis for understanding that the expectation of use is an important part of evaluation. They referred to both evaluation and research as forms of “disciplined inquiry,” noting that there were differences between these two types of disciplined inquiry. They distinguished “decision-oriented” from “conclusion-oriented” investigations. A decision-oriented study is one in which “the investigator is asked to provide information wanted by a decision maker: a school administrator, a governmental policy-maker, the manager of a project to develop a new biology textbook, or the like” (1969, p. 20). The conclusion-oriented study, by contrast, takes its direction from the investigator’s commitments and hunches and is guided by the desire to add to a disciplinary base. Thus, evaluations address the here and the now (i.e., this program at this time) and attempt to provide insights intended to lead potentially to the use of findings for program improvement either in the form of decisions or better program understanding. The final period of testing and assessment history, from 1970 to the present day, is characterized by a dramatic decline in ability testing and a concurrent steep rise in use of standardized achievement tests to hold students and schools accountable. As Cronbach (1975, p.11) noted,

“*there is a tide in the affairs of issues*”, meaning that some ideas gain traction, because of the *Zeitgeist* of a period, that would be ignored or ridiculed at a different time. The long-standing complaints against the validity of IQ tests – in the face of unequal opportunities to learn – and the harm of sorting students into dead-end school placements came to the fore during the civil rights era. The Civil Rights Act was passed in 1964, and under its authority, IQ testing was challenged both for employment and placement in special education.

In 1972, in *Larry P. Riles*, the court agreed with claims of racial bias in California’s use of IQ tests, based on evidence that smaller proportions of black students were identified as mentally retarded in states where other criteria were used such as achievement tests and teacher evaluations (Shepard, 2016).

As they had in the early 1900s, measurement experts once again became active in helping educators and policy leaders think about how to use tests. When Cronbach was AERA president in 1964-1965, he was not writing about testing, but he did so significantly in 1975 in his examination of Arthur Jensen and the longer-term history of IQ testing cited above. He was also a member of the *National Research Council (NRC) Committee on Ability Testing*, which included leading psychometricians and statisticians, Lyle Jones, Melvin Novick, Mary Tenopyr and John Tukey, along with relevant experts in each of the social sciences, including learning researcher Lauren Resnick. The *NRC Ability Testing* report (Wigdor & Garner, 1982) was cautious, attempting to balance the need for comparative data – which was possibly less biased than other indicators – against evidence of past harms from test misuse. They gave an example of a man, whose father was a mathematician, who scored 700 on the GRE mathematics test, compared to a woman of working-class background who scored 650. The committee argued that the woman could be the more exceptional candidate. The committee recommended against the use of tests as the sole criterion for selection decisions and against the use of rigid cutoff scores for special education placements. A similarly important NRC report, also published in 1982, addressed assessment issues associated with the overrepresentation of minorities in classes for the mentally retarded (Heller, Holtzman & Messick, 1982).

The civil rights movement that greatly diminished ability testing also generated renewed interest in achievement testing— because resources for equity would require accountability. The equity agenda of the 1960s launched numerous federally sponsored social programs focused on preschool education, job training, health care and housing. The first Elementary and Secondary Education Act (ESEA) of 1965 aimed at providing greater educational opportunities for low-income

students, also created the field of educational evaluation (Fitzpatrick, Sanders, & Worthen, 2011) (Shepard, 2016, p.11).

This false confidence in achievement tests as sufficient measures of learning or as adequate proxies to be used in judging educational programs is a popularized belief from those early days that still haunts us today. These examples also illustrate a pattern among measurement experts that was to acknowledge complaints about the adequacy of their measures but not to change course in their promulgation of tests and test batteries as effective tools in the service of educational improvement. This was true for both achievement measures and IQ tests, used initially as controls when evaluating the effectiveness of educational programs, but then increasingly as placement tests.

The history of intelligence testing and its connection with the American eugenics movement in the early 20th century has been told many times (Chapman, 1988; Cronbach, 1975). IQ testing did not create racism in America, but the theory of innate differences in merit among social groups was embraced because of prevailing beliefs. Moreover, the claim of scientific objectivity fed a system of beliefs – about nativist rather than opportunity explanations for attainments – that still today affects public discourse and the beliefs of educators. That’s why contemporary antiracism scholarship necessarily focuses so on understanding privilege and why effective educational interventions find it necessary to address the problem of “deficit thinking” (Valencia, 1997).

In addition to reifying racist beliefs, IQ testing created further harm by assigning low-scoring students to low-track and vocational classrooms. Terman (1919, 1922) reported IQ differences observed among various occupational groups, college students, businessmen, semiskilled, and unskilled laborers and argued in turn that different curricula and methods of instruction should be provided to children identified as “gifted,” “bright,” “average,” “slow,” and “special.” Cronbach (1975) summarized the reasoning and effects of tracking by IQ as follows: When the tests determined who would enter the college preparatory program and before that determined who would go into the “fast” section of an early grade, the tests began to determine fates. The testers’ sorting process was to shield the child destined to be a worker from the rigors of an academic curriculum. Such a plan would reduce distaste for schooling, prevent failure, and retain him in school longer. Testers said that the IQ was constant; hence to make decisions early was merciful and just (Shepard, 2016, p. 12).

Out of the trunk of social inquiry has grown a branch of evaluators who focus on concerns related to valuing or the making of judgments. Theorists on this branch believe that what distinguishes

evaluators from other researchers is that evaluators must place value on their findings and, in some cases, determine which outcomes to examine. This branch of evaluation has as its primary mainstay Michael Scriven, whose early work (1967, 1972a) was a brilliant exposition that largely defined the field in terms of its conceptual underpinnings and procedures.

1.2.3. Evaluation and its logic (Michael Scriven)

Michael Scriven is among evaluation's most influential theorists who focus the value and the making of judgments. He was the first to question the assumption that goals or objectives are crucial in the evaluation process (Scriven 1972) and the goal-free model was the outcome of this dissatisfaction. After Scriven's involvement in several evaluation projects where so-called side effects seemed more significant than the original objectives, he began to question the seemingly arbitrary distinction between intended and unintended effects.

It is Scriven (2003) who proclaims that an evaluation is not evaluation without valuing; in his words, "evaluation is the science of valuing". Shadish, Cook and Leviton (1991) noted that Scriven was "the first and only major evaluation theorist to have an explicit and general theory of valuing" (p. 94). Scriven defends that society requires valuing and it is the role of the evaluator to do that job – there is a science of valuing called evaluation. So, the evaluation should be a specific field of the education science based in scientific method (Rodrigues, 1999, p. 22). Credited with introducing discipline's foundational logic and core vocabulary, his prolific writings have significantly influenced evaluators' perspectives on what sets the educational evaluation from other disciplines that employ research tools (Alkin & Christie, 2004, p. 32).

But evaluation is not the application of social science methods to solve social problems, rather evaluator's use social science methods to examine the merit, worth and significance of a program, project or policy with the purposes of describing values associated with different stakeholder groups, as well as reaching evaluative conclusions "about good and bad solutions to social problems" (Scriven, 2012, p. 21). The transdisciplinary nature of evaluation allows its application in diverse contexts, with diverse stakeholders, to address diverse social problems, through the use of diverse methodologies. With these multiple dimensions and diversity, it should come as no surprise that there are also diverse sets of philosophical assumptions that underlie the choices that evaluators make regarding their methodologies (Mertens, 2015, p. 75).

Scriven's integrative thinking on evaluation reflects his study of mathematics, philosophy of science, psychology, physics, education, logic, and computer science. Scriven received his Ph.D. in Philosophy from Oxford University, where he studied under Gilbert Ryle. He has held faculty

appointments at numerous institutions including Claremont Graduate University, Auckland University, Indiana University, the University of California at Berkeley, Swarthmore College and the University of Western Australia. Currently, he is associate director of the Evaluation Center at Western Michigan University.

Early evaluation theorists, like Don Campbell, were strongly influenced by the positivists, who argued that facts were one thing and values quite another. Evaluators could discover the facts objectively but values (value claims) were essentially subjective, perhaps deep-seated emotions not subject to rational analysis or discussion. Hence, evaluators and social scientists should stick to the facts and leave values, as reflected in the program goals or evaluative criteria, to the politicians or program developers – in this view, evaluators simply had to accept the goals (House, 2012, pp. 11-12).

To House (2012), Scriven was the first theorist of evaluation to assert that value claims can be objective and so to reject the fact-value dichotomy. Value claims are like factual claims in that evaluators can make rational value claims by citing the evidence for particular claims. To Scriven, evaluators are particularly well equipped to do this, since they have data collection and analysis methodologies at their disposal. Furthermore, evaluators need to protect, insulate, and isolate themselves from pernicious influences, such as being pressured by political forces or becoming too friendly with the people, whose programs are being evaluated, thus eroding the evaluator's judgment inappropriately (p. 12).

According to Scriven (1991), *“evaluation is the process of determining the merit, worth, and value of things, and evaluations are the products of that process”* (p. 139). Evaluation as the *“judgment of merit and worth”* provide further insight – *merit* refers to intrinsic value and *worth* refers to extrinsic value or goodness. The notion of “worth” is, in essence, the idea that evaluations must have “utility” – the potential to be used in the given context at a given time. Some things may be intrinsically good but not of value to an individual or to an entity. More specifically, a program may have merit (i.e., it is considered to be of good quality), but it may not have value for a particular context. In its simplest forms, evaluation is *“close to the perceptual level, the near-instinctive process, sometimes inherited, sometimes learnt, which provides the kind of knowledge that links perception with action. It is, for example, learning that something is good to eat and so being able to eat it”* (Scriven, 2003, pp. 9-28). In its most complex form, *“evaluation elevates the simple process of instant or near-instant appraisal to the complex and systematic – and more reliable – investigation of elaborate systems, in this process becoming, of necessity, a complex discipline*

itself – it remains one of our key survival mechanisms, whether in the evaluation of complex weapons systems for the armed forces, or in the choice of answers to questions in a job interview (*Ibidem*).

To Scriven (1991), evaluation is a systematic reflection to determine the merits, validity and importance of things. As an activity inspired by the social sciences, evaluation can also be understood as a systematic examination of certain objects, based on scientific procedures for collecting and analyzing information about the content, structure, process and results and / or impacts of actions evaluated in reality.

According to Prestes (2012), Scriven prioritized the theoretical and methodological perspectives of evaluation, the ethical posture of the evaluator and his competence to extract the best interpretations of social reality. He preached that the evaluation is an autonomous discipline of Philosophy or Social Sciences and that its design should start from the needs of the users. In addition, evaluation practices should be outside the interests of the program and capable of pointing out effective solutions to social problems.

Scriven's major contribution is the way in which he defined the evaluator's role in valuing goals and in making value judgments. Scriven believes that an evaluator's role is to investigate and justify the value of an evaluand. Such investigation and justification shall be supported with joining empirical facts and probative reasoning. "Bad is bad and good is good and it is the job of evaluators to decide which is which" (Scriven, 1986, p.19). He rejects the notion that an evaluator's role is simply to provide information to decision-makers and claims that "the arguments for keeping science value free are in general extremely bad" (Scriven, 1969, p.36).

According to Scriven, there are four evaluator's responsibilities during an evaluation. At the first, evaluators should determine the criteria of merit from needs assessment, and the criteria of merit of an evaluand should be its capacity to meet needs. Although an evaluator can use the results of needs assessment conducted by a program developer, sometimes he/she should do an independent needs-analysis. To avoid bias, Scriven advises evaluators to conduct "goal-free" evaluation and formulate questions by ignoring the program goals and looking for all possible effects an evaluand could have.

Scriven (1991) argues that what distinguishes evaluators from other researchers is that evaluators must place value on their findings and, in some cases, determine which outcomes to examine (Alkin & Christie, 2004, p. 32). Evaluators should not examine only the stated objectives of a project, since other outcomes of the program may be equally important. So the evaluators should

launch a broad network in evaluating the results of a program, noting the intended and unintended results.

Secondly, evaluators should be setting comparative evaluation standards. A set of standards should be created by evaluators to assess the program performance. Such standards are used for comparison, either comparison with a set level of performance, or with alternative programs. The latter comparison is preferred by Scriven since he believes that an evaluator will usually make decisions about choosing among alternatives. The third evaluator's responsibility is to assess program performance. An evaluator will need to answer both the evaluative and non-evaluative questions. Evaluative questions focus on the effects of the program and should be given top priority. The evaluator should acquire the skills to collect and analyze both experimental and non-experimental data. At least, evaluator should be offering a final evaluative judgment, he should synthesize his findings into a final report and offer his/her summative judgment (Luo, 2010).

Scriven's position is new in the fact that Scriven differentiates evaluators from researchers or social scientists by emphasizing that the value judgment is an integral part of an evaluator's role and grounds such role in the logic of evaluation. His "goal-free" evaluation allows evaluators to identify possible side effects of the evaluand and address the concerns of underrepresented stakeholders (Luo, 2010).

Scriven (1986 p. 284 cit. in Christie & Alkin, 2013, p. 32) notes that the greatest failure of the evaluator is in simply providing information to decision makers and "passing the buck [for final judgment] to the non-professional." The evaluator, in valuing, must fulfill his or her role in serving the "public interest" (Scriven, 1976, p. 220). By public interest, Scriven does not restrict the evaluator's responsibility simply to clients, users, or stakeholders, but to all potential consumers. Indeed, he views the evaluator's role in valuing as similar to producing a report for Consumer Reports in which the evaluator determines the appropriate criteria for which judgments are to be made and then presents these judgments for all to see. As in Consumer Reports, there is the necessity for identifying "critical competitors," that is, competing alternatives.

According to Scriven (1991), evaluation is not the mere accumulation and summarizing of data that are clearly relevant for decision making. Gathering and analyzing the data that are needed for decision making comprises only one of the two key components in evaluation. A more straightforward approach is just to say that evaluation has two arms: only one of which is engaged in data-gathering, and other that clarifies and verifies relevant values and standards that qualify an evaluation (Scriven, 1991, pp. 4-5). Combined with the data research, another key component is

required to get to conclusions about merit and benefit or cost-effectiveness of evaluation. A third major component of evaluation, next to the valuing aspect and the systematic approach to information gathering, is the applied context.

Scriven states that it is not necessary to explain why a program or product works to determine its value. Nevertheless, he introduces an alternative to experimental and quasi-experimental design called the “modus operandi” (MO) method (Scriven, 1991, p. 234). Analogous to procedures used to profile criminal behavior, the MO method of a particular cause is an associated conGraphiction of events, processes, or properties, usually in time sequence, which can often be described as the characteristic causal chain (or certain distinctive features of this chain) connecting the cause with the effect (Scriven, 1974, p. 71).

In a manner similar to that which Scriven uses to grade critical competitors, the MO method requires that the evaluator first develop an exhaustive list of potential causes and then narrow down the potential choices. The list of potential causes is narrowed down in two steps. In the first step, the evaluator determines which potential causes were present prior to the effect. In the second step, the evaluator can determine which complete MO fits the chain of events and thus determine the true cause. To ensure accuracy and bias control, the evaluator looks for instances of “co-causation and over determination,” and calls in a “goal-free or social process expert consultant to seek undesirable effects” (Scriven, 1974, p. 76). Scriven believes that ultimately, the evaluator is able to deliver a picture of the causal connections and effects that eliminate causal competitors without introducing evaluator bias.

Furthermore, Scriven (2007) advocates “goal-free evaluation” in which the evaluator assumes the responsibility for determining which program outcomes to examine, rejecting the objectives of the program as a starting point. He maintains that by doing so, the evaluator is better able to identify the real accomplishments (and non-accomplishments) of the program. Goals are for guiding action, for planning and managing, but they are not the key issue for evaluators. The key issue for program evaluators is the effects the program actually had, measured in terms of what they meant to those you affected (and those you did not reach), whether or not you meant to have those effects; plus how much it cost to do what you did do; and what alternatives there were; and how you got to where you got (since the end does not justify all means); and other things, e.g., the potential uses of the program by contrast with the use under examination (Scriven, 2007, pp. 7-8).

1.2.3.1. Making evaluative judgments

In essential element of Scriven's valuing is the determination of a single value judgment of the program's worth ("good" or "bad"). In requiring a synthesis of multiple-outcome judgments into a single value statement, Scriven is alone among evaluation theorists. In both goal-free evaluation and in the synthesis stage, Scriven justifies his point of view by relying on the extent to which the program is able to meet "needs." Needs are the presumed cost to society and the individual and are determined through a needs assessment.

As Scriven himself notes, however, goal-free evaluation should be used to complement, not supplant, goal-based assessments. Used alone, it cannot provide sufficient information for the decision maker. Some critics have faulted Scriven for not providing more explicit directions for developing and implementing the goal-free model; as a consequence, it probably can be used only by experts who do not require explicit guidance in assessing needs and detecting effects. Scriven (1991) also popularized the terms formative and summative assessments as a way of distinguishing two types of roles that evaluators play: they can evaluate the merits of a program while it is still in development, or can evaluate the results of an already completed program.

There are two major types of evaluations: (1) those designed to distinguish worthwhile programs from ineffective ones, and (2) those designed to help improve existing ones in order to achieve certain desirable results. The former are often called formative evaluations, and they are conducted to provide information on how a program should be delivered or to furnish information for guiding program improvement. The latter are called summative evaluations, and they are conducted to determine whether a program's expectations are being met and what its consequences are. This is the kind of evaluation that our charge required (Scriven, 1991).

Summative evaluations generally focus on whether a given program (e.g., a social program, an educational intervention) is effective. For example, summative evaluations might study such issues as the program's accomplishment of its intended objectives, impacts beyond those that were intended, how effectively resources have been used, the benefits of the program and what it costs to produce these benefits, and alternative interventions that might produce similar benefits. Summative program evaluations usually focus on the effects of a program on outcomes for a client population and consider the extent to which the program changes the outcomes for participants. In practice, these two roles are not always as clearly demarcated as Scriven defended. However, this distinction between the two purposes of evaluation is still widely used today.

Scriven's strong preference for summative evaluation was developed in reaction to Lee Cronbach, who took the program or product developer as the most important interest, thus focusing on formative evaluation to the exclusion of summative evaluation, in Scriven's view. According to Scriven's view, summative evaluation corrects this bias.

To Scriven (2003, pp. 16-17) evaluation is not just the process of determining facts about things (including their effects). An evaluation must, by definition, lead to a particular type of conclusion - one about merit, worth, or significance - usually expressed in the language of good/bad, better/worse, well/ill, elegantly/poorly etc. This constraint requires that evaluations - in everyday life as well as in scientific practice - involve three components: (i) the empirical study (i.e., determining brute facts about things and their effects and perhaps their causes); (ii) collecting the set of perceived as well as defensible values that are substantially relevant to the results of the empirical study, e.g., via a needs assessment, or a legal opinion; and (iii) integrating the two into a report with an evaluative claim as its conclusion.

For example, in an evaluation of a program aimed to reduce the use of illegal drugs, the empirical study may show (i) that children increased their knowledge of illegal drugs as a result of the drug education part of the program, which is (we document by means of a survey) widely thought to be a good outcome; and (ii) that they consequently increased their level of use of those drugs, widely thought to be a bad outcome. A professional evaluator, according to the definition, should do more than just report those facts. Reporting such facts is a useful research, but is purely an empirical research, partly about effects and partly about opinions. A further effort must be made to analyze critically these results, for example, analyzing their consistency in comparison with other relevant studies, to conclude about the validity of any assumptions and their practical implications in the studied reality and our relevant knowledge's construction. Second, we must synthesize all these results with other relevant studies and values. Only these further steps can get us to an overall evaluative conclusion about the merit of the program (Scriven, 2003, p. 17). Comparisons are a key to making value judgments, and the evaluator has the responsibility for identifying the appropriate alternatives.

Scriven's realist thinking about valuing is reflective of both the ontology and epistemology of the post-positivist paradigm. He does not argue against the idea that we should be seeking an objective truth about the object being evaluated. He, in fact, offers what he believes to be a comparatively unbiased method for obtaining truth about an object's worth, and then advocates for the evaluator to make a value judgment after gathering the most credible evidence. Further, he does not reject

the idea of using experiments to determine causality, but rather argues that there should be more than one method for determining causality (Donaldson, Christie, & Mark, 2008).

Scriven auto-included his evaluation approach in the consumer-oriented evaluation model. Scriven has been committed to consumer-oriented evaluation, using *Consumer Reports* as a model, and to goal-free evaluation, again derived from a consumer rather than a producer orientation. By generalizing from the *Consumer Reports* product evaluation model Scriven has provided many insights into the nature of evaluation. And he admits in his chapter that he too took one type of evaluation to be the whole, particularly when championing the consumer as the only legitimate interest worthy of consideration (House, 2012, p. 10).

Shadish et al. (1991) maintained that Scriven's conception of valuing is dependent on his definition of needs and note that his "conception of needs implies a prescriptive theory of valuing and that he disparages descriptive statements about what people think about the program" (p. 95). Moreover, they maintain that his needs assessment is not independent of the views of the evaluator and that failing to directly reflect the views of stakeholders inhibits the potential use of evaluation findings in policymaking.

Scriven is apparently unconcerned by this, maintaining an idealist position of determining "truth" whose revelation is sufficient. Scriven's unique training in philosophy, mathematics, and mathematical logic provides him with the assurance that he can make sound, unbiased judgments. Scriven is the leading theorist of the valuing perspective and provides the "spiritual guidance" and direction for others depicted on this branch.

Luo (2010) identifies the limitations of Scriven's evaluation model: besides giving evaluators higher authority over different stakeholders in value judgment, Scriven fails to provide a solution to eliminate personal biases of evaluators. Meta-evaluation proposed by Scriven is a good attempt but still it is highly subjective and requires years of experiences and expertise for an evaluator to make a non-biased judgment. For the novice evaluator, the decision of whose needs should be considered and which merit should take higher priority can still very arbitrarily. Besides, a complete goal-free evaluation is also highly unfeasible when an evaluator is hired by his/her clients and has an obligation to answer their specific inquiries.

Patton (2012, p.97) refers to evaluation as "the contextual pragmatics of valuing", requiring flexibility and adaptability on the evaluator's part to ensure the optimal quality of the evaluative reasoning despite the constraints. He states "valuing is fundamentally about reasoning and critical

thinking. Evaluation as a field has become methodologically manic-obsessive. Too many of us, and those who commission us, think that it's all about methods. It's not. It's about reasoning" (p. 105). According to Chelimsky (2012) the role of evaluation in the development and assessment of public policy is well documented. Evaluation provides information about what works, for whom, and why, as well as determining whether the desired outcomes and impacts of public policy are being achieved. For evaluation to be viewed as a credible contributor to public policy, evaluative conclusions need to be robust. Further, because evaluative judgements are "consequential" (Greene J. C., 2011, p. 90) they need to be defensible.

1.2.4. Evaluation and decision (Daniel L. Stufflebeam)

Daniel L. Stufflebeam was another contemporary exponent of evaluation theories, presented his conceptions "strongly supported by democratic principles of equity and social justice" (Freitas, Pontual, Koyama, 2007, p. 122). Following Cronbach's seminal call for evaluations to guide decision making, Stufflebeam also argued that evaluations should help program personnel make decision keyed to meeting recipients' needs. Although advocated an improvement orientation to evaluation, he also stressed that evaluators should both inform decisions and provide information to accountability (Stufflebeam, 1971). He emphasized further that evaluators should interact with and report to the full range of stakeholders who need to make judgements and choices about a program.

To Pacheco (2011) and Prestes (2012) the theoretical axis of Stufflebeam defends the "evaluations centered on the transforming goal" (1971) and the use of methodologies capable of contemplating the approaches of social agendas, of expressing rights defense and generating social changes. Understanding that all processes revolve around decision-making, Stufflebeam defends evaluation processes and procedures as a political and pedagogical act, conducted through four types of decision: context, input, process and product.

Stufflebeam's Context, Input, Process, and Product (CIPP) evaluation model is recommended as a framework to systematically guide the conception, design, implementation, and assessment of service-learning projects, and provide feedback and judgment of the project's effectiveness for continuous improvement (Zhang, et al., 2011).

To House (2012) the CIPP approach has been used in a large number of evaluation studies across all types of subject areas and settings. Stufflebeam has developed it carefully over the years so that the approach has been highly refined and elaborated by constant use. The aim of the evaluation is not to prove, but to improve, he says. CIPP was originally formulated for evaluating educational

programs. The evaluator looked at the context of the program, the inputs to the program, the processes of the program, and the product, or outcomes, both outcomes anticipated and unanticipated. Originally, the information from the evaluation was to be used to inform decision makers about what to do. Since that time CIPP has been used in a variety of settings outside education, and Stufflebeam's concept of decision makers has been broadened to include a range of stakeholders, not just the program administrators (p. 10).

Shaw, Greene and Mark (2005) consider Stufflebeam the "critical evaluation theorist" by proposing indicators that unite the efficiency and technical rationality of evaluation with the concrete reading of the manifold aspects of social reality contained in social programs, one position still defended actually by Prestes (2012) in the face of advances and refinement of the methods and techniques of assessments used by the major international comparative evaluation systems, such as the International Student Assessment Program (PISA), sponsored by the Organization for Economic Co-operation and Development (OECD).

One of those responsible for the development and implementation of the PISA, understands that public policy evaluation specialists from different countries are striving to improve their design, methods, techniques and analysis, which leads them to consider the evaluative processes of the PISA and its results is "highly valid and reliable" (p.31). There are, however, many technical problems, including the development of an evaluation and evaluation framework, capable of looking at the political perspectives of educational evaluation and issues related to equity and the educational success of those assessed (Prestes, 2012).

Afonso (2009) defends that philosophical, theoretical, methodological and political doubts persist as to the relevance of evaluation procedures, to extend the democratic dimension and social justice of evaluation. Thus, the author also concludes the adoption of technical advances in evaluation continues without providing answers to the major social, cultural, and political questions of the processes of participation, negotiation, and justification of explicit models of justice and social equity (p. 25).

1.2.4.1. The CIPP evaluation model's theoretical roots and applications

The need for rigorous and authentic assessment of service learning outcomes has been increasingly recognized, and the many challenges in assessing service-learning have been enumerated. Service-learning is a complex approach to teaching and learning; it needs and

deserves approaches to assessment, evaluation, and reporting that are capable of capturing that complexity (Zhang, et al., 2011, pp. 59-60).

With the passage of legislation in the United States in the 1960s under the Great Society initiative, evaluators realized that their work had the potential to inform policy decisions at the highest level. Several evaluation scholars, such as Daniel Stufflebeam (1980), Carol Weiss (1998) and Michael Patton (2010), raised the consciousness of the evaluation community regarding the use (or nonuse) of their findings.

To Mertens (2015) the pragmatic paradigm aligns closely with the Use Branch in the sense that the focus is on the conduct of evaluations that can provide information in a way that the intended stakeholders can use the results as a basis for informing decision making.

The pragmatic paradigm began in the second half of the 19th century with the contributions of William James, John Dewey, George Herbert Mead and Arthur F. Bentley (Mertens & Wilson, 2012). These scholars rejected the idea that truth could be discovered by scientific methods. Neopragmatism emerged in the 1960s, with scholars such as Abraham Kaplan, Richard Rorty, and Cornel West. "These philosophers have distinguished themselves from the early pragmatists by their emphasis on common sense and practical thinking" (*Ibid.*, p. 89).

Fitzpatrick, Sanders and Worthen (2011) classify evaluation approaches into six groups such as *objectives-oriented*, *management-oriented*, *consumer-oriented*, *expertise-oriented*, *adversary-oriented* and *participant-oriented* approaches. Management-oriented evaluation approach is one of the most important approaches serving managers who are responsible for planning, implementing and evaluating programs. In education, management-oriented evaluation approach provides managers with the information about the implemented program. Hence, the information obtained from evaluation must be the essential part of the decision process and evaluators must contribute to education serving managers, school administrations, teachers and people who need evaluation in education. In this approach, the objectives of the program are not the focal point of evaluation. Stufflebeam has been the pioneer of management-oriented evaluation approach so as to help managers be able to make correct decisions about the program with his CIPP evaluation model (Fitzpatrick, Sanders, & Worthen, 2011).

The CIPP model has been extensively developed and widely implemented in different areas (Hakan & Seval, 2011; Tokmak, Baturay & Fadde, 2013; Tiantong & Tongchin, 2013). Steinert, Cruess, Cruess and Snell (2005) used the CIPP model to evaluate a faculty development program designed to promote the teaching of professionalism to medical students and residents. The authors

conducted all four elements of the CIPP model and also provided preliminary evaluations of their program. Their faculty development initiative was, therefore, evaluated from the initial steps of its planning to the implementation and evaluation of its educational benefits and impacts. Steinert et al. (2005) suggest more rigorous evaluations of these faculty development initiatives should be conducted.

The CIPP evaluation model belongs in the improvement, accountability approach, and is one of the most widely applied evaluation models. The improvement/accountability category is oriented toward determining the merit and worth of the project or entity being evaluated, and encompasses three approaches: decision- and accountability-oriented studies, consumer-oriented studies, and accreditation and certification (Stufflebeam & Shinkfield, 2007).

The CIPP evaluation model has a strong orientation to consumer, service and the principles of a free society. It calls for evaluators and clients to identify and involve rightful beneficiaries, clarify their needs for service, obtain information of use in designing responsive projects and other services, assess and help guide effective implementation of service and ultimately assess the services' merit, worth, significance and probity (Oliveira, 2013). The thrust of CIPP evaluations is to provide sound information that will help service providers regularly assess and improve services and make effective and efficient use of resources, time, and technology in order to serve the well-being and targeted needs of rightful beneficiaries appropriately and equitably (Stufflebeam, 2007, p. 330).

According to Cookingham (2012) Stufflebeam uses the term "evaluation approaches" instead of "models" in a comprehensive analysis, because, in general "evaluation model" refers to an idealized view of conducting evaluation. He does not use the term "model" because some of the approaches he examined are not legitimate (Alkin & King, 2016, p. 20).

1.2.5. Responsive evaluation (Robert E. Stake)

Robert Stake (1974) made a major contribution to curriculum evaluation in his development of the responsive model, because the responsive model is based explicitly on the assumption that the concerns of the stakeholders – those for whom the evaluation is done – should be paramount in determining the evaluation issues.

Responsive evaluation, as a doctrine extending and disciplining common sense, has an intellectual history, some of it passing through Robert Stake's work in the late 60's (Abma & Stake, 2001).

Robert E. Stake is a professor emeritus of education at the University of Illinois, Urbana-Champaign. He completed his PhD in psychometrics at Princeton University in 1958 and then became an assistant professor of educational psychology at the University of Nebraska, Lincoln.

In 1963, he moved to the University of Illinois as an associate professor and associate director of the Illinois State Testing Program. When the University's Center for Instructional Research and Curriculum Evaluation (CIRCE) absorbed the testing center in 1969, Stake served first as the co-director and then the director of CIRCE, a role he continues to this day. He retired from the university in 1998 and received many honors, including the American Evaluation Association's Lazarsfeld Award (1988), the honorary doctorates from the University of Uppsala (1994) and the University of Valladolid (2009) (Miller, King, Mark, & Caracelli, 2016, p. 287).

Stake began as a specialist in psychometrics and instructional research. His focus evolved over the years to a specialization in program evaluation theory and practice, including naturalistic or ethnographic field study, particularly of classrooms. Among his many contributions are the *Countenance Model of Educational Evaluation* (1967), *Responsive Evaluation* (1974), and *The Art of Case Study Research* (1995).

Stake's writings and thinking in program evaluation have influenced the work of numerous scholars in the field. He is well known as a creator of the metaphor that captures the distinction between formative and summative evaluation: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." (Miller, King, Mark, & Caracelli, 2016, pp. 287-8).

Stake describes his evaluation experiences and interactions with key others who developed the field, beginning in the 1960s at the University of Illinois, including Lee Cronbach, Michael Scriven and Ralph Tyler. He describes the creation of the *countenance model*. His current desire is to make evaluators aware of the continuing challenges the field has faced since its inception.

To emphasize evaluation issues that are important for each particular program, Stake recommends the responsive evaluation approach. It is an approach that trades off some measurement precision in order to increase the usefulness of the findings to persons in and around the program. An educational evaluation is a responsive evaluation if it orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspectives present are referred to in reporting the success and failure of the program. (p. 14).

Early on Stake pointed out that responsive evaluation assumes value pluralism and that the evaluator should not press for consensus. He put it as follows:

"A work of art has no single true value. A program has no single true value. Yet both have value. The value of an art-in-education program will be different for different people, for different purposes. . . . Whatever consensus in values there is . . . should be discovered. The evaluator should not create a consensus that does not exist" (1975, pp. 25-6).

These are because Stake believed there are different constituencies and different stakeholders who have different expectations and different values. The evaluator should understand those things and be in a good position to illustrate them, to represent them to readers and outsiders as part of the evaluation task (Abma & Stake, 2001).

According to Stufflebeam and Shinkfield (2007), Stake recommends an interactive and recursive evaluation process that embodies nine steps or duties that evaluator should implement: (i) meets with clients, staff, and audiences to gain a sense of their perspectives on and intentions regarding the evaluation; (ii) draws on such discussions and the analysis of any documents to determine the scope of the evaluation project; (iii) observes the program closely to get a sense of its operation and to note any unintended deviations from announced intents; (iv) discovers the stated and real purposes of the project and the concerns that various audiences have about it and the evaluation; (v) identifies the issues and problems with which the evaluation should be concerned. For each issue and problem, the evaluator develops an evaluation design, specifying the kinds of data needed; (vi) selects the means needed to acquire the data desired. Most often, the means will be human observers or judges; (vii) implements the data-collection procedures; (viii) organizes the information into themes and prepares "portrayals" that communicate in natural ways the thematic reports (the portrayals may involve videotapes, artifacts, case studies, or other "faithful representations"); (ix) by again being sensitive to the concerns of the stakeholders, the evaluator decides which audiences require which reports and chooses formats most appropriate for given audiences (p. 381).

The method of responsive evaluation is characterized by being pluralistic, flexible, interactive, holistic, subjective and service-oriented. One of Stake's recommendations about evaluation is that evaluators need to work continuously with the various audiences and meet their various evaluation needs.

To Alkin and King (2016) the chief advantage of the responsive model is its sensitivity to clients. By identifying their concerns and being sensitive to their values, by involving them closely throughout the evaluation, and by adapting the form of reports to meet their needs, the model, if effectively used, should result in evaluations of high utility to clients.

The responsive model also has the virtue of flexibility:

“The evaluator is able to choose from a variety of methodologies once client concerns have been identified. Its chief weakness would seem to be its susceptibility to manipulation by clients, who in expressing their concerns might attempt to draw attention away from weaknesses they did not want exposed” (Afonso, 2009, p. 20).

Stufflebeam and Shinkfield (2007) synthesized the main aspects of the Skate’s evaluation design:

- Evaluations should help audiences to observe and improve what they are doing (hence our denomination of client-centered assessments);
- Evaluators should write programs regarding background and operations as well as outcomes;
- Side effects and accidental achievements must be as well studied as the results sought;
- Evaluators should avoid presenting summarized final conclusions, but instead should collect, analyze and reflect the judgments of a wide range of people interested in the subject of the evaluation;
- Regularized experiments and tests are often inadequate or insufficient to meet the purposes of an evaluation, and should often be replaced or supplemented with a variety of methods, including "soft" and subjective.

For the development of these aspects in an evaluation process, the method of responsive evaluation consists of the following steps: (i) description and judgment of a program; (ii) different data sources; (iii) analysis of congruences and contingencies; (iv) identification of relevant and opposing standards; (v) the multiple uses of the evaluations.

Stake suggests these steps with the aim of overcoming an evaluation view that is central to students' classifications and the validity of their results. An evaluation process that considers individual differences as a consequence of the antecedent contingencies of each one, so the method of responsive evaluation is based on the idea that the evaluators and educators must fully describe “the antecedents assumptions and realities, didactic operations and results, and to examine their congruence’s and contingencies” (Stufflebeam & Shinkfield, 2007).

Some advantages of responsive evaluation are the intention to help the evaluated ones to identify the advances and the retreats and develop a process of communication to the audience during the educational process. This evaluation model has a methodology that presents the purposes and procedures generally from the beginning and develops throughout the studies; the preferred techniques are the concrete case studies, the observations, the examinations of opposing programs, the expressive reports, among others (Stufflebeam & Shinkfield, 2007).

In 1978 Stake expressed his preference for case study methods, arguing that knowledge from case studies is concrete and contextual, and open for different interpretations. Case studies are therefore fitted to the natural ways in which people assimilate information and come to an understanding:

Case studies are the preferred method of research because they may be epistemologically in harmony with the reader's experience and thus to that person a natural basis for generalization (p.5) (...) Most case studies feature: descriptions that are complex, holistic and involving a myriad of not highly isolated variables; data that are likely to be gathered at least partly by personality observation; and a writing style that is informal, perhaps narrative, possibly with verbatim quotation, illustration, and even allusion and metaphor. Comparisons are implicit rather than explicit. Themes and hypotheses may be important, but they remain subordinate to the understanding of the case (p.7).

Stake distinguished two types of case studies he had in mind, the intrinsic case studies and the instrumental case studies (Stake, 1994). The primary purpose of an intrinsic case study is to understand a particular case rather than to investigate a certain issue (Stake & Mabry, 1995). The information about context (cultural, organizational, curricular, legal, professional, policy, and collegial) and other details are important to gain a deep understanding of a particular case.

Abma and Stake (2001) point that Stake has always wanted to make a contribution to educational practices such as discovery learning by giving a maximum amount of vicarious experience. The responsive evaluator should present not only the findings but also personal experiences. Vicarious experiences can function as a substitute for those who are missing direct experiences. This does not imply a rejection of formal scientific knowledge, but Stake points out that we often rely too exclusively on that kind of knowledge.

Stake's responsive model (1974) also dealt with context, but his focus was more on the program deliverers, in his case, teachers. The responsive model was developed in reaction to the dominant pre-ordinate approach of the time, large studies designed to test big theories with fixed, quantitative methods. Greene and Abma (2001, p. 1) note that Stake reframed evaluation "from the application of sophisticated analytic techniques that address distant policymakers' questions on program benefits and effectiveness 'on the average,' to an engagement with on-site practitioners about the quality and meanings of their practice".

Like Stufflebeam, Stakes wanted to shift evaluators from a focus entirely on objectives, but his solution was to become closer to the program, observing it, talking with program deliverers and drawing out their thoughts and perceptions, and developing case studies or in-depth descriptions of the program. Stake was moving evaluators toward seeing the program and, in so doing, to understand the program's context (Fitzpatrick, 2012).

1.3. The curriculum evaluation

In recent years the interest in curriculum evaluation in particular has seemed to increase markedly. The public's insistence on educational accountability, the experts' demands for educational reform, and the educators' concomitant need for evidence of results have all contributed to the current interest in theories and methods of curriculum evaluation. Unfortunately, much of this interest seems to have resulted in an obsession with test results.

Otherwise, evaluation has an extremely extensive territory, since it includes the substantial portion of everyday discourse devoted to proposing, attacking, and defending evaluative claims about food products, football teams, human behavior, global warming, and almost everything else. The domain of professional evaluation is still very extensive: we here distinguish seven standard sub-divisions of it, and four other specialized domains which are less commonly categorized or recognized as part of evaluation's domain, although substantially devoted to it (Scriven, 2007, p. 2).

Kliebard (1975) reminds us of the genesis of the curriculum movement in the United States, and identifies two key figures in the early part of the last century who represents the enthusiasm for the application of the scientific curriculum-making and the method to the study and implementation of the curriculum.

Werrett Charters (1923) and Franklin Bobbitt (1924) in their different ways argued for precision, objectivity, prediction and the use of the scientific method to establish once and for all what should be taught in schools and indeed how educational knowledge should be structured. Bobbitt's two major works were *The Curriculum* (1918) and *How to Make a Curriculum* (1924), and in 1913, he published a long article entitled *Some General Principles of Management Applied to the Problems of City-school Systems* (1913). Charters' two major works were *Methods of Teaching: Developed from a Functional Standpoint* (1909), and *Curriculum Construction* (1923), both of which reflected then currently fashionable ideas of structural-functionalism (Scott, 2016, p. 13).

Bobbitt defends the arguments for behavioral objectives and he is credited with developing a notion of objective analysis whereby designated skills are broken down into their constituent elements. These skills were derived from the activities of experts in a variety of fields essential to the well-being of society, and he claimed that curricular aims and objectives could be derived from an objective examination of these activities. Furthermore, these skills and their component sub-skills could be expressed as specific teaching objectives which could be so arranged that the curriculum could be designed around them. His work was behaviorist in that he understood learning as the acquiring of these skills and the evaluation of sets of behaviors so as to determine whether these

skills had been successfully acquired by the learner. It is easy to see here the origin of the behavioral objectives movement which influenced curriculum making in the 1970s and 1980s and which continues to shape global, national and local curricula round the world (Scott, 2016, p. 13). Ralph Tyler (1950) advocated a means-end approach to the development of the curriculum. He believed that educational aims could only be articulated in terms of objectives and that these preceded learning experiences and the evaluation of what is learnt. Curriculum-making was understood as a linear process which starts with the development of clear objectives or goals, proceeds through to the selection of content which is specified in behavioral terms – its acquisition must be an observable or testable process – and finishes with the evaluation of that process to see if those objectives have been met. However, he did not believe that objectives could be specified in precise behavioral terms, and he believed that they should be kept at a fairly general level. His work has influenced current models of curriculum-development, though his objectives approach has in turn been heavily criticized for its limited understanding of the enacted curriculum (Scott, 2016, p. 16).

Stake (1971) had argued effectively that standardized tests often are poor approximations of what teachers actually teach. Moreover, as has been patently clear in evaluations of programs for both disadvantaged and gifted students, norm-referenced tests often do not measure achievements well for low and high scoring students. Unfortunately, program evaluators often have made uncritical use of standardized test results to judge a program's outcomes, just because the results were conveniently available and had face validity for the public. Often, the contents of such tests do not match the program's objectives.

Other theorists such as Popham (2001) were less discriminating about the use of behavioural objectives and were enthusiastic advocates of a scientific view of curriculum making. Such a position was underpinned by a view of knowledge that colored their perception of the curriculum. To Scott (2016) in his *An Evaluation Guidebook: A Set of Practical Guidelines for the Educational Evaluator* (1972), Popham argues strongly for a behavioural objectives model of teaching and learning, an approach that has had a considerable influence on the field of curriculum, culminating in the development of a national curriculum in the United Kingdom in the 1990s and similar policy initiatives round the world. Though educational theorists such as Popham embraced a technicity model of curriculum inherent in the specification of behavioural objectives, other curriculum theorists associated with this approach argued for weaker versions.

Ralph Tyler (1950), for example, suggested that specifying objectives was the only logical way of determining learning experiences. However, he did not subscribe to the view that they could be broken down into thousands of detailed educational sub-purposes, because he felt that this would unnecessarily restrict the teachers, and overwhelm their capacity to use the curriculum objectives. The rationale for developing this type of curriculum model was to provide clarity of purpose where none had previously existed. Popham defends the major advantages of such objectives is that they promote increased clarity regarding educational intents, whereas vague and unmeasurable objectives yield considerable ambiguity and, as a consequence, the possibility of many interpretations not only of what the objective means but, perhaps more importantly, whether it has been accomplished (Scott, 2016).

To Popham (2001) behavioral objectives have a number of features. First, they have to be unambiguously stated so that they provide explicit descriptions of the behaviors that should occur after instruction has taken place. These behaviors furthermore have to be stated so that any group of reasonable observers would agree that the individual has shown herself capable of performing them. Second, those behaviors have to refer to the learner and not the teacher.

The teacher may devise systems of instruction that have merit; however, if they do not lead to the desired and pre-specified behaviors in learners, then they cannot be considered useful. Third, those behaviors should be expressed so that they can be measured; clarity is therefore reduced to measurability. Popham's third proposition in relation to behavioral objectives is that "the educational evaluator must identify criteria of adequacy when using instructional objectives which require constructed responses from learners".

Popham (2001) provides no guidance for determining whether objectives should be specific or general, but suggests only that teachers may prefer to work at a level of generality and as a consequence this should not be ruled out. Popham makes a further suggestion to the effect that behavioural objectives should take account of proficiency levels of performance, and that they should refer to either the individual learner or the class as a whole. Objectives therefore can be formulated so that they are only partially achieved, but this does not rule out their usefulness as curriculum tools.

Drawing on Bloom's Taxonomy of Educational Objectives with regards to cognition (Bloom & Krathwohl, 1956) and Krathwohl's Taxonomy in relation to the affective domain (Krathwohl *et al.* 1964), Popham (2001) argues that curriculum-makers should use these to develop their lists of behavioural objectives. Three types of objectives are identified: the cognitive, the affective and the

psychomotor, and these in turn are broken down into six cognitive domains (knowledge, comprehension, application, analysis, synthesis and evaluation), five affective domains (receiving, responding, valuing, organizing and characterizing by a value or value complex), and five psychomotor domains (perception, set, guided response, mechanism and complex overt response) (Scott, 2016, p. 16).

After defining the evaluation concept and analyze the different conceptual models of evaluation, emerges the following question: *How can the merit and worth of such aspects of curriculum be determined?* Evaluation specialists have proposed an array of models, an examination of which can provide useful background for the process presented in this work.

1.3.1. Program evaluation

Although its historical roots extend to the 17th century, widespread systematic evaluation research is a relatively modern 20th-century development. The application of social research methods to program evaluation coincides with the growth and refinement of the research methods themselves as well as with ideological, political, and demographic changes.

The first evaluative initiatives of social policies inspired by scientific procedures emerged in the United States in the 1950s and were intended to support the welfare policies implemented by different administrations. In those postwar years, the USA administration, willing to implement capable policies of solving its social problems, found difficulties in determining the effectiveness of its actions, and resorted to information based on scientific methods to remedy these difficulties. This attitude helped to legitimize their actions, even requiring more empirical analysis of the socioeconomic connections of the phenomena evaluated.

Starting in the 1950s and becoming established in the middle of the 1960s, evaluation was part of a much larger stream of ideas to make government more scientific. The public sector would perform much better with a proper dose of trustworthy scientific findings about the real results of adopted policies and programs. Given externally set goals, professional academic researchers should be commissioned to scientifically evaluate appropriate means to reach these goals through controlled two-group experimentation. Evaluation was based on means–ends rationality. While goals were considered subjective, the means to reach the goals could be ascertained in an objective and scientific way.

Already in the early 1970s faith in methods-driven scientification of government started to languish. Mistrust of experimental evaluation gained momentum. Demands were voiced for more participation by diverse groups and more dialogue and communication in evaluations. Supporters

of this dialogue-oriented wave often characterized it as democratic evaluation. Representative democracy might be supplemented by evaluative arenas where users, citizens and other stakeholding audiences deliberated effects and implementation of public interventions.

Although considerably older than evaluation, the stakeholder-dialogue idea was incorporated into evaluation discourse and practice at about this time. And it has stayed there ever since. This second wave was also driven by an ideological current from the left, but sprinkled with some green, environmentalist tenets.

Around 1980, the neo-liberal wave came rolling in, this time from the right of center on the political scale. The neo-liberal wave was based on a mistrust of central planning but saw the remedy not in dialogue and participation but in more market orientation. Deregulation, privatization, efficiency and customer orientation became new key words. Evaluation came to be included in a neo-liberal, market-oriented train of thought called.

While gaining in strength, the fourth evaluation wave is not yet as strong as the scientific wave of the 1960s and the neo-liberal wave that grew in popularity from 1980 onwards. Characteristic of this evidence wave is an effort to make government more scientific and based on real empirical evidence. It is concerned with what works. This can be interpreted as a renaissance of science and randomized experimentation. It is basically driven from the right-of-center end of the political spectrum (Vedung, 2010, pp. 275-276).

Most developments in program evaluation took place in the United States and were "exported" to other parts of the world, sometimes only ten or twenty years later. In Europe, for instance, the major concern was – and in some countries still is – testing and student assessment, although tests and other achievement measures have begun to be used for other purposes. Gradually, tests came to be used as outcome measures for other evaluation objects, such as programs, schools, and education systems, sometimes alongside other information regarding the objects' goals and processes (Alkin & King, 2016).

The models that emerged were developed in response to the need to provide accountability for large USA government program expenditures in health, education, and welfare during this period. Evaluation is especially important in public sector organizations because they are required, for political accountability or legislative reasons, to demonstrate the benefits of their actions to the public (Vedung, 2010). Judgments can also be made regarding the cost effectiveness of programs and plans.

Evaluating means judging the value of an object, and evaluation in the sense of a particular type of disciplined inquiry emphasizes that this “judging” and “valuing” is based on some kind of systematic information gathering approach. In the case where this systematic information gathering is formalized according to the criteria for social scientific inquiry the term evaluation research is appropriate.

Following this reasoning, Weiss (1972) showed how a model could be developed and tested to explain how a chain of events in a teacher home visit program, for example, could lead to the ultimate objective of improving children’s reading achievement. This early work led to the development of an approach known today as theory-based evaluation, theory-driven evaluation or program theory evaluation (PTE).

PTE consists of two basic elements: an explicit theory or model of how the program causes the intended or observed outcomes and an actual evaluation that is at least guided by the model (Rogers et al., 2000). The program model, often called a logic model, is typically developed by the evaluator in collaboration with the program developers, either before the evaluation takes place or afterwards. Evaluators then collect evidence to test the validity of the model. PTE does not suggest a methodology for testing the model, although it is often associated with qualitative methodology. Cook (2000) argues that program theory evaluators who use qualitative methods cannot establish that the observed program outcomes were caused by the program itself, as causality can only be established through experimental design. Generally speaking, the contribution of PTE is that it forces evaluators to move beyond treating the program as a black box and leads them to examining why observed changes arising from a program occurred (Owston, 2008, p. 607).

The connection between the (researcher’s) practice of evaluation and evaluative knowledge represents an innovative perspective conceiving of evaluation as a practice-based approach. In this way, Hammersley (2013) conceives the evaluation as contributing to organizational transformation by mediating the inevitable political, cultural and relational implications that are characteristic of all evaluations. This means, on the one hand, taking into account the embeddedness of evaluation in organizations whose dynamics are shaped by stakeholders with their own interests. On the other hand, this conception of evaluation involves questioning the idea that evaluation research should directly serve the decisions of policy-making and practice (Hammersley, 2013).

In this regard, as Contandriopoulos and Brousselle (2012) suggest, evaluation results need to be embedded in a real-life context in order to be relevant. In particular, they point out that the use and the usability of the evaluation depend on both the distance between actors’ different readings of

problems, and the balance that they attribute to the cost/benefit ratio of investing the energy and resources that evaluation requires. The process and the results of evaluation practice, therefore, depend on how different and often divergent opinions, criteria and interests are combined in a context. It is this that determines the specific conditions of knowledge-use and exchange, including the use of knowledge produced through evaluation.

To Carol Weiss (1998, p. 4) program evaluation can be broadly defined as the “systematic assessment of the operations and/or outcomes of a program, compared to a set of explicit or implicit stands, as a means of contributing to the improvement of the program.” There are several key elements in this definition.

First, program evaluation requires a systematic assessment that is governed by acceptable social science research methods. Evaluation is considered an empirically oriented discipline that generates information about programs in order to improve the program or guide future decisions (Pal, 2014).

Second, there is an emphasis on both program operation and outcomes. That is, evaluation is not only concerned with program effectiveness but also the process of delivering programs such as the organizational methods used to deliver the program, program inputs (e.g., resources), program outputs (e.g., tangible measures of a program), and cost effectiveness.

Finally, program evaluation is used to help make programs work both efficiently and effectively (Weiss 1998), and as a means to ensure accountability and quality assurance (Cousins, Goh, Elliott, Aubry & Gilbert, 2014; Pal, 2014).

There are generally two main types of program evaluation – formative and summative. Evaluations that focus on improving the performance of a program are known as formative. A formative evaluation provides feedback in order to improve the outcomes of programs or to increase its efficiency. Formative evaluations generate information to influence immediate decisions about a program, such as improving component parts and processes (Shadish, Cook & Leviton, 1991).

On the other hand, evaluations that focus on outcomes are known as summative and occur once a program is complete or substantially complete. Summative evaluations provide information to decision makers regarding whether a program has achieved its stated goals or is worthwhile to continue (Shadish, Cook & Leviton, 1991).

The ways in which evaluation is designed and takes place can be categorized in four models: post positivism, pragmatism, interpretivist and critical normative science, inspired in Greene (1994) and synthesized in the next table.

Table 1. Epistemological models of evaluation

Philosophical framework	Key values promoted	Key audience	Preferred methods
Postpositivism	Effectiveness, efficiency, causal knowledge	Decision makers	Quantitative: experiments and quasi experiments, cost-benefit analysis
Pragmatism	Management, practicality, quality control	Program managers, administrators, and other decision makers	Surveys, questionnaires, interviews, observations
Interpretivism	Pluralism, understanding, diversity	Program staff, program beneficiaries	Qualitative: case studies, interviews, document reviews
Critical, normative science	Emancipation, empowerment, social change	Program beneficiaries and other “powerless” groups	Participatory: stakeholder participation in qualitative and quantitative designs

Source: Adopted from Greene (1994).

The first, and historically dominant, evaluation model is science-driven and highly technical with a strong emphasis on quantitative methods. Scientific research methods and techniques such as systematic randomized research designs and experiments are considered integral to the evaluation process (Alkin, 2013). This approach to evaluation focuses on measuring effectiveness and efficiencies (Greene, 1994). For example, a common measure for evaluation might include the extent to which programs promote or impede the realization of goals or objectives. Such an approach requires that evaluators clearly identify goals and objectives, and be able to measure them through quantitative processes (Guyadeen & Seasons, 2016).

The second evaluation model emerged as a response to the overreliance on scientific research and the difficulties associated with identifying objectives to be evaluated (Alkin, 2013). This model adopts a pragmatic approach to evaluation and argues that evaluation methods should be matched with the program being evaluated (Greene, 1994). The CIPP model (Context, Input, Process, and Product evaluations) is an example of a pragmatic evaluation model that was developed to engage decision makers in the evaluation process. The intent of the CIPP model is to provide support for efficient and effective program management by providing continuous information to decision makers (Alkin, 2013; Greene 1994).

The third evaluation model is grounded in the interpretivist philosophical framework. This model places a strong emphasis on pluralism and in understanding the diverse stakeholders involved in an evaluation (Greene, 1994). Qualitative methods are often used to enhance the understanding of programs from the perspectives of the stakeholders directly involved in the program (Greene, 1994). Stakeholders are considered critical as they are seen as having a direct stake (e.g., money and vested interest) in the evaluation. Guba and Lincoln’s (1989) fourth generation evaluations is

an example of the interpretivist framework whereby the claims, concerns, and issues of stakeholders involved in a program are considered central to the evaluation (cit. in Guyadeen & Seasons, 2016, pp. 1-13).

The fourth evaluation model follows a normative approach (Greene, 1994) and emphasizes collaboration and negotiation among stakeholders (e.g., decision makers, program recipients, and evaluators) during the evaluation process. In this phase, evaluators attempt to acknowledge and recognize the multiple realities and stakeholder perspectives associated with the evaluation process (Alkin, 2013). Participatory evaluation (Cousins, 2004), empowerment evaluation (Fetterman, 2004), collaborative evaluation (Rodriguez-Campos, 2012), and more recently, developmental evaluation (Patton, 2011) closely aligns with this category.

According to Guyadeen and Seasons (2016) these models demonstrate the established nature of program evaluation when compared with evaluation in planning. The authors' argument is that there is value to further strengthening the linkages to program evaluation in an effort to advance evaluation in planning. It is important to note that there has been, and continues to be, much debate regarding the appropriateness of these models.

The Tyler's objectives-centered model, Stufflebeam's context, input, process, product model, Scriven's goal-free model, Stake's responsive model, and Eisner's connoisseurship model) presented above give some support to Bradley's effectiveness model.

1.3.2. Bradley's effectiveness model

How can a developed curriculum be assessed and evaluated for effectiveness? Bradley's (1985) book *Curriculum Leadership and Development Handbook* provides 10 key indicators that can be used to measure the effectiveness of a developed curriculum.

The table 2. is designed to help one identify perceptions regarding the 10 indicators to appraise curriculum effectiveness in a school building or district. To assess how a school or district meets each of the indicators, evaluator responds with a Yes or No in the column provided.

Table 2. Bradley's effectiveness model for curriculum development indicators

Indicator	Description	Yes or No
Vertical curriculum continuity	The course of study reflects a K–12 format that enables teachers to have quick and constant access to what is being taught in the grade levels below and above them. Also, upward spiraling prevents undue or useless curricular repetition.	
Horizontal curriculum continuity	The course of study developed provides content and objectives that are common to all classrooms of the same grade level. Also, daily lesson plans reflect a commonality for the same grade level.	
Instruction based on curriculum	Lesson plans are derived from the course of study, and curriculum materials used are correlated with the content, objectives, and authentic tasks developed.	
Curriculum priority	Philosophical and financial commitments are evident. Clerical assistance is provided and reasonable stipends are paid to teachers for work during the summer months. In addition, curriculum topics appear on school board agendas, administrative meeting agendas, and building-staff meeting agendas.	
Broad involvement	Buildings in the district have teacher representatives on the curricular committees; elementary, middle level or junior high, and high school principals (or designees) are represented; and school board members are apprised of and approve the course of study.	
Long-range planning	Each program in the district is included in the 5-year sequence and review cycle. Also, a philosophy of education and theory of curriculum permeate the entire school district.	
Decision making clarity	Controversies that occur during the development of a program center on the nature of the decision, and not on who makes the decision.	
Positive human relations	Also, the initial thoughts about the curriculum come from teachers, principals, and the curriculum leader. All participating members are willing to risk disagreeing with anyone else; however, communication lines are not allowed to break down.	
Theory-intopractice approach	The district philosophy, vision, mission, exit (graduation) outcomes, program philosophy, rationale statement, program goals, program objectives, learning outcomes, and authentic tasks are consistent and recognizable.	
Planned change	Tangible evidence shows that the internal and external publics accept the developed program course of study for the school district. The process of developing a course of study for each program or discipline in a school district is no longer one of determining how to do it, but one of determining how to do it better.	
If any of the 10 indicators are identified with a No (negative), consideration should be given to make it a Yes (positive) indicator.		

Source: The 10 indicators of effective curriculum development were adapted from Curriculum Leadership and Development Handbook (pp. 141–146), by L. H. Bradley, 1985, Englewood Cliffs, NJ: Prentice Hall.

The indicators for effective curriculum development represent working characteristics that any complex organization must have in order to be responsive and responsible to its clients. Further, the measurement can be oriented to meet the needs of any school district – from large to small – and it can focus on a specific evaluation of a district’s curriculum area, such as reading, language arts, math, or any content area designated.

1.3.3. Eisner's connoisseurship model

Elliot Eisner (1979) drew from his background in aesthetics and art education in developing his “connoisseurship” model, an approach to evaluation that emphasizes qualitative appreciation.

The Eisner model is built on two closely related constructs: connoisseurship and criticism. Connoisseurship, in Eisner's terms, is the art of appreciation—recognizing and appreciating through perceptual memory, drawing from experience to appreciate what is significant. It is the ability both to perceive the particulars of educational life and to understand how those particulars form part of a classroom structure. Criticism, to Eisner, is the art of disclosing qualities of an entity that connoisseurship perceives. In such a disclosure, the educational critic is more likely to use what Eisner calls “nondiscursive”—a language that is metaphorical, connotative, and symbolic. It uses linguistic forms to present, rather than represent, conception or feeling.

Educational criticism, in Eisner's formulation, has three aspects. The descriptive aspect is an attempt to characterize and portray the relevant qualities of educational life—the rules, the regularities, the underlying architecture. The interpretive aspect uses ideas from the social sciences to explore meanings and develop alternative explanations—to explicate social phenomena. The evaluative aspect makes judgments to improve the educational processes and provides grounds for the value choices made so that others might better disagree.

The chief contribution of the Eisner model is that it breaks sharply with the traditional scientific models and offers a radically different view of what evaluation might be. In doing so, it broadens the evaluator's perspective and enriches his or her repertoire by drawing from a rich tradition of artistic criticism. Its critics have faulted it for its lack of methodological rigor, although Eisner has attempted to refute such charges. Critics have also argued that use of the model requires a great deal of expertise, noting the seeming elitism implied in the term connoisseurship.

1.3.4. Michael Patton's utilization-focused evaluation

Michael Patton's original utilization-focused model (Patton, 1976) followed Wholey's focus on managers and encouraged use by identifying one key manager who has the position to do something with the evaluation and the disposition, or interest, to use it. Patton's model moved evaluators to begin thinking more about the organizational contexts in which their evaluations were used and to explore those contexts in thinking about utility. He made different and more specific suggestions for evaluators to use in working with managers (Fitzpatrick, 2012).

The innovative practices to which many educators aspire can accommodate and build on more traditional mandates (Ferrero, 2006). Although the models above seem sharply distinct from one another, some evidence of congruence exists in current theories of evaluation.

This congruence is quite evident in the ASCD monograph *Applied Strategies for Curriculum Evaluation* (Brandt, 1981) in which seven experts in evaluation were asked to explain how their “evaluation model” would be used in evaluating a secondary humanities course.

While the models proposed by the experts (Tyler, Cronbach, Scriven, Stufflebeam, Stake, Eisner and Worthen) differed in many of their details, several common emphases emerged in the approaches, such as: study the context, determine client concerns, use qualitative methods, assess opportunity cost (what other opportunities the student is missing by taking this course), be sensitive to unintended effects, and develop different reports for different audiences.

By using these common emphases, along with insights generated from analyzing other models, it is possible to develop a list of criteria that can be used in both assessing and developing evaluation models. Such a list is shown in next table.

Table 3. Criteria for a curriculum evaluation model

An effective curriculum evaluation model does the following:
1. Can be implemented without making inordinate demands upon district resources
2. Can be applied to all levels of curriculum—programs of study, fields of study, courses of study
3. Makes provisions for assessing all significant aspects of curriculum—the written, the taught, the supported, the tested, and the learned curricula
4. Makes useful distinctions between merit (intrinsic value) and worth (value for a given context)
5. Is responsive to the special concerns of district stakeholders and is able to provide them with the data they need for decision making
6. Is goal oriented, emphasizing objectives and outcomes
7. Is sensitive to and makes appropriate provisions for assessing unintended effects
8. Pays due attention to and makes provisions for assessing formative aspects of evaluation
9. Is sensitive to and makes provisions for assessing the special context for the curriculum
10. Is sensitive to and makes provisions for assessing the aesthetic or qualitative aspects of the curriculum
11. Makes provisions for assessing opportunity cost—the opportunities lost by those studying this curriculum
12. Uses both quantitative and qualitative methods for gathering and analyzing data
13. Presents findings in reports responsive to the special needs of several audiences.

Source: CHAPTER 12 Curriculum Evaluation, p. 365.

Such an eclectic process has been used successfully in evaluating a field of study; this same process also can be used to evaluate a course of study with the scope of the evaluation reduced. The key issues are the establishment of more effective and accurate ways with which we can measure the real benefits of education and measure the true associated costs in money and time spent learning about curriculum and instruction.

1.4. Teacher performance evaluation

We find that while the idea that teachers' performance, like that of other professionals, should be subject to systematic evaluations, the nature and complexity of the teaching function itself does not allow us to clearly determine the purpose of the evaluation.

In this regard, Simões (2000, p. 12) recalls that if the assessment "presupposes the collection of information to judge the merit and / or value of the teacher", the existence of different conceptions of teaching work, or different conceptions of the which is to be a teacher, "imply different ways of collecting information and different ways of making value judgments." This is one of the main difficulties in assessing teacher performance because it conditions the collection and use of information to judge the merit and value of each teacher.

Therefore, finding a unanimous definition is as difficult as prescribing the operationalization of the concept itself, since ambiguity and polysemy dominate. Hence there is a need to clarify what we refer to as the evaluation of teacher performance.

Hadi (1994) considers that the evaluation of teacher performance results from the comparison between the expected (referential) performance and the actual performance (referred to) through the "reading of an observable reality", based on information from different moments of the teacher's activity. It is, according to the author, to attribute a value or, in a broader sense, a meaning "to a real situation in the light of a desired situation, when confronting the field of concrete reality with that of expectations" (pp.32-33).

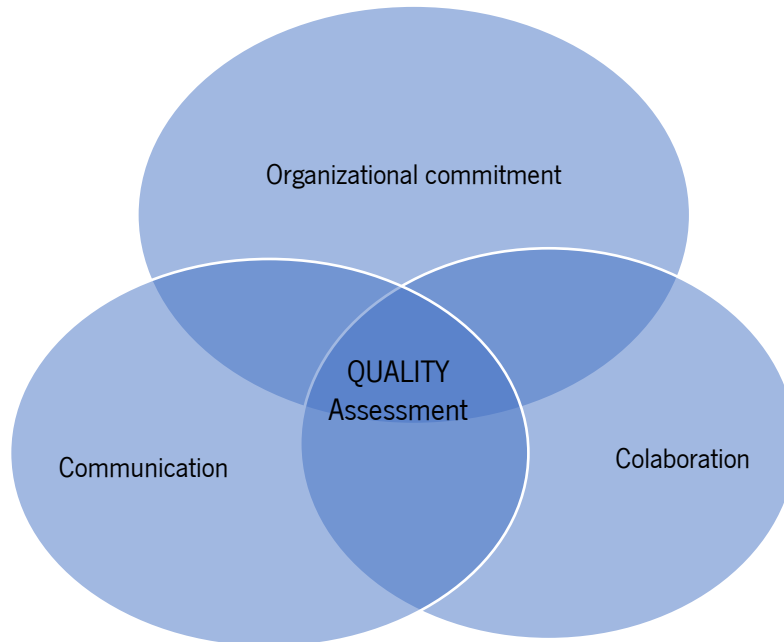
According to Rodrigues and Peralta (2008), the assessment of teacher performance "involves the observation, description, analysis and interpretation of professional activity to make decisions regarding the teacher" (pp. 11-12) which decisions can be educational, administrative or salary.

From the perspective of the authors, the evaluation tries to infer if the teachers have acquired and developed the competences foreseen in the evaluation references, in a process in which, in addition to the comparison with external data, the individual performances are also evaluated on the basis of intrinsic data. The decisions it has made about its development as a professional, the objectives it has defined, the projects in which it has decided to commit itself, the innovations it intends to introduce in its practices, among others (Rodrigues & Peralta, 2008).

Graça et al. (2011, p. 22) pointed out that "the implementation of an evaluation process of teacher performance that is consistent and aims to achieve objectives should be resulted from a process of sharing, reflection and consultation, based on dialogue and involvement of all, from the

management of the school, to the evaluator and the evaluator". The authors synthesized these ideas in diagram that figures next.

Figure 1. Relationship in 3 C's to a quality of the evaluation



Source: Graça, et al. 2011, p.22.

The success of any performance evaluation system is based on the interdependence of a set of approaches and procedures, which are: communication between the participants; peer collaboration, in view of the common goal of achieving the quality of teacher education and the quality of student learning; and organizational commitment to the school and the education system as a whole. From the interconnection and interdependence of these approaches, quality assessment is achieved (Stronge & Tucker, 2003).

In *Linking Teacher Evaluation and Student Learning*, researchers Tucker and Stronge (2005) show that including measures of student achievement in teacher evaluations can help schools focus their efforts to meet higher standards. For any school that is working to meet higher standards, linking teacher evaluation to measures of student learning is a powerful way to refocus professional development and improve student achievement.

Performance appraisal, teacher standards and professional learning are difficult to isolate, these influences need to be considered as “part of a wider, more complex web of factors that impact in significant ways upon the work of teachers, and the learning that happens in schools” (Tuinamuana, 2011, p. 79).

Teacher's evaluation has a positive influence on teacher's professional development (Stronge & Tucker, Handbook on teacher evaluation: Assessing and improving performance, 2003) and is a vehicle to promote education quality (Danielson & McGreal, 2000).

The teacher performance evaluation is fundamentally based on two perspectives: professional development and control, or in the terminology of Day (2001) "one oriented to the process, another to the product" (p.150).

McLaughlin (1997) considers that teacher assessment does not produce more responsibility or improvement of practice in schools, which is why De Ketele (2010) defends we should talk more about professional development and teaching evaluation than evaluation of teachers, assuming the assessment two positions: the control and the recognition.

The recognition posture encourages professional development, since the evaluation should focus on the effects of the practices and not on the practices themselves, so that from the effects can be traced back to the impact, a fundamental process for greater awareness of the actions of the teacher and of all the other actors in the teaching-learning process. From the point of view of professional development, the main concern of the evaluation is to gather information to help teachers who are at least competent (Pacheco & Flores, 1999, p. 167).

Performance evaluation has played a key role in providing background information to support decision-making in human resource management, particularly in terms of career selection and promotion (such as determining whether salary increases and Provide feedback among supervisors, or evaluators, and employees) (Coutts & Schneider, 2004).

The evaluation of performance then integrated into several dimensions and became more extensive. On the one hand, in addition to the management function, namely the monitoring function, the performance evaluation started to intervene in making decisions, in control and in accountability (Pacheco & Flores, 1999, 167).

Studies on the appraisal process predominantly identify three models of appraisal, namely, the accountability model, the professional development model and the mix model, which we will now briefly review.

1.4.1. The accountability model

The accountability model is managerial, control oriented, judgmental and hierarchical (Monyatsi, 2006). Goddard and Emerson (1995, p. 15) summarize the essence of the accountability model of appraisal when they state that, in its purest form, it identifies incompetent teachers, identifies

weaknesses in teachers' performance, assesses performance for the purposes of pay and promotion and provides evidence for disciplinary procedures.

The accountability model has been unpopular with teachers and their unions because of the following: first, its key characteristic has been seen as imposition since the philosophy is the checking of competence; second, it is designed to bring about a better relationship between pay, responsibilities and performance; third, it is judgmental, and teachers have questioned the capabilities of those making judgements, and the validity and reliability of the instruments used; fourth, the model fosters defensiveness because teachers fight to serve their interests and not those of the clients - students; and fifth, it provides evidence for disciplinary procedures (Monyatsi, Steyn, & Kamper, 2006) (Monyatsi, 2006).

Accountability evaluation became a permanent feature of performance management and outsourcing, in the new public management movement. Evaluation took on new expressions in the form of customer-oriented evaluation. Value-for-money evaluation in the form of cost-effectiveness and productivity studies was highly regarded.

The deepening of the relationship between the education system and society has made schools increasingly subject to social demands and, at the same time, public scrutiny of the quality of service they provide. Accountability has become a social requirement, aiming to guarantee citizens satisfaction levels that are adequate to their real needs (Morgado & Carvalho, 2012).

According to Graça et al. (2011, p.23) this model presents itself as "an instrument that aims to measure the performance of the teacher and has characteristics, mainly, summative, and in which the score reached by the evaluated serves to position him on a scale of evaluation." According to the accountability model, the evaluator should be a senior professional, that is, a teacher in a higher level of the career.

Accountability-centered assessment is usually defined in terms of minimum competencies and is seen as quality control from above, linking to the concept of accountability (Simões, 2000, p. 15). Teaching is thus understood as a job, as it requires mechanisms of supervision and control by the management bodies, which end up leading to dismissal or promotion decisions.

1.4.2. The professional development model

In the professional development model the objective is to develop the teacher professionally. The function of evaluator is "assumed by a panel of teachers who are in charge of coordinating the different disciplinary groups, and there is no need for the figure of the senior teacher" (Graça, et al. 2011, p. 90).

Evaluations are conducted annually or even in less time, depending on previous evaluations. In this evaluation model, the evaluator assumes a partner and collaborator position, not just the position of examiner and observer. The evaluation procedures are as follows: the prior interview, the information collection and the final interview as in the accountability model, however the result of the evaluation is descriptive, not being valued the quantitative component of the evaluation result (*Ibidem*).

In a first stage of the evaluation, evaluator and evaluated must meet and define the areas in which the evaluation will be focused, always based on the service distributed to the evaluated teacher. Then information is collected "which presupposes the follow-up of the work that the teacher develops" (Graça, et al. 2011, p. 23) and later the teacher performance indexes are compared with those established by the school.

After obtaining the performance result, the evaluator and the evaluated must meet and the evaluator gives feedback on the performance of the evaluated teacher. The positive aspects of performance and the areas needs to be improved should be recognized. At the least, an improvement plan and procedures for following and monitoring the progress of the teacher evaluated may also be elaborated (Coelho, 2011, p. 90).

According to Morais and Medeiros (2007, p.27), "the approach to professional development, centered on teachers' conceptions of teaching and learning processes, starts from the beginning of the teacher's career and extends throughout the course of his professional life."

In the same sense, Oliveira-Formosinho (2009, p. 226) argues that

"professional development is like a continuous process of improvement of teaching practices, centered on the teacher, or a group of teachers interacting, including formal and non-formal moments, With the concern of promoting educational changes for the benefit of students, families and communities".

This definition presupposes that the great purpose of professional development processes is not only personal enrichment but also to the benefit of students.

To Amelvoort et al. (2009 cit. by Coelho, 2011, p.90), evaluating teachers for professional development would have as a main result "to be able to assure feedback to the teacher about their performance, as well as their contribution to the school, and would result in a strictly qualitative assessment (without any quantitative classification) and in a professional development plan, which would integrate the professional registers of the teacher.

According to Day et al. (Day, Sammons, Stobart, Kington & Gu, 2007) "the management of evaluation for professional development, it is necessary to take into account the relationship of

evaluation with the autonomy of the teacher, as well as reflection, learning and change of the teacher”.

1.4.3. The mixed model

The mixed model "is a model that seeks, on the one hand, to privilege the quality of the teacher's performance and, on the other hand, to ensure that the professional development needs of the teacher are ensured throughout his career" (Graça et al. 2011, p. 24).

In this model, evaluation moments according to the accountability model and moments of evaluation according to the professional development model occur in an intercalated way. "The results of the career advancement assessment can provide information for each teacher's professional development plan and provide useful feedback for improving evaluation processes for development" (Coelho, 2011, p.90). This model does not overwhelm the senior evaluator and simultaneously builds a "collaborative culture of professional development" (Graça et al., 2011, p.23).

Ramalho's theoretical proposal (2014) looks at teaching performance as a multi-referenceable object of evaluation, with the confluence of synchronic and diachronically related media, institutions and actors (either by convergence or divergence) among students, educators, teachers, coordinators, directors, political systems, ideological postulates, economics and society in general, attributing to the processes of referentialization the character of dynamic and necessarily multi-discursive realities. It is projected in a complex way and with very diffuse meanings, whose references and referents cannot be reducible to each other, nor mutually excluded. It is therefore the prospect of an object of evaluation that is multi-referenciable through a diversity of texts, contexts, actors and perspectives very diversified and contemplated in narratives and multiple rationalities of the conception of teaching performance.

The proposal of (re)conceptualization the teaching performance is configured around four references and respective narratives, namely: i) the bureaucratic rational reference implying the normalization narrative; ii) the managerial technical reference bearer of a management narrative; iii) professional technical reference with the narrative of professionalism; and, finally, iv) the political frame of reference aligned with the narrative of dissent (Ramalho, 2014).

Quality of education encompasses "different interpretations and covers different conceptual and political frameworks of conceiving education, relating it to the type of society and citizenship that one wants to construct" (Candau, 2013, p. 10). This is why it is a concept "socially constructed and in constant reformulation" (Candau, 2013) and its polysemy has generated frequent debates,

therefore there is a need to clarify what is meant by quality in education when we refer to this concept.

For Stronge (2010) a quality assessment system also emphasizes the relationship between teacher evaluation and the organization's objectives and the importance of the context, highlighting several contextual aspects such as the disciplines that the teacher teaches, the size of the class, the conditions and quality of the classroom, the resources for the exercise of the teaching profession and the resources for professional teacher development.

Salgueiro and Costa (2013) identify three logics of organizational action in the evaluation of teachers. On the one hand, a "bureaucratic logic" resulting from an overly complex process of operationalization that resulted, above all, from the need to over-involve schools and teachers in the construction, application and completion of documents (individual objectives, lesson plans, self-assessment sheet, classroom observation grid, etc.).

On the other hand, a "conflictual logic" deriving mainly from peer evaluation and the competitiveness inherent in the evaluation process, through which hierarchical factors were created within a very equal professional culture. And finally, an "artificial logic" marked by the awareness of the different actors that teacher evaluation is not aimed at professional and organizational development, but rather a way to control, monitor and "supervise" the work of teachers.

For the same authors,

"in addition to the bureaucratization of the process that occupied excessive time for teachers, the two most controversial factors and the greatest difficulties created were the division of the career and the attribution of the evaluative function to the titular teachers, within one of the Major innovations of this system - peer review" (Salgueiro & Costa, 2013, p. 104).

For Alves and Aguiar (2013, p. 146) the evaluation of teachers in Portugal provoked a climate of tensions, since the almost absolute centrality of the school in the organization of the evaluation process led to "an accumulation of bureaucratic work and responsibilities, to an environment of distrust, demotivation and distance between teachers, which often generates isolation and competitiveness among teachers".

Other studies on the Portuguese case (Filipe, 2011) (Oliveira, 2012) (Ribeiro, 2011) show that one of the perverse effects of school-centered evaluation is the decrease in collaborative practices due, on the one hand, to the lack of time resulting from excessive "bureaucracy", and on the other hand due to the increase of the competitiveness between the pairs.

To Marcos (2013), internal and peer conflicts contribute to a reduction of supervisory logics between evaluators and evaluated, jeopardizing one of the greatest potentialities of “evaluation in situ”.

This configuration of the Portuguese teacher evaluation system is marked by a set of tensions, some of which are of a constitutive nature and which are transversal to the solutions found in several European countries. From the point of view of legitimacy, this is “a set of options that reinforce the control of performance and increase the process of accountability through results, putting itself in line with current changes in the teaching profession” (Machado, Abelha, Barreira, & Salgueiro, 2012, p. 64).

Competitiveness and isolation is, according to Alves and Aguiar (2013, p. 146), a negative effect resulting from the fact that performance evaluation is performed by peers, thus not contributing to the “development of the teacher, either as a professional or as a person” leading to “weakening of the relationship between teachers and, consequently, the deterioration of the school environment”. To Araújo (2013) the lack of specific training and skills for this "new" function of peer evaluation is another constraint to this teacher evaluation system, both by teachers evaluated and by evaluating teachers.

With regard to the question "who should the evaluators be?" The opinions oscillate around the following criteria: i) teachers outside the school; ii) a mixed team of outside teachers and teachers of the school; iii) higher education teachers with training in the area of evaluation (Gomes, 2010) (Cardoso, 2012) (Carneiro, 2011).

According to Garcia (2011), peer evaluation has adopted classification and personnel selection objectives, and is seen by teachers as a disruptive personal relationship and a source of conflicts (Cardoso, 2012) that weakened the development of collaborative dynamics. The inadequacies of the selection process of the evaluators and the lack of specialized training of the evaluators will have also shaken the legitimacy for the exercise of the function and the usefulness of the teacher evaluation process itself (Coelho, 2011).

The results of the referred investigations point to peer evaluation as a difficult process and inducing a climate of uneasiness among teachers. For this reason, teachers point to the preference for "evaluative modes outside the context, when the importance of knowledge of the context is repeatedly affirmed for a more rigorous evaluation" (Roldão, 2013, p. 172).

Currently the evaluation of teacher performance constitutes the most demanding professional dimension and the most challenged by teachers in school every day. Successive educational

reforms and measures implemented to make public schools more effective and efficient have interfered with the internal climate of schools, marked less by hope and more by professional discouragement and dissatisfaction.

According to Day (2004), this ongoing transformation is marked by the substitution of strong liberal and humanist traditions, based on a non-instrumental vision and the intrinsic value of education, by functionalist and performative educational policies that seek to respond to the increase of economic competitiveness through reforms centered on accountability the salary associated with merit, and indirect and recentralized forms of regulation (Afonso, 2011).

The emergence of a “generalized performative logic” (Machado, Abelha, Barreira & Salgueiro, 2012, p. 58) that take place under the sign of “obligation of results” and “accountability”, translating into “evaluations essentially aimed at the teachers determine the effectiveness of their pedagogical practices - in support of learning - and their evaluation practices” (Tardif & Foucher, 2010).

Indeed, the various figures of the teaching profession, from the "Master" (Steiner, 2005) to that of the “reflective practitioner” (Estrela, 2011) rooted in a humanist rhetoric of the teacher as an agent of the progress of subjects, Humanity, have lost (or are in the process of losing) their professional, scientific and social relevance vis-a-vis the figure of the “performative teacher” (Alves, Flores, & Machado, 2011). The evaluation of teacher performance has worked both as a legitimating discourse and as a regulator of action, consubstantiating the prevalence of a “metanarrative of control” (Alves & Machado, 2011).

Machado (2013) points to the establishment of a “regime of omni-evaluation”, which results in at least three evident consequences in the educational field.

First, the shift from “evaluation policies” to “evaluation as a policy” (Lima, 2011) recognizing that there is no turning away, but a reconGraphiction of the role of the State in the regulation of education that passes through more subtle strategies and in which the centrality of evaluation prevails. Secondly, the transition from a logic of separation and specialization to a logic of democratization and universalization of the functions of evaluators and evaluators (although they are both evaluators and evaluators), which allows all a (re)centralized state regulation. Finally, in the third place, the reinforcement of the practices of self-assessment, the last avatar of omni-evaluation, conciliating the individualist legitimacy of the present democratic societies with internalization, in the subject itself, of evaluative schizophrenia: more than each being evaluator of

others and evaluated Of the others, what matters is each one is, finally, evaluator and evaluated of itself (Machado & Abelha, 2014, p. 58).

To Machado and Abelha (2014) the evaluation of teachers assumes, in this context of "omniavaluation", a central role in educational policies and in the construction of the identity of the teaching profession, being torn between performance pressures and emancipatory desires.

In this respect, changes in Portugal, especially in the period 2007-2013, contribute to the discussion of the logics, processes and devices for teacher evaluation, both with regard to the endogenous implications and with regard to the exogenous effects to the own Evaluation process. The main lessons we draw from the analysis of these two aspects that have deserved particular attention in the literature: i) school-centered evaluation and ii) peer evaluation.

1.5. The professional development

The professional development takes place broadly in five areas: *pedagogical development*, which focuses on capacity building; *professional development*, which focuses on individual teacher growth within the context of their professional role; *organizational development*, highlighting the needs, priorities and organization of the institution; *development of the professional trajectory*, which focuses on the preparation for career advancement; *personal development*, which highlights the importance of planning, interpersonal skills and growth of the teacher as an individual (Riegle, 1987, cit. por Graça et al. 2011, p. 25).

This approach aims to consolidate respect for the professionalism of teachers, from the point of view of their pedagogical responsibilities, as well as the autonomy of schools.

From the epistemological point of view, effective professional development is considered: (i) experiential, involving teachers in concrete questions of teaching, observation and reflection; (ii) inquiring, is based on reflection and experimentation; (iii) collaborative and interactive, it involves the sharing of knowledge among educators and focused mainly on group practices rather than individual practices; (iv) articulated and resulting from the work that teachers do with students; (v) grounded, continuous, intensive and sustained by collaborative work and collective resolution of specific problems of practice; (vi) articulated with other aspects of change at the organizational level (Darling-Hammond & McLaughlin, 1995 cit. por Morais & Medeiros, 2007, p. 27).

For Pacheco and Flores (1999, p. 168) professional development is "a continuous process of learning that includes, on the one hand, the acquisition of new skills resulting from practices of school innovation and, on the other hand, the consolidation of skills acquired and maintained throughout the career".

There are three types of factors that contribute to the inefficiency of the evaluation aimed at professional development: those related to the context (organizational climate, resources and leadership); factors related to the procedures (instruments used for the collection of information and use of other sources of feedback); factors related to the teacher (such as motivation and effectiveness) (Pacheco, 2011).

The position of De Ketele (2010) gains relevance, which argues for the need to

...speak more about professional development and evaluation of teaching than about evaluation of teachers, since evaluation is not only a process leading to an assessment or a judgment but rather a process of assessment and evaluation. In order to collect a set of relevant, valid and reliable information and to compare this set of information with a set of criteria, which must be consistent with a relevant reference to support decision-making appropriate to the function envisaged (pp. 9-10).

The concepts of professional development, quality and quality development have been included in purposes and practices of change and improvement of teaching and learning, which demands an analysis of evaluation procedures that contribute to it.

In education, it is precisely at the level of learning and teacher professional development that the evaluation can make the difference, especially if contributes to improving the scientific and pedagogical training of teachers, helps to consolidate commitments and contributes to renew mentalities (Morgado, 2014).

One of the most controversial and most critical issues in this area has been the tendency to establish teacher training in the light of what is commonly referred to as performativity culture, that is, training that enables the acquisition of predetermined competencies, performances and results, embodying controlled professionalism, which emphasize the importance of "increasingly sophisticated management processes and systems of outcome measures", emphasizing accountability and competitiveness (Candau, 2013, p. 13). To change this situation, it is necessary to develop training policies that create conditions for teachers to renew and strengthen their professionalism, through autonomous, competent and creative actions, capable of giving them another leading role in terms of curricular decision (Morgado, 2014).

The professional development is a process in which teachers gradually acquire a body of knowledge and skills to improve the quality of teaching and to enhance learner outcomes. This acknowledgement is centrally important in maintaining and enhancing the quality of teaching and learning in schools (Craft 2000; Harland and Kinder, 1997; Harris 2002).

To Fullan (2011) professional development is a career-long project as it is a continuous, interactive, cumulative learning that combines a variety of learning formats.

Appraisal for professional development has therefore gained a good deal of popularity from both teachers and their organizations, including school managers (Duke & Stiggins, 1990:116; Duke, 1995b:6). This may be a result of various factors. First, the staff of professional development model is viewed as a genuine two-way process between appraisers and appraisee. Second, it takes place in an atmosphere of trust and confidentiality. Reflection is the buzzword (Cosh, 1999:260).

Murdock (2000:55) points out that a modern system of evaluation should encourage teachers to become reflective practitioners. Third, it is based on the belief that teachers wish to improve their performance in order to enhance the students' learning. Fourth, the key characteristic of the model is negotiation and the philosophy is the supporting of teaching and managerial development. Teachers are involved and this is supported by Murdock (2000, p.55) who maintains that participation by staff in initiating and contributing to the instruments and procedures used to evaluate their performance leads to motivation and empowerment as teachers develop a full understanding of the whole programme. Fifth, it identifies the teacher's potential for career development.

Professionalism has been the subject of many studies over the last century. Adopting a macro-perspective, Hargreaves (2000) has presented the development of professionalism as passing through four historical ages in many countries: the 'pre professional' (managerially demanding but technically simple in terms of pedagogy); the 'autonomous' (marked by a challenge to the uniform view of pedagogy, teacher individualism in and wide areas for discretionary decision taking); 'collegial' (the building of strong collaborative cultures alongside role expansion, diffusion and intensification); and the 'post-professional' (where teachers struggle to counter centralized curricula, testing regimes and external surveillance, and the economic imperatives of marketization) (Hargreaves, 2000, p. 153).

Intensification and bureaucratization increased forms of managerialism, and greater accountability and public scrutiny are but a few examples of the most recent changes. Over the last two decades, schools, and therefore teachers, have been confronted with new challenges, such as increasing roles and responsibilities; changes occurring in social agencies; greater influence of the mass media on the education of children and young people; the co-existence of different educational models in a multicultural society; the fragmentation of teachers' work; growing opportunities for learning outside school owing to the development of information and communication technologies;

and increasing contractual accountability, bureaucracy and public scrutiny (Day, 2010) (Day, 2002) (Day, Flores & Viana, 2007) (Esteves, 2000).

The international research literature has consistently shown that professional development is an essential component of successful school level change and development (Day, 2002) (Goodson, 2014). It has confirmed that where teachers are able to access new ideas and to share experiences more readily, there is greater potential for school and classroom improvement. Improving schools invest in the development of their staff and create opportunities for teachers to collaborate and to share best practice. Evidence also suggests that attention to teacher learning can impact directly upon improvements in student learning and achievement (Hargreaves, 2000). Where teachers expand and develop their own teaching repertoires and are clear in their purposes, it is more likely that they will provide an increased range of learning opportunities for students (Joyce et al. 1999). The research literature demonstrates that professional development can have a positive impact on curriculum, pedagogy, as well as teachers' sense of commitment and their relationships with students (Talbert and McLaughlin 1994).

Much research literature demonstrates that events and experiences in the personal lives of teachers are intimately linked to the performance of their professional roles (Ball & Goodson, 1985) (Goodson, 2014). Teachers' sense of professional, personal identity is a key variable in their motivation, job fulfilment, commitment and self-efficacy; and these will themselves be affected by the extent to which teachers' own needs for autonomy, competence and relatedness are met. Reforms have an impact upon teachers' identities and because these are both cognitive and emotional, create reactions which are both rational and non-rational. Thus, the ways and extent to which reforms are received, adopted, adapted and sustained or not sustained will be influenced by the extent to which they challenge existing identities (Day, 2002).

Fernet, Gagné and Austin (2010) showed that peer relations at school were positively associated with employees' autonomous motivation and negatively with burnout. Quality of relations with staff has also been positively associated with autonomous motivation and occupational commitment, and negatively with controlled motivation and emotional exhaustion (Fernet, Guay, Senécal & Austin, 2012).

Fernet, Guay Senécal and Austin (2012) showed that higher job demands (classroom overload, students' disruptive behaviors) over a school year reduced teachers' autonomous motivation (relative to controlled) in the classroom, which in turn predicted greater emotional exhaustion. Their analysis took into consideration the role of teachers' feelings of self-efficacy in class. The same

authors found that job resources (job control, recognition, and quality of relationships with the staff) increased autonomous motivation and reduced controlled motivation over a 9-month period in French-Canadian school principals. They also found that autonomous motivation fostered occupational commitment and helped prevent emotional exhaustion, whereas controlled motivation led to emotional exhaustion over the same period (Fernet, Guay, Senécal & Austin, 2012).

Nie, Chua, Yeung, Ryan & Chan (2015) showed that teachers' perceptions of the support provided by their immediate superior were positively associated with intrinsic, identified, and introjected regulation, but negatively with external regulation.

Yet sustaining a positive sense of identity to subject, relationships and roles is important to maintaining motivation, self-esteem or self-efficacy, job satisfaction, and commitment to teaching; and although research shows consistently that identity is affected, positively and negatively, by classroom experiences, organizational culture and situation specific events which may threaten existing norms and practices (Day, Kington, Stobart & Sammons, 2006) (Flores, 2002).

Successive reform implementation strategies have failed to address its key role in effective teaching. Reform which addresses key issues of professional identity, commitment and change is more likely to meet the standards raising recruitment and retention agendas more efficiently and more effectively than current efforts which, though well intentioned, appear from empirical data to be failing to connect with the long term learning and achievement needs of teachers and students (Day, 2002).

A study about experienced teachers in England and Australia (were interviewed about their understandings of commitment) suggest that commitment may be better understood as a nested phenomenon at the center of which is a set of core, relatively permanent values based upon personal beliefs, images of self, role and identity which are subject to challenge by change which is socio-politically constructed (Day, Elliot & Alison, 2005).

Day and Gu (2007), by this turn, demonstrated that teachers do not necessarily learn through experience; that expertise is not acquired in an even, incremental way; and that teachers are at greater risk of being less effective in later phases of their professional lives. The paper develops these findings. Moreover, it argues that the contexts for teachers' professional learning and development are, by definition, different from those who do not work in human service organizations, since teachers are essentially engaged in work which has fundamental moral and ethical as well as instrumental purposes. Their capacity to exercise these effectively relates to their ability to manage positive and negative 'scenarios' in different professional life phases. It suggests,

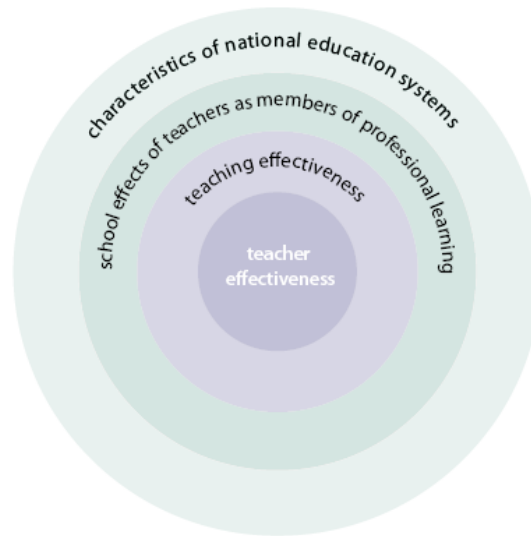
therefore, that to be effective, professional learning opportunities must be designed which take account of the personal, workplace and external scenarios which challenge their commitment to these core purposes.

Bolívar (2007) points the existence of partnerships and networks between schools and social associations work together to disseminate knowledge and good educative practices and then a constitution of a professional community with the mission of increment the professional learning and the professional development.

It is the new educational contract already proposed by Nóvoa (1999) for the future of education (4th historical time in Portugal) and the re-creation of the school as a public space for collective decision based on a new conception of citizenship also defended by Barroso (2005, p.746). In the same way, Mesquita et al. (2012) emphasize the constitution of communities of practice, educational communities, educational teams and / or collaborative cultures. They also emphasize that the relationship between solitary knowledge and collective knowledge calls for collaborative work around inclusive curricular projects and considers teacher cooperation as an integral part of change in education (p.11).

Literature on educational effectiveness research and teacher professional development seems to outline a conceptual framework that can be described as an 'onion-rings' model, going from the micro-level to the macro-level perspective – with individual teachers' personal characteristics (competences, beliefs and attitudes) at the core, a second layer concerning teaching effectiveness in the classroom (instructional repertoires), a further layer about teachers' cooperation in school contexts, and finally considering national policies and organizational features (including issues of autonomy, accountability, evaluation in education systems) as the outer layer (Richardson & Placier, 2001) (Clarke & Hollingsworth, 2002).

Figure 2. Layers of analysis in identifying contents and forms of teachers' professional development



Source: *Teachers' Professional Development: Europe in international comparison, a secondary analysis based on the TALIS dataset*. Ed.: Jaap Scheerens. European Commission, Luxembourg 2010.

Teachers' professional learning and development are essential components of school development, as well as of teachers' professional growth, well-being and success. Muijs and his collaborators (Muijs, Day, Harris & G. Lindsay, 2004, p. 291) argued that teachers' ability to reflect, access new ideas, experiment and share experiences within school cultures gives greater potential for school and classroom improvement. Moreover, teachers' learning at work also affects students' attitudes toward learning, teaching processes and achievements.

Day and his collaborators (Day, Sammons, Stobart, Kington & Gu, 2007) found that teachers in various phases of their professional careers consider continuing professional development activities to be important; these activities represent an investment in their professional lives and are a "means of recharging their batteries" (p. 148). A general feature of the knowledge society is the demand for continual learning and development of professionals, including teachers, after their period of professional preparation is finished (Day & Sachs, 2004, pp. 3-5). Because teachers benefit from professional preparation and learning, and are demanded to develop and learn, their ability to do so thus becomes an important research topic.

The concept of continuing professional development (CPD) in education is used to describe all the activities in which teachers engage during the course of a career which are designed to enhance their work. Day's (1999) definition of CPD encompasses all behaviors which are intended to effect change in the classroom:

Professional development consists of all natural learning experiences and those conscious and planned activities which are intended to be of direct or indirect benefit to the individual, group or school, which contribute, through these, to the quality of education in the classroom. It is the process by which, alone and with others, teachers review, renew and extend their commitment as change agents to the moral purpose of teaching; and by which they acquire and develop critically the knowledge, skills and emotional intelligence essential to good professional thinking, planning and practice with children, young people and colleagues throughout each phase of their teaching lives. (Day, 1999, p. 4).

Teachers' perceptions of what activities constitute CPD is frequently limited to attendance at courses, conferences and whole-school INSET days, often to meet national requirements. Professional learning, or "on the job" learning is regularly seen by teachers as separate from CPD, and something that is just done as part of the job (Edmonds & Lee, 2002; Hustler et al. 2003; Robinson & Sebba, 2004). However, the literature points to several facets of effective CPD, many of which are far removed from the commonly-held perceptions of CPD as one-off events.

"The idea of lifelong learning takes on a great centrality and is the subject of great debate, particularly in the European context, reinforcing the relevance attributed to the education sector in contemporary societies and current educational policies" (Alves M, 2010, p. 8). Here, teachers assume increased responsibilities, since they continue to be seen as "effective agents of change, largely dependent on the transformations that need to be made in school and teaching, as well as on the educational success of students and their Achievement as people " (Morgado, 2011, p. 795).

It is within this framework of increased responsibility that professional development of teachers, since it is a process that, according to Christopher Day (2001, p. 16), allows a constant scientific and pedagogical update, helps to consolidate commitments, contributes to renew mentalities and requires a predisposition for the learning. It is precisely at the level of learning and professional development of teachers that evaluation can make a difference.

Englund (1996) points the distinction between (teacher) professionalization and professionalism – the first should be viewed as a sociological process and the second as a pedagogical process, concerned with internal quality of teaching as a profession. Professionalization is a measure of the societal strength and authority of an occupational group, while "professionalism focuses on the question of what qualifications, acquired capacities and competences are required for the successful teaching" (p.76).

To Goodson (2000) professionalization is related to "promoting the material and ideal interests of an occupational group" (p.182).

The dominant discourses in the field of education indicate that teacher professionalism is associated with improving the quality and standards of teachers' works and their public image. Multiple approaches are common in the sense that teacher professionalism means meeting certain standards in education and related to proficiency. However, the meaning of the term and status of teaching profession is considered to be highly problematic and polarized in various spheres (Demirkasimođlul, 2010).

1.6. Empirical studies

Several empirical studies have attempted to identify and compare prominent characteristics of teacher's competences in higher education as evaluated by students and self-evaluated by teachers. A number of studies in the past have also looked at whether the personal characteristics of the student and of the teacher influence the evaluation ratings.

1.6.1. Teacher's evaluation by students

Higher education institutions engage in advancing good teaching because first, they are interested in demonstrating that they are reliable providers of good quality education, while serving multiple stakeholders with different expectations. Second, they are also required to respond to the increasing demand for meaningful and relevant teaching. Students, employers and policy makers want to assure that education will prepare students for rewarding employment and for a professional growth. Third, research performance is no longer sufficient to maintain the reputation of the higher education institutions and balancing performance on teaching and learning achievements with research performance has become essential.

When students assess their teachers, they emphasize mostly on the personal and social factors (Magno & Sembrano, 2007). Predominant characteristics evaluated by students included the personal and social competences (Alhija, 2017) which seem to show that students share their perceptions about good teaching and good instructors differently. Students attribute greatest importance to teacher's ability to properly explain the subjects (Alhija, 2017). Teaching in a clear and understandable manner is critical for transmitting learning material effectively as well as for communicating to students the teaching goals, expectations, and anticipated outcomes. Being helpful to students, accessible, respectful and understand students were the highest personal and social competences valuated by students, as had seen before in Al-Mohaimeed and Khan (2014) and Alhija (2017). Otherwise, students recognize a highest level of dignity, honesty, justice, fairness

and self-confidence of UNTL teachers (personal competences) but point a lack of authority (Alhija, 2017).

Alhija (2017) explores students' perceptions of major dimensions characterizing good teaching and examines whether students' perceptions differ according to gender, type of educational institution, discipline, degree and year of study. This study was conducted in the Israeli higher education system, which has undergone significant expansion and diversification since the 1990s. A 36 item questionnaire was implemented in a 2475 Israeli students' sample in order to conceptualize and examine five dimensions relating to good teaching: goals to be achieved; long-term students' development; teaching methods and characteristics; relationships with students; assessment.

The students perceived the assessment dimension as the most important followed, in descending order, by goals to be achieved, relations with students, and teaching methods and characteristics. The long-term student development was the least important dimension. The most important goal to be achieved by good instructors is "stimulating motivation and interest in the course subject". They also viewed 'promoting creative and innovative thinking capabilities' as the most important practice for long-term student development dimension. As to the teaching methods and qualities dimension, students attribute the highest importance to "teaching in a clear and understandable manner". Students point low importance to "fostering oral capabilities", "promoting written capabilities" and "advancing in public presentation capabilities". "Allowing continuous communication with students" gained the highest rating among the four items measuring instructor's relations with students. As for the features within the assessment dimension, students attributed the highest importance to "grading tasks, projects and examinations objectively and fairly" (Alhija, 2017).

The gender and field of study have a significant effect on students' perceptions of good teaching. Gender differences were found regarding all five dimensions of good teaching, whereby female students assign to all of them greater importance than did male students (Alhija, 2017).

Other previous studies have already demonstrated gender differences. Morgan (2001) found that female students receive more positive feedback from teacher making them higher rated than do male students. Anderson et al. (2012) saw in their study among doctoral students women were more likely to endorse the traits of professional, expert and student centered as characteristics of good instructors. Lavin, Korte and Davies (2012) as well Korte, Lavin and Davies (2013) found that female students tend to assign a higher rank to traits related to effective teaching than did male students (pedagogical and professional competences). Al-Mohaimed & Khan (2014) didn't see

significant differences between male and female students relating to performance and personality traits. According Alhija (2017) gender has a significant effect on students' perceptions of good teaching.

The relationship between students' age and their evaluation on teacher's competences has been the subject of little empirical attention in research on teaching competences evaluation. Alhija (2017) found the older and more mature students perceived student long-term development as significantly more important, so they tend to attribute a higher value to the pedagogical and professional competences of teachers.

Compared to students in other fields of study, students in the exact sciences in general assigned less importance to four of the teaching dimensions: goals to be achieved, long-term student development, relations with students and teaching methods. Talbot (2011) finds that mathematics and science teachers are rated highly on their strategic knowledge that pertains to pedagogical and professional competence. The study of Talbot (2011) further explains that science teachers are rated higher on pedagogical and professional competency because it matches their characteristics of applying and modifying their instructional repertoire in a given teaching context. Students in the social sciences and humanities (soft) attributed to instructor-student relations and teaching methods greater importance compared to students in natural and exact sciences (hard) (Alhija, 2017). These findings also concur with results pertaining to medical students (Al-Mohaimed & Khan, 2014) who valued characteristics related to instructors' performance more than their personality traits. The medical students put more emphasis on performance than on personality, but there were no significant differences between male and female students relating to performance and personality traits, as well did not vary between the juniors and the seniors students. Students' class level did not show a significant impact on their perception of qualities of a good teacher. Students' class level did not show any influence on their perception of qualities of a good teacher. Being helpful to students, easily accessible, being respectful and understanding/relating well to students was also highly valued by students in different cultures. Some attributes like "sharing of personal experiences," "having good sense of humor", and "dressing up appropriately," "self-sacrificing" and "giving good marks" were not well regarded though they were rated high in other studies. Al-Mohaimed and Khan (2014) concluded that although students valued performance-related attributes more, some personality related attributes are highly valued too. It indicates that students hold their honor very high. Teachers who are courteous and demonstrate this trait through their behavior and actions are deemed 'good'.

The literature regarding the relationship between students' evaluation on teacher's competences and their background characteristics is not extensive. Yet, some research findings exist regarding the relationship between the students' gender, age and field of study (or faculty) and their rating evaluation on teaching competences.

1.6.2. Teachers self-evaluation

Suknaisitha, Suwimon and Piromsombatc (2014) conducted a five-step research about teaching improving by self-monitoring. The purpose of this research was to study the performance of learning educational measurement and evaluation after having been trained about the self-monitoring for performance on educational measurement and evaluation. The results showed that teachers intended to learn about learning educational measurements and evaluations through participation in activities. By applying the self-monitoring process, teachers gained further knowledge in learning measurements and evaluations, participating in action planning, as well as greater realization and awareness of the importance of learning measurements and evaluations. In this study, the researchers emphasized the educational context (by focusing on the performance of learning measurement and evaluation of teachers in schools) and concluded that there are a large variety of problems and obstructions concerning measurement and evaluation, such as: teachers lack the ability of curriculum analysis, learning standards, and level standards because they have the greater burden of other tasks to do; teachers use measurement and evaluation instruments which are not varied. Moreover, teachers still believe that measurements and evaluations are so complicated and difficult to understand and apply. For the reasons mentioned above, Suknaisitha et al. (2014) emphasize the teachers' performances adaptation in terms of educational measurement and evaluation by using the self-monitoring strategy on the educational measurement and evaluation performance of teachers.

The teachers had to think about self-evaluation to be successful every day and this would help increase their thinking skills more. If there was continuous and consistent practice, it would help teachers achieve their performance which they had to do qualitatively. The self-monitoring strategy could change behaviors to be positive as concerns the required goals and could be applied both in academic and social aspects. The crucial principles were about observation and consistently recording the behavior (Suknaisitha et al., 2014).

Faculty development encompasses a complex and multifaceted endeavor. Each contribution in this special issue highlights aspects of this complexity. First of all, there is a recognition that faculty development does not happen in isolation but is impacted by the organizational and social worlds

that faculty members inhabit. In this respect, culture, structure, and agency that impact professional learning of lecturers at Stellenbosch University in South Africa. In response to a call for increased attention to professional learning, Van Schalkwyk et al. (2015) focus on how professional learning is embedded in workplace structures and cultures, describing professional learning as the enhancement or change of teaching and assessment practices of lecturers in order to ensure quality student learning takes place. Their findings reveal the realities and choices involved in becoming or being a good teacher at a research-intensive university. Though these realities are well known, the authors reflect on ways in which dominant discourses, spaces, culture and processes could (and should) be considered and possibly changed by those aiming to support and lead instructional excellence at universities.

A study by Nevgi and Lofstrom (2015), conducted at a research-intensive university in Finland, contributes to the importance of reflection on teacher self-evaluation and self-perception. The authors explore the development of teacher identity through academics' self-image and self-efficacy as teachers, motivation to teach, develop as a teacher and task perception. To describe each teacher's identity and development, core narratives were created based on interviews and practicum reports. Four teacher identity types were identified at the end of university teacher development programme according to the academics' task perception (university teacher/researcher versus educational developer) and reflection on teaching (reflection versus no-reflection). The results indicate that the willingness to reflect is crucial for the development as a university teacher.

The Nevgi and Lofstrom's (2015) research remind us how existing workplace cultures and practices may hinder or foster professional learning of faculty. Professional learning is impacted not only by faculty development initiatives but also by the organizational structures and practices within which faculty members operate. Providers of faculty development should first understand the role of the workplace (including tenure and promotion practices) and then design their initiatives accordingly. Faculty development initiatives should, at best, incorporate reflection on teaching (beliefs) in order to stimulate an actual impact on teaching practice.

Kozikoğlu (2017) conducted a study that aims to identify cognitive constructs of prospective teachers about ideal teacher qualifications. In accordance with this purpose, the study was conducted with 36 prospective teachers by using repertory grid technique. As a result, 356 cognitive constructs were produced by prospective teachers about ideal teacher qualifications and these cognitive constructs were grouped under ten categories: communication skills, student

centeredness, innovativeness, sensitivity, humaneness/ joviality, teaching pedagogical skills, leadership/ guidance, professional content knowledge, personal values and professional values. Apart from these competences, it is seen that prospective teachers give importance to effective teacher-student interaction such as communication skills, student centeredness. The author defends that prospective teachers' cognitive constructs are influenced by their individual experiences and their learning/knowledge resulting from their interaction with the external environment, since they attend at different grade levels, differ in educational experiences and types of teachers they met during their educational life.

Cognitive constructs produced by prospective teachers and their scores were analyzed in different ways. Considering the cognitive construct categories' importance level, the first three cognitive construct categories are humaneness/joviality, teaching pedagogical skills and personal values, by this order. Similarly, considering total scores the first three cognitive construct categories are teaching pedagogical skills, humaneness/ joviality and personal values. Considering mean scores of cognitive construct categories, the first three categories are teaching pedagogical skills, humaneness/ joviality and professional content knowledge. The first four construct categories selected by a large number of participants were teaching pedagogical skills, humaneness/joviality and personal-professional values (Kozikoğlu, 2017). On the other hand, it was concluded that the least three cognitive construct categories are innovativeness, sensitivity and leadership/guidance, by this order. In accordance with these results, authors concluded that ideal teachers should have qualifications such as humaneness, joviality and personal values as well as professional knowledge (content knowledge and pedagogical skills). Teacher's personality, humanistic approach towards students, professional knowledge and skills are key elements in being perceived as ideal teacher by the students. By this way, the author suggests a new education paradigm that has been formed: while teacher was seen as authoritarian and the only source of knowledge in traditional education theories and educational philosophies, the role of teacher in modern education and contemporary educational philosophies is seen as a guide and consultant. In this case, teachers' emphasis on the personal and humanistic characteristics of the ideal teacher or good teacher and giving importance to the ethical and social aspects of the teacher are consistent with modern education and contemporary educational philosophies. Kozikoğlu (2017) defends that there should be selective courses in teacher training programs in order to develop prospective teachers' personality, personal values and teaching pedagogical skills; a warm, lively and interactive environment should

be created in the classroom so that prospective teachers can develop their sense of understanding, humor and communication skills.

1.6.3. Differences / Similarities between students and teachers competences

A study developed by Jónsson, Smith and Geirsdóttir (2018) addressed the issue of variability of perception of teachers and students regarding feedback and assessment. The study aimed to explore how teachers and students perceive assessment in three upper-secondary schools in Iceland. In this paper, feedback refers both to the information about students' work and their engagement with the feedback information.

It is at the center of formative assessment, mainly located in the dialogue between students and teachers. Smith, Gamlem and Engelsen (2017) used the term "responsive pedagogy" for "the learning dialogue" taking place between the student and the teacher about goals, competence in achieving those goals, and strategies for getting there, as well to enhancing further learning.

The Jónsson et al. (2018) survey consisted of 30 shared statements (questions) both for students and teachers, which made possible to compare their perception of how they engaged with feedback and assessment practices. The questions are condensed to four dimensions: quality of feedback, students' use of feedback, peer feedback and student involvement in assessment practice. Two statements about self-efficacy were added to the students' questionnaire. The participants were asked to consider whether they agreed with the statements or not. For each statement, the participants could select one of four options: (4) strongly agree, (3) agree to some extent, (2) mostly disagree and (1) strongly disagree. Students were requested to answer the survey for five main academic programs, or fields of study: Icelandic, mathematics, science, foreign language and social science-humanities.

The findings show a general discrepancy in how teachers and students perceive student involvement in assessment, quality of feedback and students' use of feedback. Teachers seem to overestimate how much students are involved in the feedback dialogue and the assessment process. Teachers rate the quality of the feedback and the students' use of feedback more highly than students. In relation to peer assessment, teachers and students agree on the moderate use of feedback by peers. Most of the students experience feedback as being useful for further learning and as providing information on what is expected of them, but there was also noted that some students did not agree that the feedback they received was useful. Teachers' and students' disagreement about how they perceive discussion regarding learning is prominent. When examining how teachers discuss ways to improve students' work, all teachers, except one, agree

or strongly agree to the statement, compared to 70% of students. The majority of teachers reply that they discuss with students what to address in the subject compared to only 38% of students. A rather strong formative assessment culture is predominant as illustrated by the fact that almost all teachers, independent of school, report giving purely formative feedback. Self-efficacy is rated highly by most of the students in all the three schools. Jónsson et al. (2018) suggest that policymakers should encourage the implementation of formative assessment. It is also vital that teachers get the support they need to carry out the project in the long run. The authors also argued that students seem to be more influenced by the discourse assessment compared to the teachers. It could be argued that culture around formative assessment opens space for possible renegotiation of assessment.

1.6.4. Teacher performance evaluation (interviews)

In relation to educators' role and responses to policies and accountability expectations, the authors deal with the social effects of the meaning of policy, to which educators are being called upon to respond.

Ehren and Shackleton (2016) study the impact school inspections have on schools' outcomes over a time period during which they compare schools placed in different inspection treatment categories (weak/very weak and basic) after the early warning analysis. The authors question the potential effectiveness of such inspection targeted models on student attainment and other performance indicators. Ehren and Shackleton (2016) analyzed differential changes in other performance indicators between schools in different inspection categories, expecting the most changes in weak/very weak schools on indicators part of the early warning analysis of the inspectorate of education (student numbers). Their results only show differences in changes in student satisfaction, student-staff ratios and number of full-time students in weak/very weak schools, compared to schools in the basic inspection category, suggesting that schools do not specifically target improvement on indicators in the early warning analysis. Student satisfaction declined, as well as student numbers and student-staff ratios in weak and very weak schools over time. This would suggest that students are less likely to choose schools that are evaluated as weak or very weak by the inspectorate, and students in weak and very weak schools become less satisfied when the school is assessed to be failing. Such a decline in student satisfaction may result from an overall lack of morale in the school. Ehren and Shackleton (2016) point to various unintended consequences of risk-based models of school inspection. So, the authors defend weak schools

should have accompanying accountability approaches and sanctions, to guarantee the efficacy of their actions to improve measured aspects quickly and in sustainable way.

Teodorović, Stanković, Bodroža, Milin & Đerić (2016) analyze characteristics and patterns of formation, implementation and evaluation of three major education policies in Serbia: in-service training of teachers (INSET), school development planning (SDP) and inclusive education (IE) – three educational reforms through the lens of key actors: principals, teachers and counsellors who all have key roles in their implementation. After the major political changes in 2000, Serbian educational authorities began to design and implement policy initiatives intending to contribute to economic revival, democratic development and alignment with policies of other European countries. The authors point to several barriers in the implementation processes such as a general devaluation of the work of principals, teachers and counsellors, political instability and frequent changes of policy direction, which can lead to reform fatigue, a lack of support and resources and delegated authority. While preparing to introduce reforms, they argue authorities failed to create a common ground, which has led to a mismatch of expectations at different levels of the education system. Despite reform shortcomings, however, there seems to be a general positive attitude that the initiatives represent potential improvements. Based on these findings, Teodorović et al. (2016) provide recommendations to policy makers regarding how to improve various phases of reform work. They address the importance of educators' perceptions and values in policy implementation focus on consequences if reforms are introduced too fast without creating a common ground.

Teodorović et al. (2016) argue that educators' perspectives on policy, its implementation and accountability practices are crucial. This is why attention needs to be paid to the design of policy and participatory processes in order to establish a comprehensive understanding and positive assessment by stakeholders. Important factors in the process of assessment include the perceptions of current challenges, the quality of the policy ('doing the right thing'), the design of policy implementation ('doing the right thing rightly'), the anticipation of possible policy impact (for the profession in general and for their own practices in particular), the investment-benefit relation, participation in policy development and the locus of control in policy implementation, competences and resources for implementation and improving practices, prior experiences and attitudes – not the least accompanying accountability practices in terms of how their work is followed up. The thesis is that a comprehensive understanding influences acceptance, which influences motivation, which again influences implementation activities.

Furthermore, motivation is also influenced by competences needed for implementation (in terms of knowledge, abilities and skills) and attitudes, as well as site-based management (with planning, coordination of actions, support, provided resources etc.). If such subjective theories are not considered, the processes of educators' sense making of the policy can be negatively influenced, and there is the danger that the motivation for policy implementation will decrease. Professional judgement is needed with respect to interpreting and mediating new policies and accountability expectations. The authors argue furthermore that policy and accountability need to be based on professional knowledge and expertise (from practice and research) (Teodorović et al. 2016).

CHAPTER II – METHODOLOGY

This second chapter focuses on the methodology used in this investigation. The design of the study is the survey method. The survey method is the base of the empirical study. Research design is one in which a group of people or items is studied by collecting and analyzing data from only a few people or items considered to be representative of the entire group. It specifies how such data will be collected and analyzed. This method was chosen for data collection, because it enables the researcher to solicit for information that might not be available on the pages of the text books and also to bring successful completion of the study.

2.1. The nature of research

The nature of the investigation was simultaneously a quantitative and a qualitative research. A case study of the public higher education system in East Timor, namely the National University of East Timor (UNTL, the acronym of Universidade Nacional de Timor Lorosae), was also conducted.

According to Herdeiro (2013), scientific research is considered by many experts (such as Best, 1982; Bryman, 1988; Ketele & Roegiers, 1999; Tuckman, 2002) as “a formal, systematic and intensive process of carrying out a scientific method of analysis” (p. 97). According to Tuckman (2002), “it is a systematic attempt to assign answers to questions” formulated that can be abstract and general or highly concrete and specific (p.5). The same author added that the scientific method is a fundamental characteristic of scientific investigation and its application results in a body of organized knowledge, constituting what is called science.

The subject of this investigative activity is called investigator and his objective will be to develop the different tasks necessary to reach a new knowledge. However, this requires an investigative-based attitude, essentially on a prepared look to analyze each data collected in relation to the knowledge accumulated by other scholars, allowing more (in)formation (Herdeiro, 2013, p.97). Beyond this quality, others are identified by the aforementioned author when he states that owning and cultivating a free spirit, a creative mindset open to all possibilities, as well as the presence of valuable habits and capacities are fundamental to convert a researcher into a true researcher.

Based on the knowledge of a specialized literature on research methodology (Estrela, 2004; Tuckman, 1994; Ozga, 2000; Churton & Brown, 2010; Lüdke, & Andrew, 1986; Johnson & Onwuegbuzie, 2004; Esteves, 2006; Moreira, 2006; Erasmie, and Lima, 1989; Ketele, & Roegiers, 1993; Bardin, 2007; Bogdan, Biklen, 1994; Goodson 2015, Sousa, 1997), a first study was conducted through documentary analysis focused on the legislation of the Timorese education

system in general and decision-making processes within the UNTL in particular, taking into account the benchmarks of other empirical studies conducted in other higher education contexts. The study also collected secondary data from various official documents and other sources.

Higher education institutions engage in advancing good teaching for two main reasons. First, they are interested in demonstrating that they are reliable providers of quality education, while serving multiple stakeholders with different expectations (like students, families, employers and policy makers). Second, they are required to respond to the increasing demand for meaningful and relevant teaching.

In order to describe how the UNTL teachers' competences are perceived by students, a questionnaire survey was administered to a 342 UNTL students from the five UNTL's departments. A separate questionnaire survey was also administered to 192 UNTL teachers in the same faculty over a period of 3 months (from April to June 2016) in order to know teacher's self-evaluation of their competences.

At the least, a study with a qualitative nature was carried out through in-depth interviews to both responsible persons for the management of UNTL, as the academics of the national public system. Respondents were the UNTL's teachers and lecturers, particularly those who are in management positions and higher-level management, middle managers and lower level managers of UNTL.

The research project was carried out starting with a problem statement and research questions that need a lot of quantitative and qualitative methodology. Research on professional development culture requires a multiplicity of techniques and instruments for data collection, as well as mixed methods for data analysis, in order to understand the comprehensive situation of the area of study. Along this line, the researcher adopted the research model by Albano & Estrela (2004), with the identification of three steps: structure, outlook and organization. The structure consists of files of data collection through the documents are indispensable for the characterization of the research problem. The researcher analyzed the documents used in East Timor in two professional contexts to refer to the context of evaluation and professional development culture. An empirical study was conducted, both qualitative and quantitative in nature, based on a document body (putting rules on the assessment of professional development). The phase of perspectives was implemented by carrying out empirical study on teacher performance evaluation and professional development practices in the context of UNTL.

The distinction between qualitative and quantitative research is often framed in terms of using words (qualitative) rather than numbers (quantitative) or using closed-ended questions (quantitative

hypotheses) rather than open-ended questions (qualitative interview questions). A more complete way to view the gradations of differences between them is in the basic philosophical assumptions researchers bring to the study, the types of research strategies used in the research (e.g., quantitative experiments or qualitative case studies), and the specific methods employed in conducting these strategies (e.g., collecting data quantitatively on instruments versus collecting qualitative data through observing a setting) (Creswell, 2013).

Moreover, there is a historical evolution to both approaches, with the quantitative approaches dominating the forms of research in the social sciences from the late 19th century up until the mid-20th century. During the latter half of the 20th century, interest in qualitative research increased and, along with it, the development of mixed methods research. With this background, it should prove helpful to view definitions of these three key terms as used in this research (Creswell, 2013). According to Creswell (2013, p.145) “quantitative research is an approach 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. The final written report has a set structure consisting of introduction, literature and theory, methods, results and discussion.

The same author (Creswell, 2013, p.173) stated that “qualitative research is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem”. The process of research involves emerging questions and procedures, data typically collected in the participant’s setting, data analysis inductively building from particular to general themes and the researcher making interpretations of the meaning of the data. The final written report has a flexible structure. Those who engage in this form of inquiry support a way of looking a research that honors an inductive style, a focus on an individual meaning and the importance of rendering the complexity of a situation. Like quantitative researchers, those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings.

However, Newman and Benz (1998) defend that qualitative and quantitative approaches should not be viewed as rigid, distinct categories, polar opposites, or dichotomies. Instead, they represent different ends on a continuum. A study tends to be more qualitative than quantitative, or vice versa. Mixed methods research resides in the middle of this continuum because it incorporates elements of both qualitative and quantitative approaches. To Creswell (2013, p. 203)” mixed methods research is an approach to inquiry involving collecting both quantitative and qualitative data,

integrating the two forms of data, and using distinct designs that may involve philosophical assumptions and theoretical frameworks.” The core assumption of this form of inquiry is that the combination of qualitative and quantitative approaches provides a more complete understanding of a research problem than either approach alone.

2.2. The research design

The research design is a case study of National University of East Timor (UNTL) in order to study the Timorese higher education system. Our research object is the evaluation of UNTL teacher competences: the pedagogical, professional, personal and social competences. Higher education institutions engage in advancing good teaching for two main reasons. First, they are interested in demonstrating that they are reliable providers of quality education, while serving multiple stakeholders with different expectations (like students, families, employers and policy makers). Second, they are required to respond to the increasing demand for meaningful and relevant teaching.

In order to describe how do the UNTL teacher’s competences are perceived and evaluated by students, a 28-item questionnaire survey was administered to 342 UNTL students regarding four teaching competences referring to pedagogical (9 items), professional (8 items), personal (6 items) and social competences (5 items). The researcher intended to identify the highest evaluated competences and the lowest competences that need improvements according to student’s evaluation. The relationship between the student’s evaluation and students’ background characteristics (like gender, age and faculty) were also examined.

A second quantitative study was likewise focused on a 28-item questionnaire survey administered to 192 UNTL teachers and was conducted over a period of three months (from April to June 2016). The intent of this survey was to describe the self-perception and the self-evaluation of UNTL teacher’s competences. There were items that identify the higher self-evaluated competences and the lowest self-evaluated that need improvements among the pedagogical, professional, personal and social competences. Additionally, it was intended to compare student’s evaluation and the teachers’ self-evaluation in order to identify divergences and / or convergences between the two actors, aiming at identifying the competences with similar and different evaluation among UNTL students and teachers.

A qualitative approach was also conducted to examine the implementation of teacher performance evaluation in UNTL, through a semi-structured in-depth interview questions to 16 teachers who are at the time of the study holding management positions at UNTL. This interview intended to obtain

the perceptions and attitudes of the teachers and how the Teaching Performance Evaluation (TPE) is being implemented in the educational institution, in the UNTL to which the respondents belong, and ultimately draw out their suggestions and recommendations on how to better implement teacher performance evaluation in Timorese higher education system.

Table 4 synthesizes the research design framework of quantitative and qualitative study.

Table 4. Research design framework of quantitative and qualitative study

Research questions	Objectives	Empirical data	Theoretical framework
<i>1. How do students evaluate the UNTL teachers' competences and how do students' background characteristics (like gender, age and faculty) influence their evaluation?</i>	1. To describe the students' evaluation of UNTL teachers' competences. 1.1. To identify the highest scores of teacher's competences evaluated by students. 1.2. To identify the lowest scores of teacher's competences evaluated by the students. 1.3. To analyze if the student's gender, age and faculty influence their evaluation they made on UNTL teachers' competences.	Students questionnaire (quantitative study)	Alhija (2017) Al-Mohaimeed and Khan (2014)
<i>2. How do UNTL teachers self-evaluate their competences and how does their professional background (gender, age, faculty and degree) influence their self-evaluation?</i>	2. To describe the self-evaluation of UNTL's teachers' own performance. 2.1. To identify the highest score of the self-evaluated UNTL teacher's competences. 2.2. To identify the lowest score of the self-evaluated UNTL teachers' competences. 2.3. To analyze if the age, length of service, academic degree and department of the UNTL teachers influence their self-evaluation.	Teachers questionnaire (quantitative study)	Suknaisitha, Suwimon & Piromsombatc (2014) Van Schalkwyk, Herman, Leibowitz & Farmer (2015) Nevgi & Lofstrom (2015) Kozikoğlu (2017)
<i>3. To what extent do students' evaluation of UNTL teachers and the UNTL teachers' self-evaluation competences differ or converge?</i>	3. To compare student's evaluation of UNTL teachers' competences and the teachers' self-evaluation, in order to identify divergences and / or convergences. 3.1. To identify which teachers' competences are different between student's evaluation and teachers' self-evaluation. 3.2. To identify which teachers' competences are similar according to student's evaluation and teachers' self-evaluation.	Students and teachers questionnaire (quantitative study)	Smith, Gamlem and Engelsen (2017) Jónsson, Smith and Geirsdóttir (2018)
<i>4. How does the Teaching Performance Assessment (TPE) being implemented in the educational institution to which the respondents belong?</i>	4. To know how the TPE is being implemented in the educational institution to which the respondents belong. 4.1. To identify the attitudes towards the teacher performance evaluation in Timorese higher education system: straightforward and/ or the obstacles. 4.2. To know the teacher's experiences in the implementation of teacher performance evaluation in Timorese higher education system 4.3. To identify the teaching and learning competences should be developed by teachers in a professional development program. 4.4. To draw out suggestions and recommendations from UNTL teachers in relation to effective implementation of teacher performance evaluation in higher education system.	Qualitative study: in-depth interview to 16 UNTL teachers who are at the same time holding management positions at UNTL.	Ehren and Shackleton (2016) Teodorović, Stanković, Bodroža, Milin & Đerić (2016)

Table 5 presents the hypotheses of investigation established regarding previous studies about students' evaluation and teacher's self-evaluation of teachers' competences.

Table 5. Hypothesis and the theoretical framework of quantitative study

Research questions	Hypotheses	Theoretical framework:
1. How do students evaluate the UNTL teachers' competences and how do students' background characteristics (like gender, age and faculty) influence their evaluation?	<p>H1: The personal and the social competences were the highest valued by students, rather than the pedagogical and professional competences.</p> <p>H2: There is a relationship between students' evaluation on teacher's competences and students' background characteristics:</p> <p>H2a: The female students attribute a more positive evaluation to teachers' competences than male;</p> <p>H2b: The older students give a more positive evaluation to teachers' pedagogical and professional competences than the younger students;</p> <p>H2c: Students from humanities attribute a greater evaluation to personal competences, while students in natural and exact sciences attribute a greater evaluation to pedagogical competences.</p>	<p>Alhija (2017) Al-Mohaimed and Khan (2014)</p> <p>Alhija (2017); Anderson et al. (2012); Lavin, Korte & Davies (2012); Korte, Lavin & Davies (2013)</p> <p>Alhija (2017)</p> <p>Alhija (2017) Al-Mohaimed and Khan (2014)</p>
2. How do UNTL teachers self-evaluate their competences and how do their professional background (age, time of service, academic degree and faculty) influence their self-evaluation?	<p>H3: Teachers rate their own pedagogical and professional competences lower than their own personal and social competences.</p> <p>H4: UNTL teachers' with more age and years of experience had higher self-evaluation than younger and least experienced teachers.</p> <p>H5: Teachers with a higher academic degree have the higher self-rating on their own competences.</p> <p>H6: UNTL teachers' from political science department highest self-evaluated their competences.</p>	<p>Suknaisitha, Suwimon & Piromsombatc (2014)</p> <p>Van Schalkwyk, Herman, Leibowitz & Farmer (2015)</p> <p>Nevgi & Lofstrom (2015)</p> <p>Kozikoğlu (2017)</p>

2.3. The case study of UNTL

Yin (2009) refers to the case study as a useful methodological option when analyzing a broad and complex phenomenon such as the professional development of professors in the Timorese public higher education system.

The certification of university teachers in Timor-Leste, which is compulsory, provides for the implementation of external evaluation mechanisms (course documentation, peer evaluation, hierarchical and student evaluation) and internal evaluation mechanisms (guidelines of the

instrument description of teaching staff, curriculum and self-assessment) that make it possible to gauge the quality of teaching performance and, consequently, the service provided by the teacher¹. In addition to assessing the quality of the higher education system, these mechanisms aim to regulate the access, progression and development of the university teaching career².

According to the University Teachers' Certification Manual (CEDU) in East Timor, teacher certification consists of the formal recognition of their teaching competences, defined as a set of characteristics, knowledge and competences that they must possess and demonstrate in the activity, in terms of pedagogy, professional, social and personality.

The evaluation of *pedagogical competences* focuses on three areas: the ability to design teaching and learning; the ability to implement and evaluate the teaching and learning process and its outcomes; the ability to use research results to improve the quality of teaching and learning. The assessment of *professional competences* was based on increasing the integrated capacity of knowledge on a given scientific area, on the field of knowledge application techniques and also on positive behaviors such as innovation. In relation to the evaluation of *personal competences* we refer to the values, behaviors and professional ethics of the teacher with an impact on the students, relatives, family and society, which can influence the motivation of the students for the study, as well as their personal development.

According to the CEDU Manual, teachers' competences are crucial for the quality and implementation of the three pillars of higher education: teaching/pedagogy, research and community service. The same document (approved by Ministerial Diploma No. 33 / ME / 2014, of 10 September) incorporates the evaluation carried out by the students to assess the adequacy of the teachers' competences and the quality of their work and, in this sense, questionnaire addressed to students of the Timorese public higher education, in order to be evaluate the competences of their respective university professors.

And since university teachers are one of the essential components of the Higher Education System, their functions, duties and responsibilities are crucial for the achievement of national educational policy objectives and for ensuring greater qualification of the population in various fields (technical, scientific, technological, artistic, civic, religious, etc.). Some of the strategies of the National

¹ The University Teaching Certification (CEDU) aims to: dignify the teacher as an educator agent of higher education; evaluate the professionalism of the teacher to determine if the teacher is qualified to perform his duties; raise the quality of educational processes and outcomes; accelerate the achievement of national education objectives; and to educate the teacher to perform his duties with honesty and academic ethics.

² At first, the University Teaching Career Statute (ECDU) establishes that "the entry into the university teaching career determines the need for University Teacher Certification (CEDU), carried out through the application of a system of accumulation and weighting of credits and evaluation of the performance, and which constitute obligatory conditions for career progression "(Article 31 (1) of Decree Law No. 3/2014, of January 15).

Strategic Plan for Education are precisely aimed at promoting the quality of education and improving student learning outcomes, focusing in particular on the work of teachers (Metl, 2011). The purpose of this document is to: increase the quality of teacher education, including a flexible and modular system of credits linked to the principles of lifelong learning; The development and implementation of a "measurement system" to monitor and evaluate the impact of teacher training on improving the quality of education; the strengthening of institutions dedicated to teacher training; the promotion of more flexible training courses through multiple specialization; the training of non-formal education teachers; the implementation of teacher surveys to assess competencies and performance, tools that would help to define teachers' in-service training needs and provide indicators for a fairer and more efficient career management of teachers as well as the incentive scheme (Ibid, pp. 80-163).

In order to promote priority development in the country, the National Strategic Plan for Education 2011-2030 also defined the implementation of a quality assurance system through: (i) registration of all courses in the National Qualifications Framework and ii) the continuous development of the National Agency for Academic Evaluation and Accreditation (ANAAA), within the national quality assurance framework, with the responsibility of determining the standards and criteria for quality assurance of all courses. The development of this legal framework will make it possible to improve the performance of higher education in terms of teaching and learning as well as research and development, which represents a fundamental step towards the development of competences, competences and social and professional recognition of university teachers (Metl, 2011)³.

The National Development Plan of Timor-Leste (PDN-TL) emphasizes the importance of human and social development as the key "for the development of the Nation, reducing poverty, promoting economic growth and improving the living conditions of (PDN-TL, 2002) in order to achieve the aspirations of the East Timorese expressed in "Vision 2020", particularly in the so-called priority sectors for development.

Over the past five years, therefore, many measures have been implemented by the Timorese Government to introduce higher quality into the entire higher education system. The measures implemented aimed at regulating access to and progression of the teaching career, defined the

³ The strategic document also provided for the development of an efficient management system to coordinate government interventions in higher education and to set priority targets and budgets, as well as the creation of new higher education units (technical or university) in the areas of economics that need qualified human capital (METL, 2011).

strategic objectives set for the sector, in line with national development objectives, and defined the criteria on which the evaluation of the quality of teaching work should be based.

In 2015, the results of the Pedagogical Survey to the implementation of the 2014 Curriculum in UNTL⁴ were published, under the premise that there is an intimate relationship between the didactic performance of the teacher and the performance of the student. The report aimed to evaluate the quality of content and curricular units, teachers and students. The dimensions of the survey concerning student opinion presented a positive overall assessment, with the rating between (3) "adequate or sufficient" and (4) "moderately high" compared to the evaluation of the implementation of the new 2014 curriculum of UNTL. This Pedagogical Survey to the implementation of the 2014 Curriculum in UNTL concluded about a positive perception of the competences of the teachers on the part of the students, in line with previous studies of UNESCO (2013) carried out in the scope of the quality of the public higher education, with regard to the defined objectives and goals reached for the public higher education of East Timor, as well as the pedagogical evaluation of a new curriculum (UNTIL, 2015). However, data from students and discussions during the work sessions found some distortions in school success rates in some courses (refer the courses) and warned the need to minimize one of the causes detected, namely the poor pedagogical performance of a significant number of university teachers (UNTIL, 2015, 45). Taking into account the current evaluation framework of public higher education and following the measures implemented at this level of education, the present study arises with the purpose of evaluating the current perception of the students about the competences of their teachers, so that from here we can collect indicators on the students' perception of their teacher performance at UNTL, as well as obtain information about teachers evaluating their own competences.

2.4. Sampling and respondents

The researcher used simple random sampling technique method to select a representative number of participants from the population of the study, with an equal probability to be selected. This method of sampling also allows reducing bias. The criterion of the inclusion consisted in select students and teachers from different technical-scientific areas.

⁴ From the beginning of 2014, the UNTL implemented a new curriculum contextualizing the knowledge, skills and abilities according to the profile of the trainees in face of the scientific and professional demands, and the labor market at national, regional and international level. This new curriculum redefined and readjusted the general objectives of the courses of the University's training offer, the profile of the graduates, the skills and the professional skills, in the light of the challenges of the rapid changes in society, the labor market and the conditions of the professional exercise. Scientific areas were redefined to obtain each degree or diploma, basic, professional and specialization contents with curricular units and their respective credits and workloads redefined according to internationally accepted standards and accounted for with the ECTS system (*European Credit and Transfer System*).

Pertaining to the population size, the number of participants required in a representative sample was scientifically determined according to the following formula:

$$n = \frac{N}{1 + N (e)^2}$$

Where, n = sample size
 N = population
 e = level of significance
1 = Constant

The researcher considered the 5% level of significance to determine the sample size.

In relation to the student's evaluation on UNTL teachers' competences, there were 1200 students considered for the total population of the study, corresponding to the total number of students across the four selected faculties of UNTL studied (Faculty of Philosophy, Faculty of Education, Arts and Humanities, Faculty of Medicine and Faculty of Political Science).

Using the sample formula, the following was obtained:

$$n = \frac{1200}{1 + 1200 (0,05)^2}$$

$$n = \frac{1200}{1 + 1200 (0,0025)}$$

$$n = 300$$

The recommended sample size in students' evaluation is 300 participants. The collected sample was composed of 342 students (above the minimum required) attending the four faculties of UNTL (Faculty of Philosophy, Faculty of Education Arts and Humanities, Faculty of Medicine and Faculty of Political Science). Characterization of collected sample is presented in table 6.

Table 6. Characterization of students sample (N=342)

		n	Valid Percentage (%)
Gender	Male	163	47.7
	Female	179	52.3
	Total	342	100.0
Age	18 – 24 Years	300	87.7
	25 – 31 Years	40	11.7
	32 – 45 Years	2	.6
	Total	342	100.0
Faculties	Philosophy	87	25.4
	Education, Arts and Humanities	99	28.9
	Medical Science	101	29.5
	Political Science	55	16.1
	Total	342	100.0

The students sample consisted of 52.3% female students (n=179) and 47.7 % male students (n=163). The most frequent age group was 18 to 24 years old (87.7 %; n=300) and 25 to 31 years old (11.7 %; n=40); two respondents were between 32 and 45 years (0.6 %; n = 2). The Faculty of Medicine was the most frequent (29.5 %, n=101), followed by the Faculty of Education, Arts and Humanities (28.9%, n=99), the Faculty of Philosophy 25.4 %, n=87) and, finally, the Faculty of Political Science (16.1 %, n=55).

In relation to teacher self-evaluation study, the studied population corresponds to the total of 400 professors of all faculties' UNTL. Using the formula, the following sample dimension was obtained:

$$n = \frac{400}{1 + 400 (0,05)^2}$$

$$n = \frac{400}{1 + 400 (0,0025)}$$

$$n = 200$$

The recommended sample size on self-evaluation study was 200 participants. The final collected sample was constituted by 192 professors of UNTL, slightly lower than recommended but was already sufficient.

Characterization of collected students sample is presented in table 7.

Table 7. Characterization of teachers sample (N=192)

		n	Valid Percentage (%)
Gender	Male	152	20.8
	Female	40	79.2
	Total	192	100.0
Age	< 35 Years	21	10.9
	[35 – 45[Years	70	36.5
	[45 – 55[Years	94	49.0
	≥ 55 Years	7	8.6
	Total	192	100.0
Faculties	Education, Arts and Humanities	30	15.6
	Political Science	75	39.1
	Management/Economy	23	12.0
	Agriculture	38	19.8
	Medical	23	12.0
	Philosophy	3	1.6
	Total	192	100.0
Academic degree	Bachelors	40	20.8
	Masters	138	71.9
	PhD	11	5.7
	Postgraduate	3	1.6
	Total	192	100.0
Years of working	< 5 Years	21	10.9
	[5 – 10] Years	90	46.9
	[11 – 20] Years	81	42.2
	Total	192	100.0

The UNTL teacher's sample consisted of 79.2% (n = 152) male and 20.8% (n = 40) female. A sample of 192 teachers (N = 192) from six of the nine faculties of UNTL were obtained: 39.1% (n = 75) of the teachers were from the Faculty of Political Science; 19.8% (n = 38) of the Faculty of Agriculture; 15.6% (n = 30) of the Faculty of Arts and Humanities; 12% (n = 23) of the Faculty of Management and Economics and, equal number, of the Faculty of Medicine; 1.6% (n = 3) of the Faculty of Philosophy. The most frequent age group was 45 to 55 years (49%, n = 94) and 35 to 45 years (36.5%, n = 70); 10.9% (n = 21) of teachers were less than 35 years of age and 3.6% (n = 7) over 55 years. The majority of teachers had a Master's degree (71.9%; n = 138), while 20.8% (n = 40) had a Bachelor's degree; 5.7% (n = 11) had a PhD degree and 1.6% (n = 3) had a Postgraduate Degree. The highest percentage of respondents had between 5 and 10 years of experience (46.9%, n = 90) and 11 to 20 years (42.2%, n = 81); only 10.9% (n = 21) of teachers had less than 5 years of teaching experience.

2.5. Instruments of data collection

As stated by Morgado (2016) the techniques and data collection tools used are essential because they depend on the quality and success of the research. They should therefore be drawn up and used in such a way as to capture as fully as possible all the information inherent in the phenomenon being studied and can therefore not fail to take into account the objectives pursued and the context in which the study is carried out. Furthermore, Morgado (2016) considers the collection of information as an "organized process put into practice to obtain information from multiple sources, in order to move from one level of knowledge to another level of knowledge or representation of a given situation, within the framework of a deliberate action whose objectives have been clearly defined and of the guarantees of validity" (p.71).

The same author assures that the collection of data in the field of study is only intended to make it a task and make it "as intuitive - or as radically intuitive as possible", which the work of the field bears a certain mysticism very proper and is, therefore, "impossible (...) to be taught" (Morgado, 2016, p. 71). The greater or lesser success in the adequacy of the data collection process for the subject of research depends on the knowledge, capacity and experience of each researcher.

Hence Morgado (2016) argues that at the beginning of this stage of research, the researcher has a clear idea of the research guiding questions and data collection procedures that can be used to answer these questions, as well as a clear notion of the importance of establishing a cooperative relationship with the study participants. Once the information to be collected has been identified, it is necessary to define the strategy that would allow this attempt to be achieved, which implies knowing the techniques and the collection instruments to be used. In this segment of the text we will discuss the techniques and instruments most used in the context of the case study, particularly in the field of educational research: the document analysis and observation; the questionnaire survey and the interview.

The researcher utilized two effective methods of data collection: the questionnaire and the oral interview. The researcher collected the data from the respondent's one on one. The instrument used for the quantitative data collection was the questionnaire and the instrument used for the qualitative research was the interview.

2.5.1 Documental analysis

Consultation and document analysis is another important source of information in the context of research, particularly in education research. Saint-Georges (1997) ensures documentary analysis is "a method of data collection and verification" that allows access to relevant sources, making it

"an integral part of the research heuristics" (p.30). Although De Ketele and Roegiers (1996) consider the study of documents to be fundamental and depends always on the nature of the documents to be analyzed, whether written or not, official or unofficial, public or private, individual or collective, open or closed, educational, scientific, etc., of the number of documents to analyze - which influences and determines the greater or lesser completeness of the study - and of the object and purposes of the research itself - with a more exploratory or more confirmatory character.

In a similar vein, Stake (1999) considers such documents to be an asset in any investigative process, functioning as "substitutes for activity records that the investigator cannot directly observe" (p.66).

To Reis (2010) documentary analysis as a data collection technique allows the collection of information that fits facts and responses from other sources. He also explains that his appeal can be made in two perspectives - as a complement to information obtained from other sources and as a central source of analysis. In this case, in analytical documentary research will follow as two perspectives, since there is no manual as source but also, in the analysis of semi-structured interviews, with the purpose of complementing and crossing as information.

The documental analysis' target in this investigation was governmental, institutional studies and official statistics about Timorese educational system, in general, and higher educational institutions, in particular. The Timorese educational legislation and the national Constitution were also considered. The University Teachers' Certification Manual (CEDU) was the main document consulted which contained standardized teachers' competences.

2.5.2 Questionnaire

The questionnaire was designed for the purpose of evaluating the UNTL students' perceptions and evaluation about their teachers' competences. Another questionnaire was designed to collect the UNTL teachers self-evaluation' competences.

Survey by questionnaire is a data collection technique widely used in quantitative research. A questionnaire is an orderly and coherent series of questions that are posed to a group of respondents to gather elements about their social, professional or family situation, their opinions, their attitudes and / or the way they stand in the right direction on human and social issues, events, or problems, their expectations, their level of knowledge, and any topic or subject of interest to the researcher (Quivy & Campenhoudt, 1998, p.188).

Given its applicability and its rapid coverage of respondents becomes frequent use of this medium as underlined by Fortin (1996): "the questionnaire helps organize, normalize and control data, so

that the information sought to be harvested from a rigorous" (p. 249). The researcher chose to use the questionnaire since he was interested in studying the attitudes, beliefs, behaviors, opinions, employment status and characteristics of teachers who may be affected by the proposed performance evaluation system. The questionnaire was designed based on the research design and objectives of the study. In this way, content was developed using the indicators established with closed questions. The content of the questionnaire was based on key variables in performance evaluation study supported by theories of performance evaluation in order to collect information relevant to the analysis of this issue.

Given that the majority of respondents have difficulty with the English language, the questionnaire was translated into Tetum and Portuguese language to facilitate understanding. As Fortin stressed (1996), "even if there is a questionnaire widely depending on the intended purpose, it is often necessary to translate it into the language of the country and adapt it to the new application context." (p. 250). To address these questionnaire difficulties, the questionnaires were produced in three versions: first, an English version was completed by respondents who had an average and higher level of understanding in English; second, questionnaire was translated to Tetum and was distributed among the population studied with low-level understanding of the Portuguese language; third, a Portuguese version was distributed to those who understand Portuguese language.

In order to evaluate the pedagogical, professional, social and personality competences of the teachers, a paper survey questionnaire was carried out in paper format, as proposed in the CEDU Manual and with which the students evaluated the respective university professors, through a set of: nine (9) items focused on the pedagogical competences of teachers; eight (8) items relating to professional competences; six (6) items that evaluated personality competences; five (5) items that evaluated social competences. The items of each set of competences were classified (by the students and by teachers too) according to a Likert scale of 5 points, in which 1 corresponded to "very bad", 2 "bad", 3 "sufficient", 4 "good", and 5 "very good", given an opportunity for respondents to choose the answers according to their opinions and situation.

The questionnaires were administered to the sampled population (teachers and the students at each selected faculty of UNTL) and collected in the same manner. The relevant variables of the study were effectively measured by the items contained in the questionnaire. The questionnaire's form has an introduction presenting the context and, the objective of the investigation, guaranteed anonymity and confidentiality of responses and strict compliance with the ethical and moral procedures.

The questionnaires were personally distributed and collected by the researcher. The survey was conducted on April 2016 by distributing it directly to the teachers; however, most of them were almost unwilling to answer the questionnaire, then it took time almost 3 months to answer the questionnaire because they were waiting for the official dispatch of the Rector of UNTL as a permission to conduct the research and the dispatch was only delivered on June 2016. Therefore, the questionnaires were only collected during the entire month of June.

2.5.3 Interview

One of the most used techniques to obtain and to collect qualitative information in investigation is the interview. It is a technique used to collect data whose main objective is to understand the meanings that the interviewees attribute to certain issues and/or situations. Villar Angulo (1997) assures that the interview, by allowing to uncover certain perceptions implicit in the interviewees' thoughts, contributes to "understanding their conceptions of reality and the meaning and meaning they attribute to their actions" (p.23).

Bisquerra (1989) defines it as "a conversation between two persons initiated by the writer, for the specific purpose of obtaining information relevant to an investigation" (p.103). In a broader context, De Ketele and Roegiers (1999) view the interview as "a method of collecting information consisting of oral, individual or group conversation, with several carefully selected people, in order to obtain information about facts or representations, whose degree of relevance, validity and reliability is analyzed from the perspective of the collection of information" (p.22).

Although, in the common sense of the term, the word interview assumes a non-restrictive and not univocal meaning, there is a common denominator for the various types of interviews - the interviewer asks the questions and the subject(s) of the research provides the answers. In either case, it is important to remember both the guide of the interview and the guidelines that the interviewer intends to print in this process must be in line with the object of the study and with the theoretical framework that supported it.

For this study a semi-structured interview schedule was developed and the formulation of questions in the schedule was based on the research questions, guided by what emerged from a study of the relevant literature. This is to enable the interviewees to express their opinions without bias on the teachers' professional competences. The oral interview has the advantages of giving more in-depth information about the questions asked. The participants for this interview were the influential head departments, faculties and higher level management key positions at UNTL who are also have teaching loads. The biographical data and demographic data of the participants were included in

the interview guidelines.

The researcher recorded all the interviews on his smartphone audio recorder and took careful notes to ensure no information is lost. In addition to the request for authorization to record the interview, at the beginning of the interview, the independence of the investigation was guaranteed, the objectives of the study presented, and the confidentiality of the replies guaranteed.

The interviews were collected over a period of one month, between 27th of June 2016 and 27th of July 2016, ensuring compliance with the ethical principles of privacy, anonymity, confidentiality and the protection of the personal data of study participants, as well as fair and equitable treatment of the information provided.

The data obtained from the interviews were transcribed, coded and categorized according to the professional competences in the study, thus allowing an inductive analysis. Both manifest and latent content was coded and grouped into several central themes. Evidence on the results of the interview could be found at the data presentation and the discussion of the findings section in the next chapter within this thesis.

To Berelson (1984) the qualitative method of content analysis proved to be the most oriented to the "objective, systematic and quantitative description of the manifest content of the communications and for the purpose of interpretation" (p.18). In the same sense, Bardin (2015) points out that "all the initiatives that, based on a set of partial but complementary techniques, consist in the explication and systematization of the content of the messages and the expression of this message belong to the field of content analysis, with the contribution of quantifiable or non-quantifiable indices" (p.44). The analysis of content allowed to evaluate the observed information intensively, through the application of qualitative procedures, corresponding to the analysis of a small number of complex and detailed information in order to identify "the presence or absence of a characteristic or the way in which the elements of discourse are articulated with one another" (Quivy & Campenhoudt, 2008, p. 227).

2.6. Data analysis procedures

In data analysis there are two tasks that, according to Erickson (1989) are essential for the development of this process: (i) generate affirmations through induction, which requires a detailed examination of all data corpus, and (ii) "establish a basis of evidence for the statements that are desired to be made", which forces to review, repeatedly, the set of data to "verify the validity of the statement that has been generated, while seeking arguments for and against" (p.262).

Because it is one of the crucial stages of all research, data analysis should not be disintegrated

from its inherent interpretation. In this way, it becomes a task that can make sense of what we observe and help us understand and relate the data collected during the investigation process.

In the present investigation, statistical techniques were implemented to analyze the quantitative data obtained from the questionnaires, and the content analysis was conducted to obtain the meaning of the information collected through the interviews.

2.6.1 Statistical analysis

Moreira (2006) affirms that the analysis of quantitative data, whose general objective is to confirm explicit ideas or hypotheses that the researcher has a priori (and is spoken of in these cases of a confirmatory analysis), or rather to try to use these data to obtain the said ideas and hypotheses (speaking of exploratory analysis), has always one of two specific objectives: a) to describe the distribution of the entities by the different values of the variables or b) to describe the relation between the variables (p.51).

The case study is a methodological strategy most used in qualitative investigations, which does not prevent specific situations in case studies that use a mixed or even quantitative methodology. In addition, if statistical analysis is always associated with quantitative investigations, it is also true that "statistics may be appropriate at certain stages of data analysis in qualitative research" (Coutinho, 2005, p. 139).

Some of the methods and procedures most used in the treatment and statistical data analysis, specifically when the items of the questionnaire are structured around measurable variables (quantitative variables) including the cases in which the answers involve the use of a measurement scale, such as, for example, a Likert type scale.

In this case of study, the researcher resorted only to the descriptive statistics procedures since they provided, as refers Coutinho (2015) "a first reading of the information contained in the data" and allow "to identify what is typical and atypical", to highlight differences and relationships between them and to "give indications about of the dispersion, shape and structure of its distribution" (p.140).

It is also important to mention two aspects that seem relevant in the selection and application of statistical analysis techniques. In the first place, the procedures to be used always depend on the quantity and nature of the data at our disposal. Second, and in consonance with Coutinho (2005), that the statistical concepts used in the analysis of the data of a study "has behind it complicated mathematical calculus" a complexity that, with the help of the computer, "freeing us time to focus on what is truly important" (p. 139) it is, in the logic of analysis and in the relationships that are

established between the data under analysis.

The statistical analysis was conducted with the software *Statistical Package for Social Science* (SPSS) version 24, from the IBM®, and also the *Microsoft Excel*®. There were implemented techniques of descriptive statistics which allow for the synthesis of the data and to describe the data distribution and variation.

The four competences were evaluated through a set of items: nine items related to pedagogical competences; eight items relating to professional competences; six items related to personal competences traits; five items related to social competences. Each item was measured according to a 5-point Likert scale, in which 1 corresponded to "very bad", 2 "bad", 3 "sufficient", 4 "good" and 5 "very good". One questionnaire was administered to students' sample and another to teachers' sample with different set of items (related to pedagogical, professional, personal and social competences) and the same 5-point Likert scale.

After the data were collected, the arithmetic mean of the classifications for each skill group was calculated. The descriptive statistics of each item and each competence evaluated was used to analyze the distribution of students' evaluation and teacher's self-evaluation. There were analyzed measures of central tendency (as the mean) and measures of dispersion (as the standard deviation). The dispersion measures are characterized by helping to understand to what extent the results center around the main tendency of a set of observations with a greater dispersion.

The standard deviation is a measure of dispersion that shows the greater and lesser degree to which the values of a given distribution move away from the media, safely expressing the degree of consensus among respondents (Bryman & Cramer, 1993). With regards consensus, this is operationally defined in terms of concentration of responses in one of the evaluative categories. For this purpose, the values of the standard deviation are used since they indicate the concentration (unanimity) or dispersion of the responses. If the value of the standard deviation is equal to zero there is no dispersion and, logically, the consensus is total. In order to assess the greater or lesser consensus of the respondents' responses, the variation of the standard deviation can be interpreted according to the following criteria (Table 8).

Table 8. Criteria for verifying the degree of consensus of the answers

Value of the standard deviation	Level of consensus
0.00 – 0.40	High consensus
0.41 – 0.70	Moderate/High consensus
0.71 – 1.00	Moderate/Low consensus
>1.00	Low consensus

"The arithmetic mean was obtained by adding the scores and dividing this sum by the number of them" (Coutinho, 2005, p.141) and in order to verify the distribution of agreement level of some items we use the following scale (Table 9).

Table 9. Degree of mean agreement

Mean	Level of agreement
1.00 – 2.75	Disagree
2.76 – 3.25	Indefiniteness
3.26 – 5.00	Agree

The internal consistency of each set of competences was analyzed through the Cronbach's alpha value. The strength of the correlation between the competences was analyzed through the Pearson correlation. Pearson's non-parametric test evaluates the intensity and strength of the correlation between quantitative variables (with values of r varying between -1 and +1) since the correlation between the variables is stronger as well as closer to 1 in absolute value. Positive correlation indicates that if one variable increases the other increases proportionally as well. Conversely, a coefficient of -1 indicates a perfect negative relation: if one variable increases, the other decreases by a proportional amount. A coefficient of zero indicates the absence of a linear relationship and therefore, if one variable changes, the other remains the same. The correlation between the continuous variables is stronger as well as closer to 1 and the test result of significance still indicates if the correlation between the variables is significant, for a confidence interval of 99% ($p < .01$) (Field, 2009, p.170).

In the second phase of the study, the continuous variables of the competency assessment were transformed into categorical variables: the first category ("very bad") integrated the evaluations in the range of [1-2]; The second category ("bad") integrated the evaluations in the range [2-3]; The third category ("sufficient") included evaluations in the range [3-4]; The fourth category ("good") included assessments in the range [4-5]; The fifth category ("very good") integrated the evaluations with values equal to or greater than 5.

Fisher's exact test was used to verify the statistical association between the students' evaluation on teacher competences and the variables gender, age group and students' faculty, in order to understand if these variables had an influence on teachers' evaluation. Fisher's exact test (resulting from the chi-square test) indicates whether the variables have a statistically significant relationship at a confidence level of 5% ($p < .05$) and which categories are likely to be statistically related, given the Value adjusted residuals (Adj. Res.) greater than 1.96. The null hypothesis of Fisher's exact test

(H0) indicates that the variables are independent; and alternative hypothesis indicates that the variables are not independent. The significant result of Fisher's test ($p < 0.05$) allows rejection of H0 and assuming a statistically significant association between categorical variables (Field, 2009, p.669).

The student's t-test and Anova test were used to compare the average of the competency assessments attributed between male students and female students. The null hypothesis of the parametric tests (H0) establishes that the means of groups are equal, while the alternative hypothesis (H1) assumes that the means of groups are different. The significant result of the t test ($p < 0.05$) allows rejection of H0 and decides in favor of the alternative hypothesis of significant differences between the two groups.

2.6.2 Content analysis

The analysis and interpretation of the content data are core tasks in the field of research, regardless of the methodology adopted. However, such tasks are of significant relevance in the field of qualitative research, where the data are multifaceted in nature and their collection is often concomitant with their interpretation. In any case, Estrela (2006) points what is wanted is to resort to methodological procedures that allow "to make intelligible to facts and social and human phenomena" (p.106), specifically to the domain of education. Content analysis plays a preponderant role in this regard.

As a central data-processing methodology, either as a process adopted only at a research stage or combined with other techniques and/or as a means to construct other instruments, the analysis of content marks today presence of most investigations defends Amado (2000).

With the expression *content analysis*, a set of techniques was identified to examine and make inferences about the meaning of information previously collected, and it is possible to apply to written texts, photographs, illustrations, radio programs and verbal interactions of all kinds, and in discourses as diverse as anthropology, psychology, sociology, psychiatry, history, literature, linguistics, among others (Bardin, 1995).

According to Esteves (2006, p.107), the data to be used in the analysis of content can be of diverse origin and nature, of which: (i) data invoked by researcher, that is "traces of phenomena that exist independently of their action", such as field notes, data obtained by direct observation, archival documents, normative documents, pieces of legislation, newspaper articles, books or biographies; and (ii) data generated by the researcher, that is, that depend on his activity, such as protocols for semi-structured and unstructured interviews, open answers in questionnaires, life histories,

journals, practice reports and portfolios.

Coelho (2013) defined content analysis as "a research technique that allows an objective, systematic and quantitative description of the manifest content of communications, aiming at its interpretation." (pp. 48-50). This concept of content analysis was still closely linked to the positivist paradigm and came to be modified later.

Bardin (2016) considers the content analysis as "a set of communication analysis techniques that uses systematic procedures and objectives to describe the content of the messages" (page 44).

Esteves (2006) considers content analysis as a method of the investigator to make inferences, "inferences that, because they are presented with explicit grounds, can be questioned by others, and can be corroborated or contradicted by other collection and treatment procedures of data ..." (p.108).

For Bardin (2016), content analysis is organized chronologically in three moments: the pre-analysis, the exploration of the material and finally the treatment of results, inference and interpretation. The pre-analysis is the first phase and consists of the collection of the material to be analyzed, constituting the documentary corpus, which should be related to the research objectives..." (pp. 33 – 48).

Once transcribed and validated, the first contact with interview material is "floating reading." These first readings allow, in a very "coarse" way, to sketch, even mentally, a first categorization, "(...) the investigator lets himself be impregnated by the nature of the discourses collected and by the general meanings contained therein in order to begin to glimpse the system of categories to be used for treatment." (Esteves, 2006, p.113).

The exploitation of the material corresponds to the codification and requires the definition of what is meant by unit of record, unit of context and unit of enumeration (Bardin, 2016). Cutting the registration units is a very delicate process, the clipping should have its own meaning, set up an idea that can consist of a phrase, a part of a sentence or several sentences whose meaning is completed among themselves, "(...) corresponds to the segment of content to be considered as the basic unit, aiming at categorization and frequency counting "(Bardin, 2016, p.39). It will then be up to the investigator whether or not to count such units of record. It is also up to the researcher to decide on an open or closed procedure and the type of categories that he intends to create, according to the research objectives. Carmo & Ferreira (1998) defines categories as "(...) significant headings, according to which the content will be quantified ..." (p.254). For Bardin (2016), the categories are also headings or classes that will bring together a group of elements, the

units of records. For the author the creation of categories should obey the following qualities: the mutual exclusion (an element only has fit in a single category); homogeneity (there is only one dimension of analysis in each category); relevance (taking into account the objectives of the study); objectivity and fidelity (the material analyzed must have a single codification, even if subjected to other analyzes) and productivity (the categories must provide inferences, new hypotheses and exact data).

The interviews were subject to a content analysis, which according to Bardin (2004, p.33), "appears as a set of communication analysis techniques, which uses systematic and objective procedures for describing message content." The same author emphasizes that the basis of the content analysis lies in the articulation between the surface of the texts (described and analyzed) and the factors that determined these characteristics, logically deduced. It also argues that "[t]he analyst's reading of the content of communications is not, or is not solely, a reading" read-to-read", but rather the highlighting of a meaning that is found in the second plan" (Bardin, 2004, p. 36).

Regarding the categories and subcategories, this author states that: categorization is an operation to classify constitutive elements of a set, by differentiation and, then, by regrouping according to gender (analogy), with the previously defined criteria. The categories are headings or classes, which bring together a group of elements (record units, in the case of content analysis) under a generic title, grouping them by the common characters of these elements (Bardin, 2004, p.111). Also according to Gibbs (2009), categories or concepts can originate from the literature investigated, previous studies, topics in the interview script, and perceptions about what is happening.

For Bardin (2004), the analysis of content as a method, is configured as a set of communication analysis techniques, which uses systematic procedures and objectives to describe the content of the messages. This definition of the author identifies some essential characteristics in the analysis of content, namely, to constitute a means to study the communications between the individuals, highlighting the content of the messages, privileging the written and oral language, without excluding other means of communication. The use of this method, especially on written messages, is that they are more stable and constitute an objective material which can be used where necessary.

In the present investigation, the content analysis started fundamentally from a closed procedure, a priori, that is, it was based on the literature review and the elaborated research objectives, that the categories of analysis were defined, with slight adjustments being made from the decomposition made to the speeches/emerging from the speeches, namely with regard to the subcategories.

2.7. Ethics in research

The conduct of an educational research requires the researcher to raise moral and ethical issues, thus avoiding embarrassing situations that could compromise the rigor of the investigation.

However, before presenting the set of ethical issues considered by several scholars (Bogdan & Biklen, 1994; Fortin, 1999; Tuckman, 2002; Vilelas, 2009; Freixo, 2009), is fundamental in an educational investigation, to specify the meaning of the concept of ethics, thus guaranteeing the possibility of a more adjusted methodological framework.

The dictionary of philosophy (Audi, 1999, p.284), affirms that the word *ethic* has its origin in the Greek word "*ethos*" which means a way or way of life, including the dispositions of man in life, character, customs and, of course, morality. Ethics, therefore, studies good and thus its object is virtue in the conduct of life - to facilitate the realization of people - is to attain perfection. In other words, the realization of oneself as such, as a person.

The concept of ethics is thus close to the concept of moral as a set of norms, principles, precepts, customs and values that guide the behaviour of the individual in his social group (Freixo, 2009).

In a broader sense, Fortin (1999) considers ethics the science that studies morals, which regulates our posture and our behaviour, always based on a philosophical discourse about which is more correct.

From this conceptual framework, emerge the ethical principles to be respected in an investigation, initiated with the intention of protecting the rights and the freedoms of the research participants. Following the convictions of Bogdan and Biklen (1994), Fortin (1999) and Tuckman (2002) there were considered and followed three principals, such us: the right in privacy or non-participation; the right to anonymity and confidentiality; the right to protection and fair and equitable treatment. The three principles are described below.

Right in privacy or non-participation. This principle anticipates the premise that any subject has the right to freely decide whether or not to participate in an investigation. The subjects are invited to participate in the study and without any coercive means exercised by the researcher influencing the decision, it is up to them the final decision of participation and their permanence in the investigation.

In this study, these rights were respected: at the time of the invitation to participate, the researcher was concerned to communicate the research objectives properly, without hiding any information to the participants; made reference to the freedom to cede information to where they contend, so as not to hurt the sensibilities of the intimate forum.

Right to anonymity and confidentiality. In an investigation the participants have the right to anonymity, requiring that their personal identification data not be included in any accessible part of the documents resulting from the investigation process. The right to confidentiality guarantees the study participants have the right to require that "the data are exclusively for the purpose that they were taken and not others" (Freixo, 2009, p.180), as well the guarantee that the data cannot be divulged or patched. These rights were adhered to throughout the investigation.

The right to protection and fair and equitable treatment. The participants in research have the right to be protected against any possible harm to him or to harm him, as well to be treated equally and justifiably before, during and after his participation. In this matter, the investigator must be responsible and ensure that the participants do not suffer any sequels resulting from their participation in the study.

Lima, A.J. (2006) underlined that ethical issues are placed at all stages of an investigation, from the choice of the subject and the definition of the research questions, through the selection of the participants, to the way of access to the field, the form of data collection, the procedures analysis, the writing of the text and the publication of the results. The investigator must take into account the deontological question before starting the data collection and during its course. Coelho (2013) consider that before starting an investigation, the researcher has the obligation to evaluate the ethical acceptability of his study, and if his research presents a minimal risk, it should be reported in advance to the participants, to decide whether or not to continue with their contribution. However, the authors consider that research ethics involves the dilemma of the existence of two value systems: one that translates into the need for research in itself and the other, which is the belief in human dignity and the right to privacy. The same author pointed out that, in an investigation with people, ethics requires attention to two points: informed consent and protection of subjects against harm. In this study, we sought to respect the ethical procedures of any researcher, especially respect for all stakeholders. So explained in detail the purpose of the investigation, permission was requested for the interviews to be performed and recorded on audio. On the day appointed for the interviews, at the beginning of the interviews, a script was read. Throughout the investigation, anonymity was guaranteed, and all the interveners were codified in order to make it impossible to identify them, even in the interviews where names appeared, they were coded. In all the interviews the initial commitment of the transcription was fulfilled so that it was possible to add or change some data in order to be validated.

CHAPTER III – QUANTITATIVE RESULTS ABOUT TEACHERS COMPETENCES EVALUATION

This section presents the results of the quantitative data about the student's evaluation on the perception of teachers' performance and competences and teacher's perceptions about their own competences survey. In addition, it compares the two survey results in order to identify differences and/ or similarities between the students' evaluation and the teachers' self-evaluation.

3.1 The students' evaluation on teachers competences

The descriptive statistics and the Pearson r correlation was used to analyze the data gathered through the 28-item questionnaire containing four competences of teachers evaluated by the 342 students.

The scores attributed by the students in each of the 28 items revealed an equal dispersion of data between 1 (minimum) and 5 (maximum) points, corresponding to "very bad" and "very good", respectively.

The index scores (mean responses on the items defining each factor) were used to examine the correlations among the teaching competences. The findings revealed that these correlations are all positive, moderate to high and statistically significant. The four teacher competences reflect the complexity and flexibility of teaching and of the instructor's task. The descriptive statistics indicated very close average values in the four competency groups, with personality competences (M = 3.66, SD = .71) and social competences (M = 3.63, SD = .80) with higher averages. Taking into account the average values obtained for each set of competences, these were close to the "good" classification (4). The students' competences evaluation ranged from (3) "sufficient" to (4) "good". The descriptive statistics (means, SDs and reliability coefficients) of students' responses regarding the importance they assign to the four studied competences (Table 10).

Table 10. Descriptive statistics and correlation of students' evaluation on teacher competences

Competences	N	Min.	Max.	M	SD	α	CPed	CProf	CPersn	CSoc
Pedagogical (CPed)	342	1,56	5,00	3,56	,67	,813	1			
Professional (CProf)	342	1,50	5,00	3,59	,66	,798	,675**	1		
Personal (CPersn)	342	1,67	5,00	3,66	,71	,770	,609**	,697**	1	
Social (CSoc)	342	1,00	5,00	3,63	,80	,771	,551**	,604**	,682**	1

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation. **p<.01

The Pearson correlation test indicated a positive and highly significant correlation between all items, as well as between the four competences under analysis (p<.01). The Cronbach's alpha values for each competence showed good internal consistency of the related items evaluated ($\alpha > .7$). It was

verified that the students' evaluation strongly correlated the personal competences and the professional competences ($r = .697$) and social competences ($r = .682$). According to the students, it is the social competences that exert less weight on pedagogical competences ($r = .551$).

The scores attributed by the students in each 28-item revealed an equal dispersion of data between 1 (minimum) and 5 (maximum) points, corresponding to "very bad" and "very good", respectively. The results of the descriptive statistics for each of the items also indicated very close average values and, in general, a high standard deviation (around one unit) which suggests a coefficient of variation around 30% in students' evaluation.

Regarding the pedagogical competencies, the "Adaptation of materials of the exam/ working towards the goal of the subject" (Mean = 3.83, SD = .94) and "Give materials and clear answers to questions in the class" (Mean = 3.67, SD = 1.02) were the highest evaluated items. Items with the lowest average level (3) "sufficient" were the "Capacity to create a living environment in class" (Mean = 3.49, SD = 1.05), the item "Use of media and pedagogical technology" (Mean = 3.36, SD = 1.20) and the "Assignment of feedback on the work / evaluation" (Mean = 3.33, SD = 1.10) (Table 11).

Table 11. Descriptive statistics of students' evaluation on teacher pedagogical competences

	N	Min.	Max.	M	SD
1. Preparation of lessons and practices	342	1,00	5,00	3,63	1,10
2. Organization and discipline in the classroom	342	1,00	5,00	3,58	1,02
3. Ability to create a living environment in the classroom	342	1,00	5,00	3,49	1,05
4. Give materials and clear answers to questions in the class	342	1,00	5,00	3,67	1,02
5. Use of media and educational technology	342	1,00	5,00	3,36	1,20
6. Variety of forms of assessment of learning outcomes	342	1,00	5,00	3,54	,97
7. Assignment of feedback on the work / evaluation	342	1,00	5,00	3,33	1,10
8. Adaptation of materials of the exam/ working towards the goal of the subject	342	1,00	5,00	3,83	,94
9. Harmonization of points or grades awarded on the agenda with the result of learning	342	1,00	5,00	3,55	1,03

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

In relation to professional competences, the lower scores corresponded to item "Ability of interdisciplinary explanations of the current topic with others" (M = 3.48, SD = 1.01) and "Ability to use the various communication technologies" (M = 3.38, SD = 1.13). The remaining items received an average rating very close to (4) "good". The items with higher student's evaluation were "Ability to properly explain the content / approach / topic / concept" (M = 3.80, SD = 1.01) and "Ability to give relevant examples of concepts" (M = 3.66, SD = .92) (Table 12).

Table 12. Descriptive statistics of students' evaluation on teacher professional competences

	N	Min.	Max.	M	SD
10. Ability to properly explain the content / approach / topic / concept	342	1,00	5,00	3,80	1,01
11. Ability to give relevant examples of concepts	342	1,00	5,00	3,66	,92
12. Ability to interdisciplinary explanations of the current topic with others	342	1,00	5,00	3,48	1,01
13. Ability to interdisciplinary topic of the current explanations with everyday realities	342	1,00	5,00	3,54	,96
14. Mastery of more advanced issues of the subject (materials / references of learning lessons)	342	1,00	5,00	3,63	1,01
15. Use of research results to improve the quality of learning in the classroom	342	1,00	5,00	3,57	1,06
16. Involving students in research / analysis and develop the design made by the teacher	342	1,00	5,00	3,59	1,00
17. Ability of using various communication technologies	342	1,00	5,00	3,38	1,13

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

Regarding to 6-item personal competences, the lowest evaluated item corresponded to level "3" "sufficient" and was observed in item "21. Watchword and measures" (M = 3.39, SD = 1.00). The remaining items received an average rating very close to (4) "good" (Table 13).

Table 13. Descriptive statistics of students' evaluation on teacher personal competences

	N	Min.	Max.	M	SD
18. Dignity and personality as a teacher	342	1,00	5,00	3,91	1,05
19. Honesty and fairness in the decision	342	1,00	5,00	3,60	1,09
20. Modelling by the position and behavior	342	1,00	5,00	3,63	,967
21. Watchword and measures	342	1,00	5,00	3,39	1,00
22. Self-Confidence in any situation	342	1,00	5,00	3,72	,98
23. Fair and neutral to all students	342	1,00	5,00	3,65	1,11

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

Regarding social competences, the scores attributed by the students were once again close to the "good" level (4) and no item with an average of less than 3.5 was observed (Table 14).

Table 14. Descriptive statistics of students' evaluation on teacher social competences

	N	Min.	Max.	M	SD
24. Ability of expression	342	1,00	5,00	3,50	1,12
25. Ability to accept criticism and suggestions from students	342	1,00	5,00	3,62	1,09
26. Knowing well the students of their subject	342	1,00	5,00	3,62	1,10
27. Easily get along with peers, staff and students	342	1,00	5,00	3,57	1,09
28. Tolerance in religious groups of students	342	1,00	5,00	3,81	1,08

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

The evaluation of the competences of UNTL teachers by the students of the same institution showed a positive and satisfactory classification, since the average values of the classifications (for each competence and for each set of competences) were between level (3) "sufficient" and the "good" level (4). The items of the pedagogical competences, followed by the items of professional

competences, were those that registered a greater oscillation of the students' scores. The items related to social competences obtained the most consonant average values. Table 15 presents the results of the frequency statistics of students' evaluation on UNTL teacher competences, in relation to teachers' pedagogical, professional, personality and social competences.

Table 15. Frequency statistics of students' evaluation on teacher competences

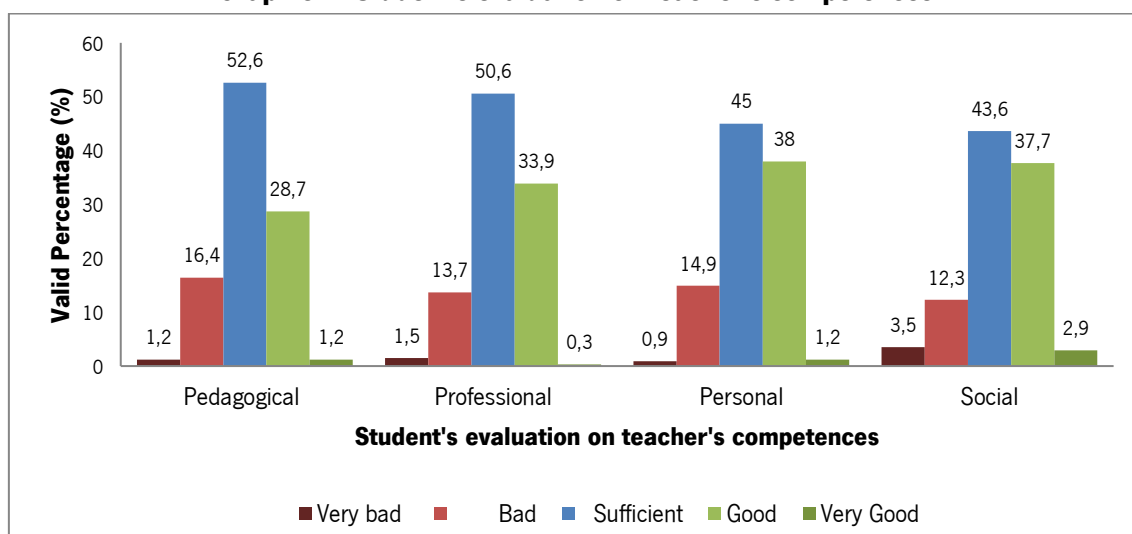
	Very bad n (%)	Bad n (%)	Sufficient n (%)	Good n (%)	Very Good n (%)
Pedagogical	4 (1,2)	56 (16,4)	180 (52,6)	98 (28,7)	4 (1,2)
Professional	5 (1,5)	47 (13,7)	173 (50,6)	116 (33,9)	1 (0,3)
Personality	3 (0,9)	51 (14,9)	154 (45,0)	130 (38,0)	4 (1,2)
Social	12 (3,5)	42 (12,3)	149 (43,6)	129 (37,7)	10 (2,9)

Legend: n-number of cases observed; % - Valid percentage.

The majority of the students evaluated as "sufficient" the pedagogical competence of UNTL teachers (52.6%; n = 180) while 28.7% (n = 98) classified it as "good" and 16.4% = 56) as "bad." The percentage of students who rated "very bad" and "very good" was only 1.2% (n = 2) in each classification category. In relation to professional competences, 50.6% of students (n = 173) classified the performance of teachers as "sufficient" and 33.9% (n = 116) as "good", although 13.7% of students = 47) has considered it "bad." The competences of the personality were evaluated with "sufficient" (45%, n = 154) and "good" (38%; n = 130), while 14.9% of the students (n = 51) classified as "bad". Social competences were mostly evaluated with "sufficient" (43.6%, n = 149) and "good" (37.7%, n = 129), while 12.3% of students (n = 42) evaluated as "bad" and 2.9% (n = 10) as "very bad".

The graphic 1 illustrates de students' evaluation in relation to each competence.

Graphic 1. Student's evaluation on teacher's competences



Among the UNTL's students there is a tendency to evaluate as "sufficient" the teacher's competences, mainly the pedagogical and professional one. About 30% of students evaluate the teacher's competences as "good". There is a more divergent student's evaluation in relation to personal and social competences.

In order to understand if students' background characteristics influence their evaluation on teacher's competences, it is important to analyze the statistical association between the students' evaluation and their personal/ academic information, like gender, age and faculty, with the Fisher's exact test.

Table 16 presents the results of the statistical association between students' gender and the evaluation of the UNTL teacher's competences.

Table 16. Statistical association between students' evaluation on teacher competences and students gender

		Male	Female	Fisher's Exact Test (df)
		n (%)	n (%)	p-value
Pedagogical Competences	Very bad	3 (75,0)	1 (25,0)	$\chi^2(4)=9,007$ $p=,045$
	Bad	34 (60,7)	22 (39,3)	
	Sufficient	83 (46,1)	97 (53,9)	
	Good	43 (43,9)	55 (56,1)	
	Very good	0 (0,0)	4 (100,0)	
Professional Competences	Very bad	4 (80,0)	1 (20,0)	$\chi^2(4)=11,347$ $p=,012$
	Bad	28 (59,6)	19 (40,4)	
	Sufficient	87 (50,3)	86 (49,7)	
	Good	43 (37,1)	73 (62,9)	
	Very good	1 (100,0)	0 (0,0)	
Personal Competences	Very bad	2 (66,7)	1 (33,3)	$\chi^2(4)=3,926$ $p=,437$
	Bad	29 (56,9)	22 (43,1)	
	Sufficient	75 (48,7)	79 (51,3)	
	Good	55 (42,3)	75 (57,7)	
	Very good	2 (50,0)	2 (50,0)	
Social Competences	Very bad	4 (33,3)	8 (66,7)	$\chi^2(4)=12,081$ $p-value=,015$
	Bad	26 (61,9)	16 (38,1)	
	Sufficient	77 (51,7)	72 (48,3)	
	Good	49 (38,0)	80 (62,0)	
	Very good	7 (70,0)	3 (30,0)	

Legend: n-number of cases observed; % - Valid percentage; Df - degrees of freedom.

The performance of pedagogical competences was rated as "very bad" by 75% (n = 3) male students and "bad" by 60.7% (n = 34) of the same gender. The classification of "sufficient" was attributed by 53.9% of female students (n = 97) and 46.1% of male students (n = 83). The evaluation of "good" was also given mostly by 56.1% of female students (n = 55) and 43.9% of male students (n = 43). Only four female students classified as "very good" the pedagogical competences of teachers.

The students' gender presented a statistically significant association with the evaluation of the pedagogical competences of the UNTL teachers ($X^2(4) = 9.007$; $p < .05$), with a significant association between the male students and the "bad" classification (Adj. Res. = 2.1). It was also observed that the most positive classifications ("good" and "very good") were attributed mostly by female students; in turn, the "very bad" and "bad" classifications were attributed mostly by male students.

Regarding professional competences, the lowest scores were mostly attributed by male students and the highest scores were mostly attributed by female students: 80% ($n = 4$) classified them as "very bad", 59.6% ($n = 28$) with "bad" and 50.3% ($n = 87$) with "sufficient". There were 62.9% ($n = 73$) of female students who assigned the classification of "good" and only one male student who evaluated the "professional performance" of the teacher as "very good". There was a statistically significant association between professional competences and gender ($X^2(4) = 11.347$; $p < .05$), but it was the female students who were associated with the classification of "good" (Adj. Res = 2,8).

The evaluation of the personal competences did not present a significant statistical association with gender of the students ($X^2(4) = 3,926$; $p < .05$). The prevalence of boys in the lowest ranks was again verified: 66.7% ($n = 2$) assessed the competencies of the teachers' personality with "very bad" and 56.9% ($n = 29$) with "bad". Female students were the most frequent in the classifications of "sufficient" (51.3%; $n = 79$) and "good" (57.7%; $n = 75$), while the evaluation of "very good" was attributed by an equal number of male and female students (50% $n = 2$).

Regarding the evaluation of social competences, it was observed that 66.7% ($n = 8$) of the female students rated it as "very bad" and 61.9% ($n = 26$) with "bad". The "sufficient" evaluation was attributed by 51.7% ($n = 77$) of male students and 48.3% ($n = 72$) of female students. The evaluation of "good" was attributed by 62% ($n = 80$) of female students and by 38% ($n = 49$) of male students; 70% ($n = 7$) of the "very good" assessments were attributed by the boys. The gender of the students presented a statistically significant association with the evaluation of the social competences of the UNTL teachers ($X^2(4) = 12,081$; $p < .05$), and the male students were associated with the classification of "bad" (Adj. Res. = 2.0) and female students were associated with the classification of "good" (Adj. Res. = 2,8).

Fisher's exact test did not reveal a statistically significant association between the age group of the students and the evaluation of teachers' competences ($p > .05$) (Table 17).

Table 17. Statistical association between students' evaluation on teacher competences and students age

		18-24 Years	25-31 Years	> 31 Years	Fisher Exact Test (df)
		n (%)	n (%)	n (%)	p-value
Pedagogical Competences	Very bad	3 (75,0)	1 (25,0)	0 (0,0)	$\chi^2(12)=20,061$ <i>p-value=,258</i>
	Bad	49 (87,5)	7 (12,5)	0 (0,0)	
	Sufficient	153 (85,0)	26 (14,4)	1 (0,6)	
	Good	91 (92,9)	6 (6,1)	1 (1,0)	
	Very good	4 (100,0)	0 (0,0)	0 (0,0)	
Professional Competences	Very bad	3 (60,0)	2 (40,0)	0 (0,0)	$\chi^2(12)=22,590$ <i>p-value=,270</i>
	Bad	39 (83,0)	8 (17,0)	0 (0,0)	
	Sufficient	155 (89,6)	17 (9,8)	1 (0,6)	
	Good	102 (87,9)	13 (11,2)	1 (0,9)	
	Very good	1 (100,0)	0 (0,0)	0 (0,0)	
Personal Competences	Very bad	2 (66,7)	1 (33,3)	0 (0,0)	$\chi^2(12)=23,886$ <i>p-value=,086</i>
	Bad	45 (88,2)	6 (11,8)	0 (0,0)	
	Sufficient	133 (86,4)	20 (13,0)	1 (0,6)	
	Good	117 (90,0)	13 (10,0)	0 (0,0)	
	Very good	3 (75,0)	0 (0,0)	1 (25,0)	
Social Competences	Very bad	11 (91,7)	1 (8,3)	0 (0,0)	$\chi^2(12)=16,486$ <i>p-value=,321</i>
	Bad	37 (88,1)	5 (11,9)	0 (0,0)	
	Sufficient	128 (85,9)	20 (13,4)	1 (0,7)	
	Good	116 (89,9)	13 (10,1)	0 (0,0)	
	Very good	8 (80,0)	1 (10,0)	1 (10,0)	

Legend: n-number of cases observed; %- valid percentage; df - degrees of freedom.

It was not possible to indicate that the students of a certain age tended to evaluate the teachers in a certain sense. This result can be understood by the fact that 87.7% of the students interviewed are between 18 and 24 years old, and therefore there is a reduced age variability of the sample. The age group of the students did not present a statistically significant association with the evaluation made to the performance of the teachers' competences ($p > .05$).

Regarding pedagogical competences, 75% ($n = 3$) of the students who evaluated them as "very bad" were between 18 and 24 years old, as well as 87.5% ($n = 49$) of the students who evaluated them as "bad", 85% ($n = 153$) of the students who rated as "sufficient", 92.9% ($n = 91$) of those who rated "good" and 100% ($n = 4$) of the students who evaluated them as "very good".

Regarding the professional competences, 89.6% ($n = 155$) of the students who evaluated them with "sufficient" were between 18 and 24 years old, as well as the 87.9% ($n = 102$) of students who assigned as "Good".

The personal competences had a more ambivalent evaluation in the 18-24 age group, since 88.2% ($n = 45$) of the students attributed the classification of "bad" and 90% ($n = 117$) of those who attributed the classification of "good" belonged to this age group.

Social competences were assessed as "bad" by 88.1% (n = 37) of students between the ages of 18 and 24; 85.9% (n = 128) of students in the same age group classified the performance of social competences as "sufficient" and 89.9% (n = 116) as "good".

Table 18 presents the results of the statistical association test between the evaluation of the teachers' competences and the students' faculty. It presents the classifications attributed by the students from Faculty of Philosophy (FP), Faculty of Education, Arts and Humanities (FEAH), Faculty of Medicine (FM) and Faculty of Political Science (FPC) to the set of teacher's pedagogical professional, personality and social competences.

Table 18. Statistical association between students' evaluation on teacher competences and students faculty

		FP <i>n (%)</i>	FEAH <i>n (%)</i>	FM <i>n (%)</i>	FPS <i>n (%)</i>	Fisher Exact Test (df) <i>p-value</i>
Pedagogical Competences	Very bad	2 (50,0)	1 (25,0)	0 (0,0)	1 (25,0)	$\chi^2(12)=36,224$ <i>p-value</i> =,000
	Bad	25 (44,6)	14 (25,0)	6 (10,7)	11 (19,6)	
	Sufficient	46 (25,6)	53 (29,4)	49 (27,2)	32 (17,8)	
	Good	14 (14,3)	30 (30,6)	43 (43,9)	11 (11,2)	
	Very good	0 (0,0)	1 (25,0)	3 (75,0)	0 (0,0)	
Professional Competences	Very bad	3 (60,0)	2 (40,0)	0 (0,0)	0 (0,0)	$\chi^2(12)=50,740$ <i>p-value</i> =,000
	Bad	23 (48,9)	10 (21,3)	4 (8,5)	10 (21,3)	
	Sufficient	45 (26,0)	54 (31,2)	42 (24,3)	32 (18,5)	
	Good	15 (12,9)	33 (28,4)	55 (47,4)	13 (11,2)	
	Very good	1 (100,0)	0 (0,0)	0 (0,0)	0 (0,0)	
Personal competences	Very bad	3 (100,0)	0 (0,0)	0 (0,0)	0 (0,0)	$\chi^2(12)=43,878$ <i>p-value</i> =,000
	Bad	20 (39,2)	12 (23,5)	5 (9,8)	14 (27,5)	
	Sufficient	42 (27,3)	44 (28,6)	40 (26,0)	28 (18,2)	
	Good	22 (16,9)	40 (30,8)	55 (42,3)	13 (10,0)	
	Very good	0 (0,0)	3 (75,0)	1 (25,0)	0 (0,0)	
Social Competences	Very bad	4 (33,3)	2 (16,7)	0 (0,0)	6 (50,0)	$\chi^2(12)=40,261$ <i>p-value</i> =,000
	Bad	11 (26,2)	11 (26,2)	7 (16,7)	13 (31,0)	
	Sufficient	44 (29,5)	45 (30,2)	35 (23,5)	25 (16,8)	
	Good	25 (19,4)	38 (29,5)	55 (42,6)	11 (8,5)	
	Very good	3 (30,0)	3 (30,0)	4 (40,0)	0 (0,0)	

Legend: n-number of cases observed; % Valid percentage; Df - degrees of freedom.

There was a statistically significant association between the Faculty and the evaluation of teachers' pedagogical competences ($\chi^2(12) = 36,224$; $p < .01$), with students from the Faculty of Philosophy (FP) being associated with the "bad" classification (44.6 %; $n=25$) (Adj. Res. = 3,6) and students of the Faculty of Medicine (FM) were statistically associated with the classification of "good" (43.3%; $n = 43$) (Adj. Res. = 3,7) and "very good" (75%; $n = 3$) (Adj. Res. = 2.0).

In relation to the evaluation of the performance of professional competences, there was also a

significant statistical association with the Faculty of students ($X^2 (12) = 50,740$, $p < .01$) and also were students of the Faculty of Philosophy who presented a significant association with the classification of "bad" (48,9 %; $n=23$) (Adj.Res.=4,0) and the students of the Faculty of Medicine associated with the classification of "good" (47.4%; $n = 55$) (Adj. Res. = 5.2).

In terms of the evaluation of personal competences, there was a significant association between the evaluation of the students of the Faculty of Philosophy and the classification of "very bad" (100 %; $n = 3$) (Adj. Res.=3.0) and "bad" (39.2%, $n = 20$) (Adj. Res. = 2.4), as well as among students of the Faculty of Political Science (FPS) for the evaluation of "bad" (27.5%; $n = 14$) (Adj. Res. = 2.4). The students of the Faculty of Medicine presented a significant association with the classification of "good" (42.3%; $n = 55$) (Adj. Res. = 4.1) and the students of the Faculty of Education, Arts and Humanities appeared statistically associated with the classification of "very good" (75%; $n = 3$) (Adj. Res. = 2.0).

The evaluation of social competences also showed a significant association with the Faculty of students ($X^2 (12) = 40,261$; $p < .01$) and the students of the Faculty of Political Science emerged associated with the evaluation of "very bad" (50%; $N = 6$) (Adj. Res. = 3,3) and "bad" (31%; $n = 13$) (Adj. Res. = 2,8), whereas the students of the Faculty of Medicine presented a significant association with the classification of "good" (42.6%, $n = 55$) (Adj. Res. = 4.1).

It was intended also to analyze the agreement proportion of students' evaluation in each faculty. The next cross tables represent the proportion of students' disagreement and agreement in relation to teacher's evaluation competences, for each faculty.

Table 19 represents the proportion of students' disagreement and agreement in relation to teacher's evaluation on pedagogical competences, for each faculty.

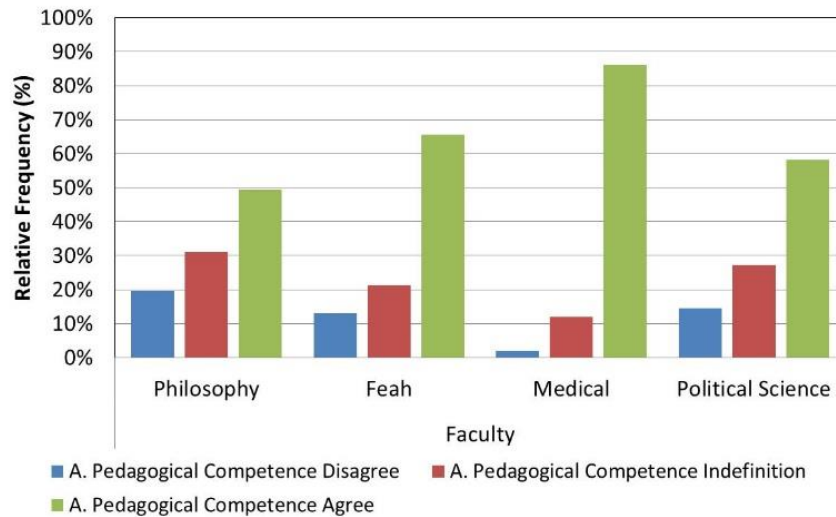
Table 19. Proportion of agreement on students' evaluation of pedagogical competencies, by faculty

		A. Pedagogical Competences		
		Disagree	Indefinition	Agree
Philosophy	N	17	27	43
	% in faculty	19,5%	31,0%	49,4%
Education, Arts and Humanities	N	13	21	65
	% in faculty	13,1%	21,2%	65,7%
Medical	N	2	12	87
	% in faculty	2,0%	11,9%	86,1%
Political Science	N	8	15	32
	% in faculty	14,5%	27,3%	58,2%

There was a highest proportion of concordant assessments in Medical Faculty (86%) followed by

Faculty of Education, Arts and Humanities (66%). Bar chart illustrates the proportion of agreement on students' evaluation of pedagogical of teachers, by faculty.

Graphic 2. Proportion of agreement on students' evaluation of pedagogical competencies, by faculty



In relation to pedagogical competences, the proportion of student's evaluation agreement was highest on Medical Faculty and lowest on Philosophy and Political Science Faculties. The proportion of student's evaluation disagreement was highest on Philosophy Faculty, as well on Political Science Faculty and lowest on Medical Faculty. These differences were statistically significant, as indicated by the chi-square teste ($\chi^2_{(6)} = 32,236$; $p < 0,001$).

Table 20 represents the proportion of students' disagreement and agreement in relation to teacher's evaluation on professional competences, for each faculty.

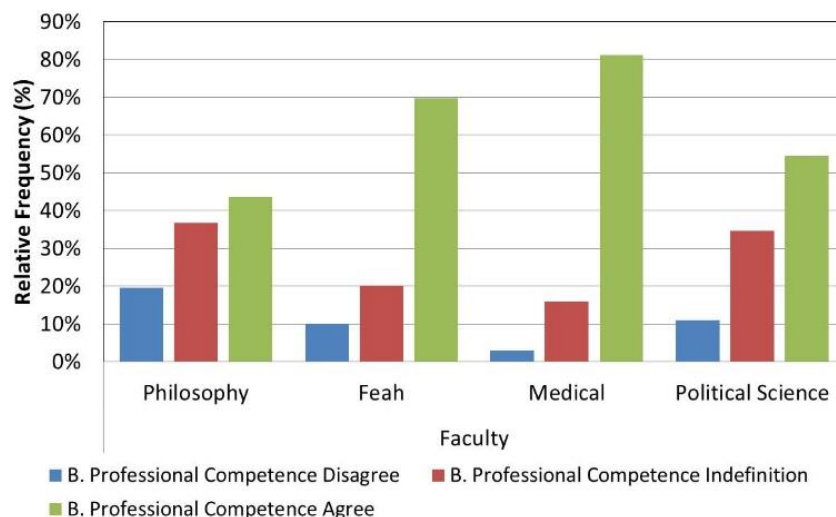
Table 20. Proportion of agreement on students' evaluation of professional competencies, by faculty

		B. Professional Competence		
		Disagree	Indefinition	Agree
Philosophy	N	17	32	38
	% in faculty	19,5%	36,8%	43,7%
Education, Arts and Humanities	N	10	20	69
	% in faculty	10,1%	20,2%	69,7%
Medical	N	3	16	82
	% in faculty	3,0%	15,8%	81,2%
Political Science	N	6	19	30
	% in faculty	10,9%	34,5%	54,5%

Students from Medical Faculty had the highest proportion of agreement regarding the evaluation of teacher's professional competences (81%), while the students from Philosophy Faculty had the

highest proportion of disagreement (20%). Bar chart illustrates the proportion of agreement on students' evaluation of professional of teachers, by faculty.

Graphic 3. Proportion of agreement on students' evaluation of professional competencies, by faculty



The proportion of student's evaluation agreement about teachers' professional competences was highest on Medical Faculty and lowest on Philosophy Faculty. The proportion of student's evaluation disagreement was highest on Philosophy Faculty and lowest on Medical Faculty. These differences were statistically significant, as indicated by the chi-square teste ($\chi^2_{(6)} = 34,676$; $p < 0,001$).

Table 21 represents the proportion of students' disagreement and agreement in relation to teacher's evaluation on personal competences, by faculty.

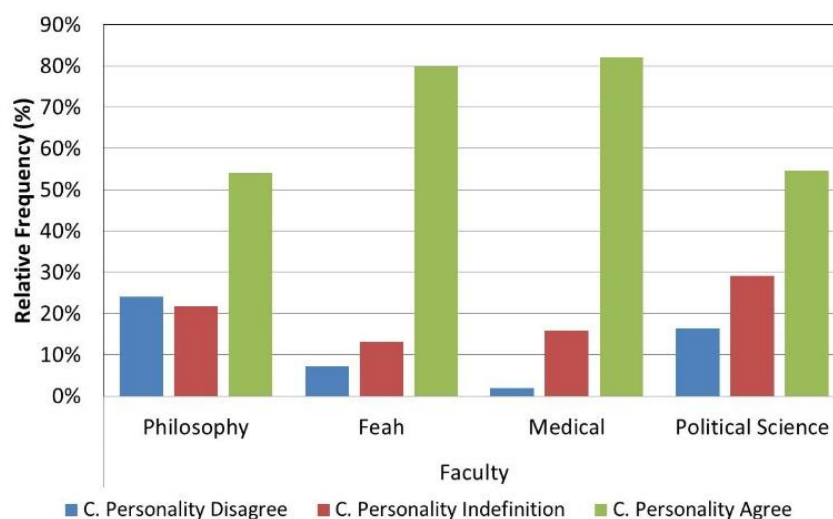
Table 21. Proportion of agreement on students' evaluation of personal competencies, by faculty

		C. Personality		
		Disagree	Indefinition	Agree
Philosophy	N	21	19	47
	% in faculty	24,1%	21,8%	54,0%
Education, Arts and Humanities	N	7	13	79
	% in faculty	7,1%	13,1%	79,8%
Medical	N	2	16	83
	% in faculty	2,0%	15,8%	82,2%
Political Science	N	9	16	30
	% in faculty	16,4%	29,1%	54,5%

Students from Medical Faculty had also the highest proportion of agreement regarding the evaluation of teacher's personal competences (82%), while the students from Philosophy Faculty had the highest proportion of disagreement (24%). Bar chart illustrates the proportion of agreement

on students' evaluation of personal of teachers, by faculty.

Graphic 4. Proportion of agreement on students' evaluation of personal competencies, by faculty



Regarding the student's evaluation of teacher's personal competences, the proportion of agreement was highest among students from Medical Faculty (82%) and from Education, Arts and Humanities Faculty (80%). The highest proportion of disagreement was among students from Philosophy Faculty, followed by Political Science Faculty. The proportion of disagreement was residual among students from Medical Faculty, the differences being statistically significant, as indicated by the chi-square teste ($\chi^2_{(6)} = 37,327$; $p < 0,001$).

Table 22 represents the proportion of students' disagreement and agreement in relation to teacher's evaluation on social competences, for each faculty.

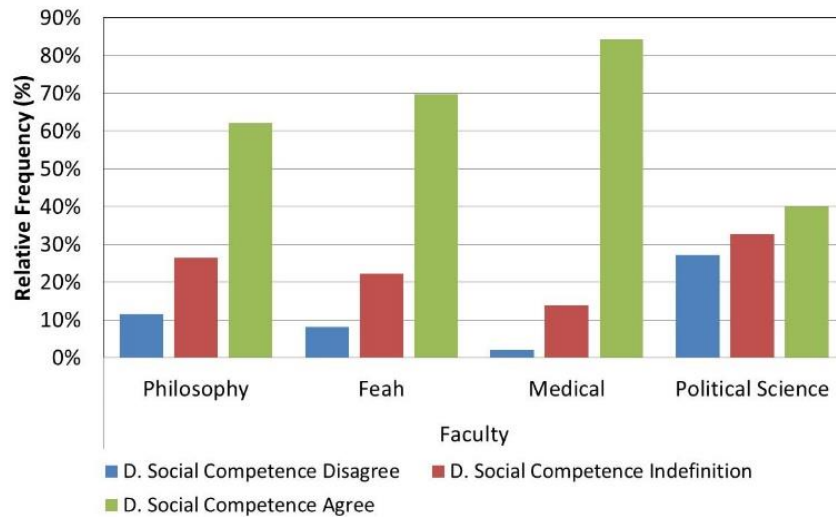
Table 22. Proportion of agreement on students' evaluation of social competencies, by faculty

		D. Social Competence		
		Disagree	Indefinition	Agree
Philosophy	N	10	23	54
	% in faculty	11,5%	26,4%	62,1%
Education, Arts and Humanities	N	8	22	69
	% in faculty	8,1%	22,2%	69,7%
Medical	N	2	14	85
	% in faculty	2,0%	13,9%	84,2%
Political Science	N	15	18	22
	% in faculty	27,3%	32,7%	40,0%

Students from Medical Faculty had also the highest proportion of agreement regarding the evaluation of teacher's social competences (84%), while the highest proportion of disagreement

was 27% in Political Science Faculty. Bar chart illustrates the proportion of agreement on students' evaluation of social of teachers, by faculty.

Graphic 5. Proportion of agreement on students' evaluation of social competencies, by faculty



The students' evaluation regarding the teachers' social competences had a high proportion in four faculties. The proportion of agreement was highest in Medical Faculty (84%) as well in Education, Arts and Humanities Faculty (70%). In terms of disagreement the highest proportion was in Political Science Faculty (27%). These differences were statistically significant as indicated by the chi-square teste ($\chi^2_{(6)} = 40,216$; $p < 0,001$).

Pertaining to the students' evaluation of teachers' competences, the proportion of agreement (equal evaluation) was also analyzed between all pairs of evaluated competences, with a contingency cross table for each pair. Table 23 shows the proportion of agreement between the evaluation of pedagogical and professional competences, made by students.

Table 23. Proportion of agreement between the student's evaluation of pedagogical and professional competences

A. Pedagogical Competences		B. Professional Competences		
		Disagree	Indefinition	Agree
Disagree	N	19	16	5
	% in group	47,5%	40,0%	12,5%
Indefinition	N	11	31	33
	% in group	14,7%	41,3%	44,0%
Agree	N	6	40	181
	% in group	2,6%	17,6%	79,7%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “A. Pedagogical Competence” show a percentage of 40,0% for indefiniteness and 12,5% of agree with “B. Professional Competence”, the ones that show indefiniteness with “A. Pedagogical Competence” show a percentage of 14,7% of disagree and 44,0% of agree with “B. Professional Competence”. The ones that agree with “A. Pedagogical Competence” show a percentage of 2,6% of disagree and 17,6% of indefiniteness with “B. Professional Competence”, however there is a statistically significant agreement, as indicated by the measure of agreement of 36,0% specified by the kappa coefficient ($p < 0,001$).

Table 24 shows the proportion of agreement between the student’s evaluation of pedagogical and personal competences.

Table 24. Proportion of agreement between the student’s evaluation of pedagogical and personal competences

A. Pedagogical Competences		C. Personal Competences		
		Disagree	Indefinition	Agree
Disagree	N	17	11	12
	% in group	42,5%	27,5%	30,0%
Indefinition	N	11	27	37
	% in group	14,7%	36,0%	49,3%
Agree	N	11	26	190
	% in group	4,8%	11,5%	83,7%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “A. Pedagogical Competence” show a percentage of 27,5% for indefiniteness and 30,0% of agree with “C. Personality”, the ones that show indefiniteness with “A. Pedagogical Competence” show a percentage of 14,7% of disagree and 49,3% of agree with “C. Personality”, the ones that agree with “A. Pedagogical Competence” show a percentage of 4,8% of disagree and 11,5% of indefiniteness with “C. Personality”, however there is a statistically significant agreement, as indicated by the measure of agreement of 34,5% specified by the kappa coefficient ($p < 0,001$).

Table 25 shows the proportion of agreement between the evaluation of pedagogical and social competences, made by students.

Table 25. Proportion of agreement between the student's evaluation of pedagogical and social competences

A. Pedagogical Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N	13	18	9
	% in group	32,5%	45,0%	22,5%
Indefinition	N	8	27	40
	% in group	10,7%	36,0%	53,3%
Agree	N	14	32	181
	% in group	6,2%	14,1%	79,7%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “A. Pedagogical Competence” show a percentage of 45,0% for indefiniteness and 22,5% of agree with “D. Social Competence”, the ones that show indefiniteness with “A. Pedagogical Competence” show a percentage of 10,7% of disagree and 53,3% of agree with “D. Social Competence”, the ones that agree with “A. Pedagogical Competence” show a percentage of 6,2% of disagree and 14,1% of indefiniteness with “D. Social Competence”, however there is a statistically significant agreement, as indicated by the measure of agreement of 28,1% specified by the kappa coefficient ($p < 0,001$). Table 26 shows the proportion of agreement between the evaluation of professional and personal competences, made by students.

Table 26. Proportion of agreement between the student's evaluation of professional and personal competences

B. Professional Competences		C. Personal Competences		
		Disagree	Indefinition	Agree
Disagree	N	22	9	5
	% in group	61,1%	25,0%	13,9%
Indefinition	N	12	32	43
	% in group	13,8%	36,8%	49,4%
Agree	N	5	23	191
	% in group	2,3%	10,5%	87,2%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “B. Professional Competence” show a percentage of 25,0% for indefiniteness and 13,9% of agree with “C. Personality”, the ones that show indefiniteness with “B. Professional Competence” show a percentage of 13,8% of disagree and 49,4% of agree with “C. Personality”, the ones that agree with “B. Professional Competence” show a percentage of 2,3% of disagree and 10,5% of indefiniteness with “C. Personality”, however there is a statistically significant agreement, as indicated by the measure of agreement of 42,5% specified by the kappa coefficient ($p < 0,001$).

Table 27 shows the proportion of agreement between the students' evaluation of professional and social competences.

Table 27. Proportion of agreement between the student's evaluation of professional and social competences

B. Professional Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N	13	12	11
	% in group	36,1%	33,3%	30,6%
Indefinition	N	11	37	39
	% in group	12,6%	42,5%	44,8%
Agree	N	11	28	180
	% in group	5,0%	12,8%	82,2%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with "B. Professional Competence" show a percentage of 33,3% for indefiniteness and 30,6% of agree with "D. Social Competence", the ones that show indefiniteness with "B. Professional Competence" show a percentage of 12,6% of disagree and 44,8% of agree with "D. Social Competence", the ones that agree with "B. Professional Competence" show a percentage of 5,0% of disagree and 12,8% of indefiniteness with "D. Social Competence", however there is a statistically significant agreement, as indicated by the measure of agreement of 34,7% specified by the kappa coefficient ($p < 0,001$).

Table 28 shows the proportion of agreement between the students' evaluation of personal and social competences.

Table 28. Proportion of agreement between the student's evaluation of personal and social competences

C. Personal Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N	19	9	11
	% in group	48,7%	23,1%	28,2%
Indefinition	N	8	32	24
	% in group	12,5%	50,0%	37,5%
Agree	N	8	36	195
	% in group	3,3%	15,1%	81,6%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with "C. Personality" show a percentage of 23,1% for indefiniteness and 28,2% of agree with "D. Social Competence", the ones that show indefiniteness with "C. Personality" show a percentage of 12,5%

of disagree and 37,5% of agree with “D. Social Competence”, the ones that agree with “C. Personality” show a percentage of 3,3% of disagree and 15,1% of indefiniteness with “D. Social Competence”, however there is a statistically significant agreement, as indicated by the measure of agreement of 41,1% specified by the kappa coefficient ($p < 0,001$).

3.2 The teachers’ competences self-evaluation

The quantitative analysis of the self-evaluation questionnaires of the teachers of the UNTL (N = 192) allowed to obtain descriptive and correlational results on the set of competences under study. Cronbach's alpha values for each competence group a high internal consistency in teachers' self-evaluation, in relation to pedagogical competences ($\alpha = .888$), as professional ($r = .794$), personal ($r = .798$) and social competences ($\alpha = .805$).

The Spearman Rho correlation test indicated a positive and highly significant correlation among all items under study and among the four set of competences ($p < .01$). There was a strong correlation between the self-evaluation of personal competences and the pedagogical ($r = .653$) and professional competences ($r = .684$). Pedagogical and professional competences have a strong correlation ($r = .716$) too, while social competences have stronger correlation with personal competences ($r = .732$) and weaker with pedagogical ($r = .512$) and professionals competences ($r = .593$) (Table 29).

Table 29. Descriptive statistics and correlation of teachers’ competences self-evaluation

Competences	N	Min.	Max.	M	SD	α	CPed	CProf	CPersn	CSoc
Pedagogical (CPed)	192	2,56	5,00	4,44	,489	,888	1			
Professional (CProf)	192	2,75	5,00	4,20	,470	,794	,716**	1		
Personal (CPersn)	192	3,00	5,00	4,50	,409	,798	,653**	,684**	1	
Social (CSoc)	192	3,00	5,00	4,54	,425	,805	,512**	,593**	,732**	1

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation. ** $p < .01$

The descriptive statistics indicated very close average values in the four competences, located in the classification (4) "good". The personal competences (M= 4.50, SD=.409) and social skills (M=4.54, SD=.425) have highest and therefore close averages. Self-assessment of personal and social competences ranged between (3) "sufficient" and (5) "very good" with a coefficient of variation of 9% and 8% respectively. Regarding the pedagogical and professional competences, there was a data dispersion of 11%, with ratings ranging from (2) "bad / low" to (5) "very good". The results of the descriptive statistics for each self-evaluated competence also indicated very close average values around the "good" (4) and, in general, a high standard deviation (close to one unit). Regarding pedagogical competences, the highest self-evaluated competences were “objectivity of

student assessment” (M = 4.51, SD=.63), “ability to guide students” (M=4.49, SD=.61) and “seriousness in the preparation of classes (programs, quality plan and its implementation)” (M=4.48, SD=.69). The “use of varied modes of assessment of student study behavior” (M=4.38, SD=.71) and the “mastering in the use of media and learning technologies” M=4.39, SD=.73) has the least self-evaluated items (Table 30).

Table 30. Descriptive statistics of teachers’ pedagogical competences self-evaluation

	N	Min.	Max.	M	SD
1. Seriousness in the preparation of classes (programs, quality plan and its implementation)	192	1,00	5,00	4,48	,69
2. Organization and discipline in the preparation of the classes (to fill the gradual meetings)	192	2,00	5,00	4,42	,68
3. Adapting classroom management to learning goals	192	2,00	5,00	4,43	,67
4. Discipline and compliance with academic regulations	192	2,00	5,00	4,46	,67
5. Mastering in the use of media and learning technologies	192	2,00	5,00	4,39	,73
6. Use of varied modes of assessment of student study behavior	192	2,00	5,00	4,38	,71
7. Objectivity of student assessment	192	2,00	5,00	4,51	,63
8. Ability to guide students	192	2,00	5,00	4,49	,61
9. Positive perception of students' ability	192	2,00	5,00	4,40	,65

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

Regarding professional competences, the “mastering the disciplinary areas of its main function” (M=4.43, SD=.59) and the “ability of interdisciplinary explanations of the current topic with others” (M=4.38, SD=.60) were the highest self-evaluated professional competences. The "ability to achieve the advancement of science and technology to enhance the learning process" (M=3.99, SD=.86) and the "involvement in scientific work and professional organizations" (M=3.64, SD=.82) were the least self-evaluated professional competences. Other items like interdisciplinary, mastering of potential questions, discussion and engage with students were self-evaluated as “good” (Table 31).

Table 31. Descriptive statistics of teachers’ professional competences self-evaluation

	N	Min.	Max.	M	SD
10. Mastering the disciplinary areas of its main function	192	3,00	5,00	4,43	,59
11. Ability of interdisciplinary explanations of the current topic with others	192	3,00	5,00	4,38	,60
12. Ability to illustrate interdisciplinary areas with everyday reality	192	3,00	5,00	4,35	,65
13. Mastering of potential questions (references) in the areas of learning	192	2,00	5,00	4,35	,67
14. Availability to reflect and discuss problems faced by peers	192	2,00	5,00	4,34	,69
15. Engage students in research or analysis and in the project developed by the teacher	192	1,00	5,00	4,08	,92
16. Ability to achieve the advancement of science and technology to enhance the learning process	192	1,00	5,00	3,99	,86
17. Involvement in scientific work and professional organizations	192	2,00	5,00	3,64	,82

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

In face to personal competences, the highest self-evaluated competences as “very good” were the “dignity and personality as teacher” (M=4.58, SD=.54), the “honesty and fairness in decision making” (M=4.55, SD=.52) and the “neutrality and fairness to peers” (M=4.54, SD=.60). The least self-evaluated personal competences were the “be a role model” (M=4.39, SD=.62), the “firm position and application of measures” (M=4.44, SD=.60) and the “self confidence in any situation” (M=4.47, SD=.59) (Table 32).

Table 32. Descriptive statistics of teachers’ personal competences self-evaluation

	N	Min.	Max.	M	SD
18. Dignity and personality as teacher	192	3,00	5,00	4,58	,54
19. Honesty and fairness in decision making	192	3,00	5,00	4,55	,52
20. Be a role model	192	1,00	5,00	4,39	,62
21. Firm position and application of measures	192	3,00	5,00	4,44	,60
22. Self confidence in any situation	192	3,00	5,00	4,47	,59
23. Neutrality and fairness to peers	192	2,00	5,00	4,54	,60

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

Social competences had also a “good” and “very good” evaluation. The “ability to accept criticism and suggestions from students” (M=4.56, SD=.57), the “easy cohabitation with peers, employees and students” (M=4.55, SD=.58) and the “tolerance for different religious denominations” (M=4.60, SD=.56) were the highest self-evaluated social competences. The “social coexistence” (M=4.47, SD=.58) and the “expression capacity” (M=4.49, SD=.58) were the least evaluated social competences but still with a “good” self-evaluation score (Table 33).

Table 33. Descriptive statistics of teachers’ social competences self-evaluation

	N	Min.	Max.	M	SD
24. Expression capacity	192	3,00	5,00	4,49	,58
25. Ability to accept criticism and suggestions from students	192	3,00	5,00	4,56	,57
26. Easy coexistence with peers, staff and students	192	3,00	5,00	4,55	,55
27. Easy social coexistence	192	3,00	5,00	4,47	,58
28. Tolerance for different religious denominations	192	3,00	5,00	4,60	,56

Legend: N- dimension of the sample; Min.-Minimum; Max.-Maximum; M-Mean; SD-Standard Deviation.

The UNTL teachers reveal a high and positive self-perception of their competences, with a highest proportion of “good” and “very good” valuation of their pedagogical, professional, personal and social competences. The pedagogical and professional competences had higher variability of teachers’ self-evaluation, while the personal and social competences had a most homogenous and highest self-evaluation.

Table 34 presents the frequency statistics of teachers’ competences self-evaluation.

Table 34. Frequency statistics of teachers' competences self-evaluation (N = 192)

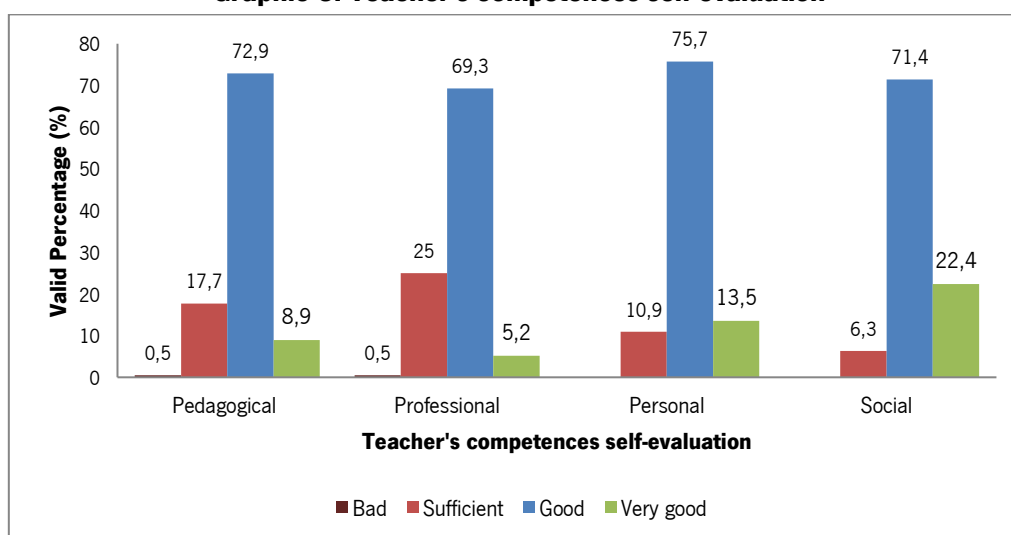
	Very bad <i>n</i> (%)	Bad <i>n</i> (%)	Sufficient <i>n</i> (%)	Good <i>n</i> (%)	Very good <i>n</i> (%)
Pedagogical	0 (0,0)	1 (0,5)	34 (17,7)	140 (72,9)	17 (8,9)
Professional	0 (0,0)	1 (0,5)	48 (25,0)	133 (69,3)	10 (5,2)
Personal	0 (0,0)	0 (0,0)	21 (10,9)	145 (75,5)	26 (13,5)
Social	0 (0,0)	0 (0,0)	12 (6,3)	137 (71,4)	43 (22,4)

The majority of teachers evaluated their pedagogical performance as "good" (72.9%, n=140) while 17.7% (n = 34) as "sufficient" and 8.9% (n=17) as "very good". Only one teacher self-evaluated "bad." In terms of professional competences, 69.3% (n=133) of teachers attributed the classification of "good", 25% (n=48) "sufficient" and 5.2% (n=10) as "very good"; only one teacher attributes the classification "bad".

The personal competences were self-evaluated as "good" by the majority of teachers (75.5%, n=145) and as "very good" (13.5%; n=26), while 10.9% (n=21) self-evaluated that like "sufficient". The performance of social competences had a "good" rating for 71.4% (n = 137) and "very good" for 22.4% (n = 43) of teachers; only 6.3% (n=12) self-evaluated the social competences as "sufficient".

The graphic 6 represents the teacher's self-evaluation in relation to pedagogical, professional, personal and social competences.

Graphic 6. Teacher's competences self-evaluation



There was a high tendency among UNTL's teachers to self-evaluate their competences as "good" with percentages around 70% and 75%, followed by 18% and 25% of teachers who evaluated their pedagogical and professional competences as "sufficient" respectively. In relation to personal and social competences the second higher percentage refers to "very good" self-evaluation, 14% and

22% respectively. There was a residual and inexpressive “bad” self-evaluation and there was no “very bad” self-evaluation among teachers. The UNTL teacher’s self-perception was very homogenous.

In order to understand if teacher’s background characteristics (like gender, age, time of experience, academic degree and faculty) influence their self-evaluation competences were analyzed the statistical association between the competences’ self-evaluation and teachers’ personal/ academic information, with the Fisher’s exact test. Table 35 presents the statistical association results between teacher’s gender and the self-evaluation of their competences.

Table 35. Statistical association between teachers’ competences self-evaluation and teachers’ gender

		Male	Female	Fisher Exact Test (df)
		n (%)	n (%)	p-value
Pedagogical Competences	Bad	1 (0,7)	0 (00,0)	$\chi^2(3)=1,489$ $p=,690$
	Sufficient	29 (19,1)	5 (12,5)	
	Good	108 (71,1)	32 (80,0)	
	Very good	14 (9,2)	3 (7,5)	
Professional Competences	Bad	1 (0,7)	0 (0,00)	$\chi^2(3)=1,739$ $p=,589$
	Sufficient	35 (23,0)	13 (32,5)	
	Good	108 (71,1)	25 (62,5)	
	Very good	8 (5,3)	2 (5,0)	
Personal Competences	Sufficient	17 (11,2)	4 (10,0)	$\chi^2(2)=,686$ $p=,688$
	Good	116 (76,3)	29 (72,5)	
	Very good	19 (12,5)	7 (17,5)	
Social Competences	Sufficient	10 (6,6)	2 (5,0)	$\chi^2(2)=,819$ $p=,709$
	Good	110 (72,4)	27 (67,5)	
	Very good	32 (21,1)	11 (27,5)	

Legend: n-number of cases observed; % - Valid percentage; Df - degrees of freedom.

The results of the association test showed that teachers’ gender did not influence teachers’ self-assessment ($p>.05$) which can be explained by the fact that the self-assessment ranks had a low variability between the teachers who participated in the study. The majority of male ($n=108$; 71,1%) and female teachers ($n=32$; 80%) self-evaluated their pedagogical competences as “good”, as well their professional, personal and social competences. The self-evaluation was very homogenous between male and female teachers, with the majority of both with a “good” self-perception.

Table 36 presents the statistical association results between teacher’s age group and the self-evaluation of their competences.

Table 36. Statistical association between teachers' competences self-evaluation and teachers' age

		<35 Years	[35-45] Years	≥ 45 Years	Fisher Exact Test (df)
		n (%)	n (%)	n (%)	p-value
Pedagogical Competences	Bad	1 (4,8)	0 (0,0)	0 (0,0)	$\chi^2(6)=10,613$ $p=,326$
	Sufficient	3 (14,3)	13 (18,6)	18 (17,8)	
	Good	17 (81,0)	50 (71,4)	73 (72,3)	
	Very good	0 (0,0)	7 (10,0)	10 (9,9)	
Professional Competences	Bad	1 (4,8)	0 (0,0)	0 (0,0)	$\chi^2(6)=11,643$ $p=,241$
	Sufficient	4 (19,0)	21 (30,0)	23 (22,8)	
	Good	16 (76,2)	44 (62,9)	73 (72,3)	
	Very good	0 (0,0)	5 (7,1)	5 (5,0)	
Personal Competences	Sufficient	3 (14,3)	9 (12,9)	9 (8,9)	$\chi^2(4)=1,672$ $p=,736$
	Good	15 (71,4)	50 (71,4)	80 (79,2)	
	Very good	3 (14,3)	11 (15,7)	12 (11,9)	
Social Competences	Sufficient	2 (9,5)	6 (8,6)	4 (4,0)	$\chi^2(4)=2,840$ $p=,532$
	Good	16 (76,2)	47 (67,1)	74 (73,3)	
	Very good	3 (14,3)	17 (24,3)	23 (22,8)	

Legend: n-number of cases observed; % - Valid percentage; Df - degrees of freedom.

The results of the association test showed that teachers' age did not influence teachers' self-evaluation ($p>.05$). The competences' self-evaluation was homogenous and with no differences between the three age groups. The majority of teachers with less than 35 years self-evaluated their competences as "good" (among 70% to 80%). In the age from 35 to 44 years there was a "good" self-evaluation of competences (among 60% to 70%). The same was noted in the older group with a "good" self-evaluation in four competences (around 70%).

The teacher's self-evaluation was compared in relation to teacher's time of experience with Anova test (Table 37).

Table 37. Statistical comparison of teachers' competences self-evaluation and teacher's time of experience

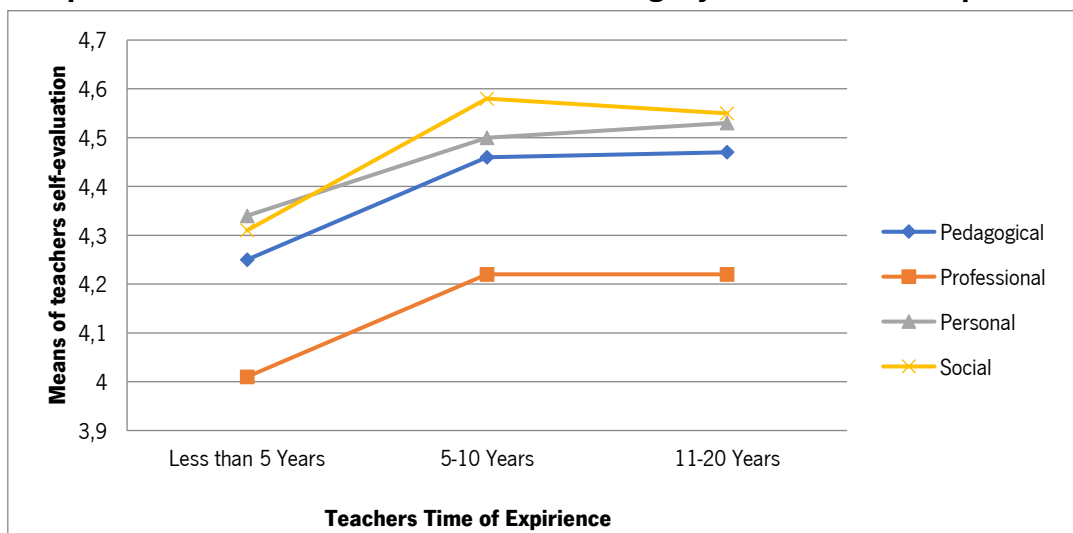
		n	M	SD	F (df) p-value	Differences
Pedagogical Competences	Less than 5 Years	21	4,25	,54	$F(2,189) = 1,817$ $p=0,165$	<i>No differences</i>
	5-10 Years	90	4,46	,48		
	11-20 Years	81	4,47	,48		
Professional Competences	Less than 5 Years	21	4,01	,57	$F(2, 53) = 1,245$ $p=0,296$	<i>No differences</i>
	5-10 Years	90	4,22	,47		
	11-20 Years	81	4,22	,44		
Personal Competences	Less than 5 Years	21	4,34	,56	$F(2,189) = 1,819$ $p=0,165$	<i>No differences</i>
	5-10 Years	90	4,50	,41		
	11-20 Years	81	4,53	,35		
Social Competences	Less than 5 Years	21	4,31	,61	$F(2, 189) = 3,731$ $p=0,026$	Less than 5 Years < 5-10 Years
	5-10 Years	90	4,58	,40		
	11-20 Years	81	4,55	,39		

Legend: n-number of cases observed; M – Mean; SD – Standard deviation; F- Anova test; Df - degrees of freedom.

Teachers' self-evaluation on pedagogical, professional and personal competences was no different between the three periods of length of service. Teacher's self-evaluation regarding social competences was different between teachers with less than 5 years of experience and teachers with 5 to 10 years of experience ($F(2, 189) = 3,731, p < ,05$). UNTL teachers with 5-10 years of experience ($M = 4.58, SD = .40$) has a highest self-evaluation on social competences than teachers with less than 5 years of experience ($M = 4.31, SD = .61$).

Graphic 7 illustrates the means of teachers' self-evaluation ratings by teacher's time of experience. The means of teacher's self-evaluation competences were similar between teachers with different time experiences. Social competences has highest self-evaluated mean in teachers with 5-10 years of experience.

Graphic 7. Means of teachers' self-evaluation ratings by teacher's time of experience



The self-evaluation was equal among different teachers degree in relation to professional, personal and social competences. There were only differences in relation to pedagogical competences (Table 38).

Table 38. Statistical comparison of teachers' competences self-evaluation and teacher's degree

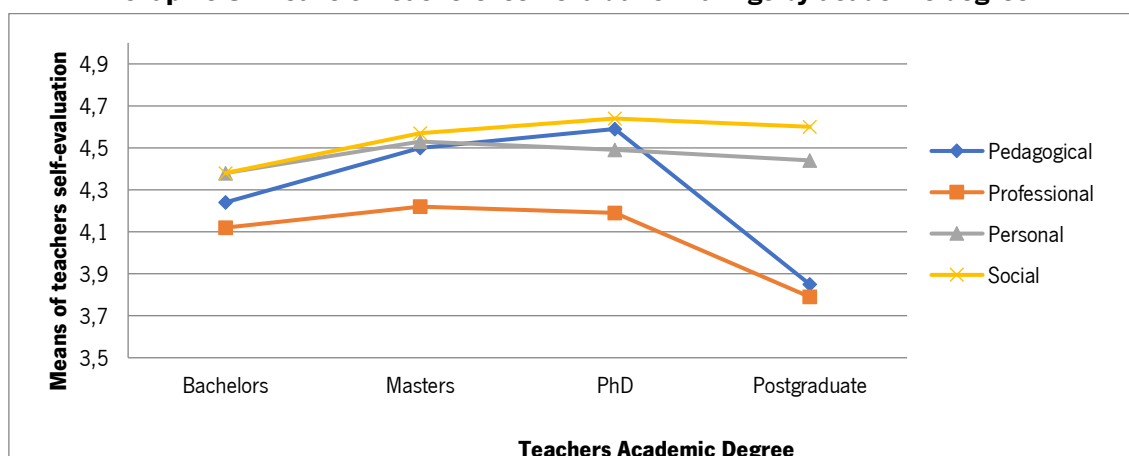
		n	M	SD	F (df) p-value	Differences
Pedagogical Competences	Bachelors	40	4,24	,57	F (3,188) = 2,842 <i>p</i> =0,039	Bachelors<Master
	Masters	138	4,50	,45		Postgraduate <Master
	PhD	11	4,59	,33		Bachelors<PhD
	Postgraduate	3	3,85	,32		Postgraduate <PhD
Professional Competences	Bachelors	40	4,12	,51	F (3,188) = ,439 <i>p</i> =0,725	No differences
	Masters	138	4,22	,47		
	PhD	11	4,19	,26		
	Postgraduate	3	3,79	,47		
Personal Competences	Bachelors	40	4,38	,44	F (3,188) = ,082 <i>p</i> =0,970	No differences
	Masters	138	4,53	,39		
	PhD	11	4,49	,41		
	Postgraduate	3	4,44	,59		
Social Competences	Bachelors	40	4,38	,55	F (3,188) = ,433 <i>p</i> =0,730	No differences
	Masters	138	4,57	,39		
	PhD	11	4,64	,22		
	Postgraduate	3	4,60	,53		

Legend: n-number of cases observed; M – Mean; SD – Standard deviation; F- Anova test; Df - degrees of freedom.

The Anova test revealed significant differences on pedagogical competences' self-evaluation among teachers with different academic degrees (F (3,188) = 2,842; $p < ,05$). The self-evaluation of pedagogical skills was statistically higher among teachers with a PhD degree (M = 4,586, SD = .33) and with a Master's degree (M = 4.50, SD = .45) than teachers with a Bachelors (M = 4.24, SD = .57) and with Postgraduate (M = 3.85, SD = .32). Regarding to professional, personal and social competences there were no differences between teachers with different academic degree ($p > 0,05$). The teachers' competences self-evaluation is equal and independent from their academic degree (bachelor, master, PhD or Postgraduate degree).

Graphic 8 illustrates the means of teachers' self-evaluation ratings by teacher's academic degree. The means' differences were more accentuated in pedagogical competences self-evaluation, with higher ratings among teachers with master or PhD degree.

Graphic 8. Means of teachers' self-evaluation ratings by academic degree



In relation professional competences teachers with a postgraduate level had the lowest self-evaluation score, but these differences were not statistically different.

The parametric ANOVA test was also used to compare teacher's self-evaluation competences in different faculties. Table 39 presents descriptive statistics and ANOVA test to compare teacher's self-evaluation competences between their faculties.

Table 39. Statistical comparison of teachers' competences self-evaluation and teacher's faculty

		n	M	SD	X2 (df) p-value	Differences
Pedagogical Competences	FEAH	30	4,05	,59	F (5, 186)=10,275 $p=0,000$	PS > FEAH PS > M Economy > FEAH
	Political Science	75	4,66	,30		
	Management / Economy	23	4,55	,54		
	Agriculture	38	4,40	,34		
	Medical	23	4,23	,62		
	Philosophy	3	4,11	,19		
Professional Competences	FEAH	30	3,98	,50	F (5, 186)=8,762 $p=0,000$	PS > FEAH Economy > FEAH Economy > Agriculture PS > Agriculture PS > Medicine
	Political Science	75	4,40	,36		
	Management / Economy	23	4,38	,42		
	Agriculture	38	3,98	,34		
	Medical	23	4,03	,65		
	Philosophy	3	3,83	,14		
Personal Competences	FEAH	30	4,32	,48	F (5, 186)=4,651 $p=0,001$	PS > FEAH
	Political Science	75	4,61	,28		
	Management / Economy	23	4,59	,39		
	Agriculture	38	4,36	,38		
	Medical	23	4,53	,55		
	Philosophy	3	4,00	,44		
Social Competences	FEAH	30	4,38	,53	F (5, 186)=4,093 $p=0,002$	PS > Agriculture
	Political Science	75	4,67	,28		
	Management / Economy	23	4,63	,41		
	Agriculture	38	4,38	,35		
	Medical	23	4,52	,62		
	Philosophy	3	4,20	,53		

Legend: n-number of cases observed; M – Mean; SD – Standard deviation; F- Anova test; Df - degrees of freedom.

The faculty had a significant influence on teachers' competences self-evaluation ($F(5, 186) = 8,762, p < .01$). Teachers of the Faculty of Political Science ($M=4.66, SD=.30$) have a statistically higher self-evaluation on their pedagogical competences than teachers of Faculty of Education, Arts and Humanities ($M=4.05, SD=.59$) and teachers of Faculty of Medicine ($M=4.23, SD=.62$). Teachers of faculty of Management and Economy ($M=4.55, SD=.54$) have also a highest self-evaluation on their pedagogical competences than teacher of Education, Arts and Humanities. Regarding professional competences, there were too many differences dependent on teachers'

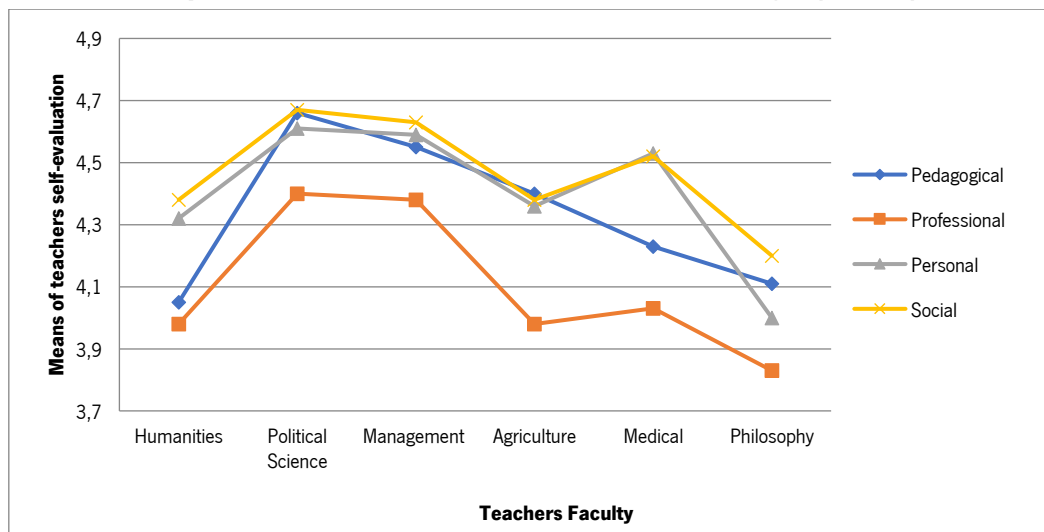
faculty ($F(5, 186) = 8,762, p < .01$). Again teachers of the Faculty of Political Science ($M=4.40, SD=.36$) have a statistically higher self-evaluation than teachers of Faculty of Education, Arts and Humanities ($M=3.98, SD=.50$), faculty of Agriculture ($M=3.98, SD=.34$) and faculty of Medicine ($M=4.03, SD=.65$). Teachers of Faculty of Management and Economy ($M=4.59, SD=.39$) have also a statistically higher self-evaluation than teachers of Faculty of Education, Arts and Humanities and faculty of Agriculture.

The personal competences have significant differences only between teachers of the Faculty of Political Science ($M=4.61, SD=.28$) and Faculty of Education, Arts and Humanities ($M=4.32, SD=.48$) and teachers of humanities have lowest self-evaluation scores than teachers of political science ($F(5, 186) = 4,651, p < 0,01$). UNTL teachers of other faculties have similar self-evaluation scores regarding their personal competences.

Teachers of the Faculty of Political Science ($M=4.67, SD=.28$) have also a statistically higher self-evaluation on social competences than teachers of Faculty of Agriculture ($M=4.38, SD=.35$) ($F(5, 186) = 4,093, p < 0,01$).

Teachers of Faculty of Political Science have a significant higher self-evaluation in all competences. Political Science teachers have then a more positive self-perception of their competences than everyone else. Graphic 9 illustrates the mean scores of teacher's self-evaluation by faculty.

Graphic 9. Means of teachers' self-evaluation ratings by faculty



The mean value of teachers' self-evaluation on pedagogical competences was higher for teachers from Political Science Faculty and lower for Faculty of Education, Arts and Humanities and Philosophy Faculty. The self-evaluation of professional, personal and social competences was statistically higher for Political Science and Management/ Economy Faculties ($p < .01$).

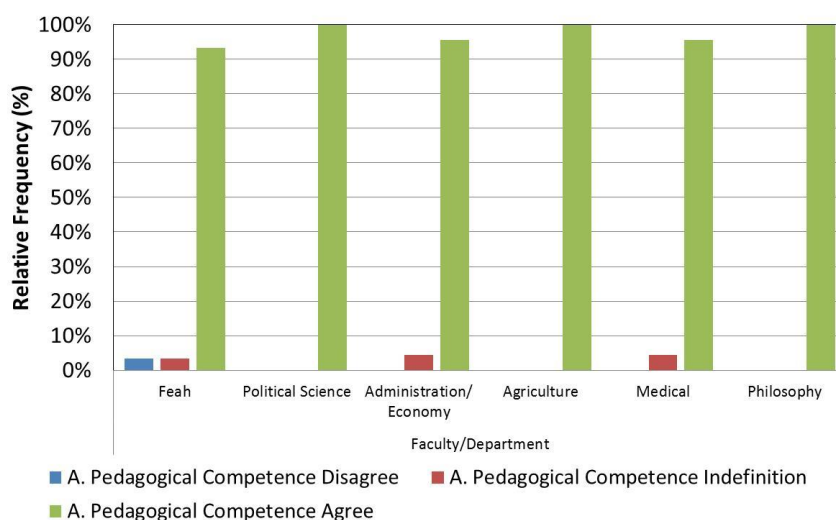
We intended also to analyze the agreement proportion of teachers' self-evaluation in each faculty. The next cross tables represent the proportion of teacher's self-evaluation disagreement and agreement, for each faculty. Table 40 represents the proportion of teacher's self-evaluation disagreement and agreement in relation to pedagogical competences, for each faculty.

Table 40. Proportion of teacher's self-evaluation agreement in relation to pedagogical competences, by faculty

		A. Pedagogical Competence		
		Disagree	Indefinition	Agree
Education, Arts and Humanities	N	1	1	28
	% in faculty	3,3%	3,3%	93,3%
Political Science	N	0	0	75
	% in faculty	0%	0%	100,0%
Administration/ Economy	N	0	1	22
	% in faculty	0%	4,3%	95,7%
Agriculture	N	0	0	38
	% in faculty	0%	0%	100,0%
Medical	N	0	1	22
	% in faculty	0%	4,3%	95,7%
Philosophy	N	0	0	3
	% in faculty	0%	0%	100,0%

There was total agreement in teacher's self-evaluation of their pedagogical competences in Political Science and Agriculture faculties. In other faculties the proportion of agreement was close to 100%. The Bar chart illustrates the proportion of teacher's self-evaluation agreement in relation to pedagogical competences, by faculty.

Graphic 10. Proportion of teacher's self-evaluation agreement in relation to pedagogical competences, by faculty



There were a total of agreement in all faculties (around or equal to 100%) so the differences were not statistically significant, as indicated by the chi-square teste ($\chi^2_{(10)} = 10,238$; $p=0,420$).

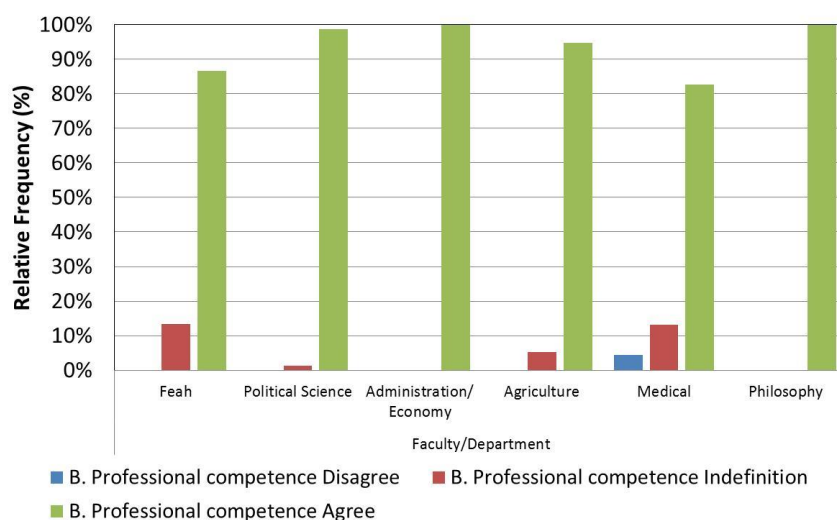
Table 41 represents the proportion of teacher's self-evaluation disagreement and agreement in relation to professional competences, for each faculty.

Table 41. Proportion of teacher's self-evaluation disagreement and agreement in relation to professional competences

		B. Professional Competence		
		Disagree	Indefinition	Agree
Education, Arts and Humanities	N	0	4	26
	% in faculty	0%	13,3%	86,7%
Political Science	N	0	1	74
	% in faculty	0%	1,3%	98,7%
Administration/ Economy	N	0	0	23
	% in faculty	0%	0%	100,0%
Agriculture	N	0	2	36
	% in faculty	0%	5,3%	94,7%
Medical	N	1	3	19
	% in faculty	4,3%	13,0%	82,6%
Philosophy	N	0	0	3
	% in faculty	0%	0%	100,0%

Again, teacher's self-evaluation agreement is near to 100% in all faculties. Bar chart illustrates the proportion of teacher's self-evaluation agreement in relation to professional competences, by faculty.

Graphic 11. Proportion of teacher's self-evaluation agreement in relation to professional competences, by faculty



The teacher's self-evaluation agreement was close to 100% in all faculties, with lowest values in

Medical and Education, Arts and Humanities Faculties, but differences were not statistically significant, as indicated by the chi-square teste ($\chi^2_{(10)} = 18,139$; $p=0,053$).

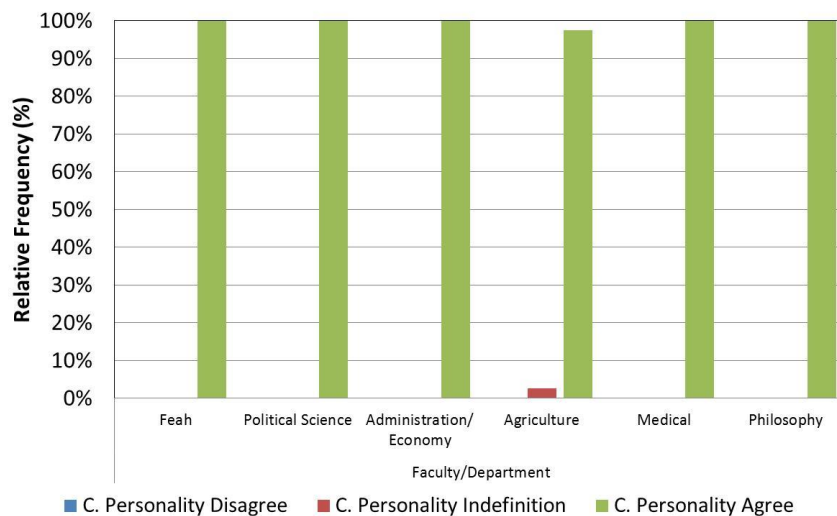
Table 42 represents the proportion of teacher's self-evaluation disagreement and agreement in relation to personal competences, for each faculty.

Table 42. Proportion of teacher's self-evaluation agreement in relation to personal competences

		C. Personal Competences		
		Disagree	Indefinition	Agree
Education, Arts and Humanities	N	0	0	30
	% in faculty	0%	0%	100,0%
Political Science	N	0	0	75
	% in faculty	0%	0%	100,0%
Administration/ Economy	N	0	0	23
	% in faculty	0%	0%	100,0%
Agriculture	N	0	1	37
	% in faculty	0%	2,6%	97,4%
Medical	N	0	0	23
	% in faculty	0%	0%	100,0%
Philosophy	N	0	0	3
	% in faculty	0%	0%	100,0%

The personal competences were total agreement in teacher's self-evaluation in all faculties. Bar chart illustrates the proportion of teacher's self-evaluation agreement in relation to personal competences, by faculty.

Graphic 12. Proportion of teacher's self-evaluation agreement in relation to personal competences, by faculty



The proportion of agreement of teacher's self-evaluated personal competences was close to 100%

in all faculties, so the differences were not statistically significant, as indicated by the chi-square teste ($\chi^2_{(10)} = 4,074$; $p=0,539$).

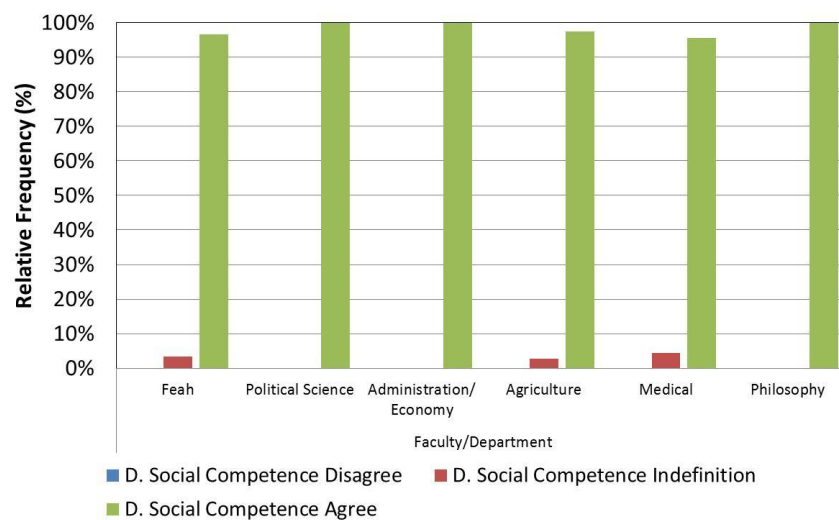
Table 43 represents the proportion of teacher's self-evaluation disagreement and agreement in relation to social competences, for each faculty.

Table 43. Proportion of teacher's self-evaluation disagreement & agreement in relation to social competences

		D. Social Competences		
		Disagree	Indefinition	Agree
Education, Arts and Humanities	N	0	1	29
	% in faculty	0%	3,3%	96,7%
Political Science	N	0	0	75
	% in faculty	0%	0%	100,0%
Administration/ Economy	N	0	0	23
	% in faculty	0%	0%	100,0%
Agriculture	N	0	1	37
	% in faculty	0%	2,6%	97,4%
Medical	N	0	1	22
	% in faculty	0%	4,3%	95,7%
Philosophy	N	0	0	3
	% in faculty	0%	0%	100,0%

In relation to social competences, there was a high level of concordance (close to 100%) among teachers independent from faculty. Bar chart illustrates the proportion of teacher's self-evaluation agreement in relation to social competences, by faculty.

Graphic 13. Proportion of teacher's self-evaluation disagreement & agreement in relation to social competences, by faculty



The proportion of agreement of teacher's self-evaluated social competences was close to 100% in

all faculties, so the differences were not statistically significant, as indicated by the chi-square teste ($\chi^2_{(10)} = 3,657$; $p=0,600$).

In the second stage of teachers self-evaluation analysis, there were studied the level of agreement between the all pairs of self-evaluated competences, with a contingency cross table. Table 44 shows the proportion of agreement between the teacher's self-evaluation of pedagogical and professional competences.

Table 44. Proportion of agreement between the teacher's self-evaluation of pedagogical and professional competences

A. Pedagogical Competences		B. Professional Competences		
		Disagree	Indefinition	Agree
Disagree	N	0	0	1
	% in group	0%	0%	100,0%
Indefinition	N	0	2	1
	% in group	0%	66,7%	33,3%
Agree	N	1	8	179
	% in group	,5%	4,3%	95,2%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with "A. Pedagogical Competence" show a percentage of 100% (only one element) of agree with "B. Professional Competence", the ones that show indefiniteness with "A. Pedagogical Competence" show a percentage of 33,3% of agree with "B. Professional Competence", the ones that agree with "A. Pedagogical Competence" show a percentage of 0,5% of disagree and 4,3% of indefiniteness with "B. Professional Competence", however there is a statistically significant agreement, as indicated by the measure of agreement of 24,7% specified by the kappa coefficient ($p<0,001$).

Table 45 shows the proportion of agreement between the self-evaluation of pedagogical and personal competences.

Table 45. Proportion of agreement between the teacher's self-evaluation of pedagogical and personal competences

A. Pedagogical Competences		C. Personal Competences		
		Disagree	Indefinition	Agree
Disagree	N	0	0	1
	% in group	0%	0%	100,0%
Indefinition	N	0	0	3
	% in group	0%	0%	100,0%
Agree	N	0	1	187
	% in group	0%	,5%	99,5%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “A. Pedagogical Competence” show a percentage of 100% (only one element) of agree with “C. Personality”, the ones that show indefiniteness with “A. Pedagogical Competences” show a percentage of 100% of agree with “C. Personality”, the ones that agree with “A. Pedagogical Competence” show a percentage of 0,5% of indefiniteness with “Personal Competences”, but the statistically agreement cannot be determined by the kappa coefficient due to the reduced number of observations for non-agreement.

Table 46 shows the proportion of agreement between the self-evaluation of pedagogical and social competences.

Table 46. Proportion of agreement between the teacher’s self-evaluation of pedagogical and social competences

A. Pedagogical Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N	0	0	1
	% in group	0%	0%	100,0%
Indefinition	N	0	0	3
	% in group	0%	0%	100,0%
Agree	N	0	3	185
	% in group	0%	1,6%	98,4%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “A. Pedagogical Competence” show a percentage of 100% (only one element) of agree with “D. Social Competence”, the ones that show indefiniteness with “A. Pedagogical Competence” show a percentage of 100% of agree with “D. Social Competence”, the ones that agree with “A. Pedagogical Competence” show a percentage of 1,6% of indefiniteness with “D. Social Competence”, but the statistically agreement cannot be determined by the kappa coefficient due to the reduced number of observations for non-agreement.

Table 47 shows the proportion of agreement between the self-evaluation of professional and personal competences.

Table 47. Proportion of agreement between the teacher's self-evaluation of professional and personal competences

B. Professional Competences		C. Personal Competences		
		Disagree	Indefinition	Agree
Disagree	N	0	0	1
	% in group	0%	0%	100,0%
Indefinition	N	0	1	9
	% in group	0%	10,0%	90,0%
Agree	N	0	0	181
	% in group	0%	0%	100,0%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “B. Professional Competence” show a percentage of 100% (only one element) of agree with “C. Personality”, the ones that show indefiniteness with “B. Professional Competence” show a percentage of 90,9% of agree with “C. Personality”, the ones that agree with “B. Professional Competence” also agree with “C. Personality”, but the statistically agreement cannot be determined by the kappa coefficient due to the reduced number of observations for non-agreement. Table 48 shows the proportion of agreement between the self-evaluation of professional and social competences.

Table 48. Proportion of agreement between the teacher's self-evaluation of professional and social competences

B. Professional Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N	0	1	0
	% in group	0%	100,0%	0%
Indefinition	N	0	1	9
	% in group	0%	10,0%	90,0%
Agree	N	0	1	180
	% in group	0%	,6%	99,4%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that disagree with “B. Professional Competence” show a percentage of 100% (only one element) for indefiniteness with “D. Social Competence”, the ones that show indefiniteness with “B. Professional Competence” show a percentage of 90,9% of agree with “D. Social Competence”, the ones that agree with “B. Professional Competence” show a percentage of 0,6% of indefiniteness with “D. Social Competence”, but the statistically agreement cannot be determined by the kappa coefficient due to the reduced number of observations for non-agreement.

Table 49 shows the proportion of agreement between the self-evaluation of personal and social competences.

Table 49. Proportion of agreement between the teacher’s self-evaluation of personal and social competences

C. Personal Competences		D. Social Competences		
		Disagree	Indefinition	Agree
Disagree	N			
	% in group			
Indefinition	N	0	1	0
	% in group	0%	100,0%	0%
Agree	N	0	2	189
	% in group	0%	1,0%	99,0%

The bold percentages in the table and the gradation columns in the graphic show the similar responses in each dimension. For the dissimilar responses, the ones that show indefiniteness with “C. Personality” give the same appreciation to “D. Social Competence”, the ones that agree with “C. Personality” show a percentage of 100% of indefiniteness with “D. Social Competence”, however there is a statistically significant agreement, as indicated by the measure of agreement of 49,6% specified by the kappa coefficient ($p < 0,001$).

3.3 Comparison between the students’ evaluation and teachers’ self-evaluation on teachers’ competences

The UNTL students evaluation about teachers competences (N=342) were analyzed and compared with the UNTL teachers self-evaluation (N=192). At the first place, there was a more positive self-perception of teachers than the student’s evaluation, as noted by Belo (2016). Teacher’s self-evaluation scores were higher than student’s evaluation.

The UNTL teachers evaluated in a very positive way, between "sufficient" and "very good"; only two teachers registered a classification of "bad" in relation to pedagogical and professional competences. In the evaluation of the students, the classification of bad had a more expressive percentage, superior to 10% in each one of the four groups of competences.

Regarding the pedagogical competences of the teachers, the lowest classification attributed by the students was in relation to item "7. Attribution of feedback on the work / evaluation "and in relation to" item 5. Providing the means and pedagogical technology "(M = 3.36, SD = 1.20). For students, these are the teacher’s pedagogical competences that need to be improved.

Teachers self-evaluate their theological competences and the ability to guide students as "good." The ability to use the various communication technologies (M = 3.38, SD = 1.13) as well as the

ability to present interdisciplinary explanations about a given topic ($M = 3.48$, $SD = 1.01$) get the lowest scores assigned by students. The overall assessment of students was at level 3 ("sufficient") while teacher self-assessment recorded an overall average of 4 ("good").

For teachers, professional skills with less evaluation were those that referred to scientific research, both from the point of view of teaching and the ability to involve students in scientific and technical research in organizations.

From the perspective of the teacher, the lower average classifications referred to the ability to achieve the advancement of science and technology to enhance the learning process ($M = 3.99$, $SD = .86$), the ability to be involved in scientific work and in the professional organizations ($M = 3.64$, $SD = .82$) and the students' ability to engage in scientific and technological work ($M = 3.64$, $SD = .82$). Since these were the competencies with lower self-evaluation ratings of teachers, it was concluded that they should be prioritized in terms of improvement and professional development practices.

For the perspective of the students, teachers' competences in the area of scientific research, production and knowledge sharing also registered lower ratings in the group of professional skills. Students thus converged their assessment classifications with teacher self-rating classifications by considering only as "sufficient" the ability of teachers to use research results in promoting learning and engaging students in research studies.

In terms of personality competencies, we again observed a difference between the classifications of the students and the classifications of the teachers. The students assessed each of the skills with level 3 ("sufficient") and in some cases close to level 4 ("good"), particularly with respect to dignity, personality, justice, exemplary behavior and self-confidence. In the self-assessment study, teachers were even more positive since the average for each competency was at level 4 ("good") and in some cases very close to level 5 ("very good") which referred to dignity, personality, sense of neutrality and justice and self-confidence.

The social skills of teachers were classified by students with level 4 ("good") and by teachers with a difference of one unit in the average of each competency. The students emphasized the religious tolerance of teachers ($M = 3.81$, $SD = 1.08$), as well as the ability to accept students' criticisms and suggestions ($M = 3.62$, $SD = 1.09$) and knowledge of students in each class ($M = 3.62$, $SD = 1.10$). In the self-evaluation of teachers, the most valued competences were also religious tolerance ($M = 4.60$, $SD = .56$) and acceptance of criticisms and suggestions of students ($M = 4.56$, $SD = .57$). The capacity of expression was the social competence of the teachers that registered among

the students the lowest classification ($M = 3.50$, $SD = 1.12$), while for teachers, besides the capacity of expression ($M = 4.49$; 58) also had a lower social rank ($M = 4.47$, $SD = 58$), which means that there are two skills that need to be improved.

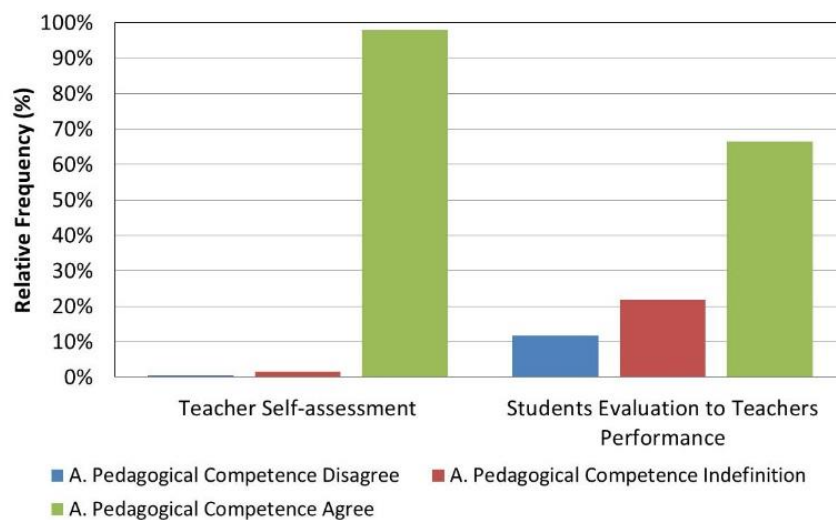
It was intended also to analyze the proportion of agreement between students' evaluation and teachers' self-evaluation in relation to each competence. The next cross tables represent the proportion disagreement and agreement between the students' evaluation and the teacher's self-evaluation, by competences. Table 50 shows the proportion of disagreement and agreement of students' and teacher's evaluation in relation to pedagogical competences.

Table 50. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to pedagogical competences

		A. Pedagogical Competences		
		Disagree	Indefinition	Agree
Teacher Self-assessment	N	1	3	188
	% in group	,5%	1,6%	97,9%
Students Evaluation	N	40	75	227
	% in group	11,7%	21,9%	66,4%

Graphic 14 represents the proportion of agreement between students' evaluation and teachers' self-evaluation in relation to pedagogical competences.

Graphic 14. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to pedagogical competences



The proportion of agreement higher for teachers' self-assessment and the percentages that disagree or show indefiniteness is higher for students' evaluation, and differences were statistically significant, as indicated by the chi-square teste ($\chi^2_{(2)} = 70,665$; $p < 0,001$).

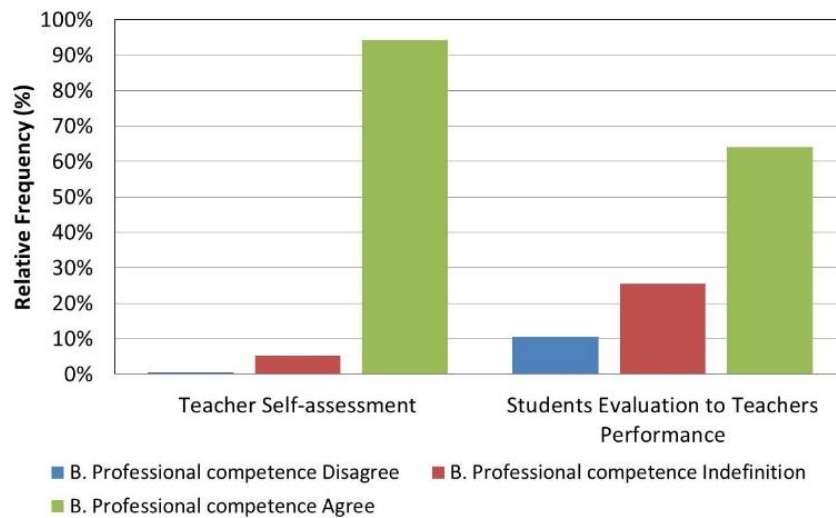
Table 51 shows the proportion of disagreement and agreement of students' and teacher's evaluation in relation to professional competences.

Table 51. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to professional competences

		Professional competences		
		Disagree	Indefinition	Agree
Teacher Self-assessment	N	1	10	181
	% in group	,5%	5,2%	94,3%
Students Evaluation	N	36	87	219
	% in group	10,5%	25,4%	64,0%

Bar chart 15 represents the proportion of agreement between students' evaluation and teachers' self-evaluation in relation to professional competences.

Graphic 15. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to professional competences



The percentages that agree is higher for Teacher Self-assessment and the percentages that disagree or show indefiniteness is higher for Students Evaluation, and the differences are statistically significant, as indicated by the chi-square teste ($\chi^2_{(2)} = 60,479$; $p < 0,001$).

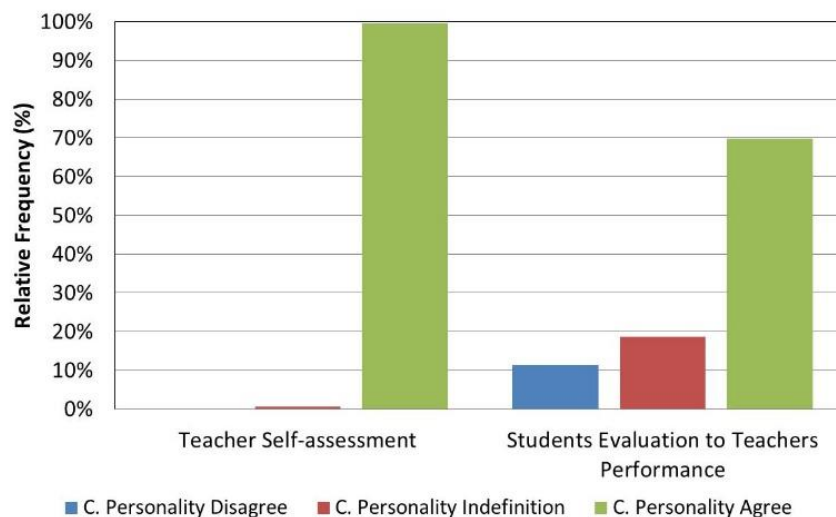
Table 52 shows the proportion of disagreement and agreement of students' and teacher's evaluation in relation to personal competences.

Table 52. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to personal competences

		B. Personal competences		
		Disagree	Indefinition	Agree
Teacher Self-assessment	N	0	1	191
	% in group	0%	,5%	99,5%
Students Evaluation	N	39	64	239
	% in group	11,4%	18,7%	69,9%

Graphic 16 represent the proportion of agreement between students' evaluation and teachers' self-evaluation in relation to personal competences.

Graphic 16. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to personal competences



The percentages that agree is higher for Teacher Self-assessment and the percentages that disagree or show indefiniteness is higher for Students Evaluation, and the differences are statistically significant, as indicated by the chi-square teste ($\chi^2_{(2)} = 68,706$; $p < 0,001$).

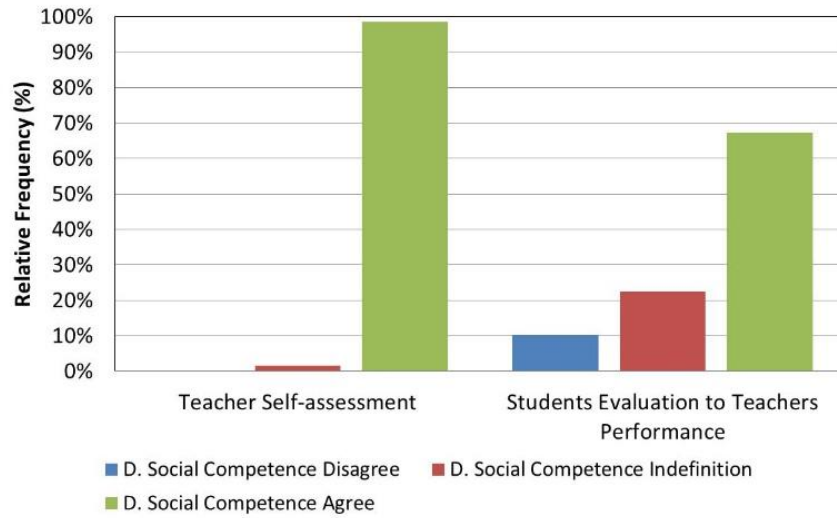
Table 53 shows the proportion of disagreement and agreement of students' and teacher's evaluation in relation to social competences.

Table 53. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to social competences

		B. Social competences		
		Disagree	Indefinition	Agree
Teacher Self-assessment	N	0	3	189
	% in group	0%	1,6%	98,4%
Students Evaluation	N	35	77	230
	% in group	10,2%	22,5%	67,3%

Bar chart represents the proportion of agreement between students' evaluation and teachers' self-evaluation in relation to social competences

Graphic 17. Proportion of agreement between students' evaluation and teachers' self-evaluation in relation to social competences



The percentages that agree is higher for Teacher Self-assessment and the percentages that disagree or show indefiniteness is higher for Students Evaluation, and the differences are statistically significant, as indicated by the chi-square teste ($\chi^2_{(2)} = 70,923$; $p < 0,001$). Again, for all the dimensions, the Students Evaluation to Teachers Performance is significantly lower than the Lecturers/ Teachers Self-Assessment.

CHAPTER IV – QUALITATIVE DATA ANALYSIS RESULTS

The qualitative study of the case study of UNTL, the public university of East Timor, intends to describe how the Teaching Performance Assessment (TPA) is being implemented in the educational institution: to identify the attitudes towards the TPE in Timorese higher education system, like the obstacles and the contributions to performance and teaching development; to know the teacher's experiences in the implementation of TPE in Timorese higher education system; to identify the teaching and learning competences should be developed by teachers in a professional development program; draw suggestions and recommendations from UNTL teachers in relation to effective implementation of TPE in higher education system. These findings will allow us to identify the teaching and learning competences should be developed by teachers in a professional development program, as well teachers' suggestions and recommendations to carry a better teacher performance evaluation in Timorese higher education system.

The study involves the participation of 16 UNTL teachers who holding management positions at UNTL at the time of the interview. From the 16 participants there were 14 male and two female teachers, so the majority of UNTL teachers was male. Teachers' age varies between 35 (minimum) and 66 years (maximum). The mean age was 50 years (Mean=49,625 ± 8,42) and had a coefficient variation of 17% revealing variability and heterogeneous ages among UNTL teachers. In relation to their professional background, length of service varies between two (minimum) and 16 years (maximum) – four teachers had 16 years of experience as a teacher, one of them has 15 years of experience, other has 11 years, four teachers had a length of service of 10 years and the others had less than 10 years. The mean length of service was 10 years (Mean=10,438 ± 4,63) and had a higher coefficient variation of 44%.

None of the participants have previous experience in the area of teacher performance evaluation. In terms of their educational background, 11 teachers have a master degree while only five have a PhD/Doctorate degree, and no other academic degree were observed (table 54).

Table 54. Sociodemographic, professional and academic background characteristics of the interviewees

No. of Interviewees	Personal background		Professional background		Academic Formation	
	Gender	Age	Length of service	Experience in TPE	Educational degree	Other academic background
E1	Male	42	10	None	PhD/Doctorate	None
E2	Male	47	16	None	Master	None
E3	Male	57	13	None	Master	None
E4	Male	48	10	None	PhD/Doctorate	None
E5	Male	54	16	None	Master	None
E6	Male	57	6	None	Master	None
E7	Male	45	11	None	Master	None
E8	Male	52	16	None	Master	None
E9	Male	43	10	None	Master	None
E10	Female	66	10	None	PhD/Doctorate	None
E11	Male	42	2	None	Masters	None
E12	Male	65	5	None	Masters	None
E13	Male	47	15	None	Masters	None
E14	Male	47	6	None	PhD/Doctorate	None
E15	Male	47	16	None	PhD/Doctorate	None
E16	Female	35	5	None	Masters	None

The qualitative study was conducted through the follow matrix of analysis (table 55). There were analyzed three main categories: the recognition and importance attributed to the Teacher Performance Evaluation (TPE); attitudes and the positioning of UNTL teachers towards TPE; Practices for improving teacher performance. Each of three categories includes subcategories that emerged from the interviews' content.

Table 55. Matrix of qualitative analysis

Category	Subcategories
Recognition and importance attributed to the Teacher Performance Evaluation (TPE)	A. Contributions from TPE B. Documents to the existing TPE in the Institution (Regulations and Legislation) C. Implementation of the TPE in the Teaching Institution
Attitudes and the positioning of UNTL teachers towards TPE	A. Obstacles identified for its implementation B. Recommendations
Practices for improving teacher performance	A. Valuing the formative/training component B. Examples of practices that enhance the teaching and learning process / academic success

The results of the analysis are presented in tables, in which the interviewees, in the total of 16 UNTL teachers with, are designated E1, E2, E3, E4 E16, followed by a synthesis of the most relevant aspects extracted from the interviews.

4.1. Category 1: Recognition and importance attributed to the Teacher Performance Evaluation (TPE)

Subcategories

- A. Contributions from TPE
- B. Documents to the existing TPE in the Institution (Regulations and Legislation)
- C. Implementation of the TPE in the Teaching Institution

Table 56. Recognition and importance attributed to the Teacher Performance Evaluation (TPE)

Category	Subcategories	Indicators	No. of respondents
Recognition and importance attributed to the Teacher Performance Evaluation	A. Contributions from TPE	Improving teacher performance	E1; E5; E6; E14
		Professional development through training in pedagogical skills	E2; E4; E5; E6; E14
		Decision-making to continue or not to teach	E3
		More careful selection of teachers	E4
		Identification of the strengths and weaknesses of teaching practice	E6; E15
	B. Documents for the TPE existing in the Teaching Institution (Regulations and Legislation)	Legislation exists (not specified) but not enforced	E7
		No specific legislation / Not known	E6; E9; E16
		Public Service Commission / General regime from the civil service	E2; E5; E6; E14; E16
		Certification Manual, which contains some items for evaluation	E2; E3
		Regulation for Quality Control	E1; E15
		Statute of UNTL	E9
		Code of Conduct	E5; E9
		Special Career Regime/Statute of the Teaching Career (Career Regime Law, General Career Regime or Special Career Regime / Career Regime), where explicit pedagogical and investigation procedures are found.	E3; E8; E13; E14; E15
		Decree-Law no. 13/2014, of January 15 (recommends the creation of a University Teaching Certification Office - GAB CEDU) by the University	E11
	C. Implementation of TPE in the teaching institution	Is not implemented	E1; E2; E3; E4; E6; E7; E8; E9; E10; E12; E13; E14; E16
		Self-assessment carried out by teachers in order to identify the existing gaps (2012)	E8; E10
		Monitoring system for attendance and punctuality of teachers, but it does not apply to all faculties and to all teaching classes, without consequences in terms of continuity in teaching.	E13

Category	Subcategories	Indicators	No. of respondents
		The performance evaluation of UNTL teachers and civil servants is based on the standard form of the Public Service Commission of Timor-Leste, however it is not the most appropriate.	E15; E16
		Establishment of the CPAI (Institutional Standing Evaluation Committee) approved by Law No. 21 dealing with the TPE.	E6; E9
		A process of identification of teachers is in progress, with gaps in pedagogical skills, for subsequent participation in continuous training programs (SWOT analysis)	E6

As shown in table 56, respondents attribute the professional development of the teacher (E1, E5, E6, E14) as a main contribution of the Performance Evaluation, through participation in pedagogical training (E2, E4, E5, E6 and E14). They also highlight their relevance for decision-making on whether or not to continue teaching (E3) and for the selection of teachers when applying for university admission (E4).

While some respondents state that there is no specific legislation in the area (E6, E9, E16), others refer to some documents and legislation of a general nature, which can be considered as precursors and guiding principles, namely the General Regime issued by the public service (E2, E6, E14, E16) and the Special Career Regime (E3, E13, E14, E15), among others. Thus, it is apparent that the evaluation of teacher performance is currently governed by the principles defined by the Public Service Commission of Timor-Leste (E15, E16), general and non-specific for evaluation of teaching practice, and other existing documents / regulations in the University, such as the Code of Conduct and the Certification Manual. Respondent 6 states that a process is under way to identify existing gaps in pedagogical competencies in order to promote the participation of teachers in continuing education programs.

They also present specific procedures for performance evaluation, which are more procedural and bureaucratic in nature, namely, the existence of a system for monitoring attendance and punctuality of teachers (E13), but not applied to all faculties and teachers, and an experience occurred in 2012 at the level of self-assessment (E8; E10). The creation of the CPAI (Standing Evaluation Committee Institutional) referred to by E6 and E9 could be a driving force for the implementation of the TPE. Given the above, we can affirm that respondents recognize and attach significant importance to TPE, although it is not yet implemented in the University.

4.2. Category 2: Attitudes / Positioning towards TPE

Subcategories

A. Obstacles identified for its implementation

B. Recommendations

Table 57. Attitudes towards TPE

Category	Subcategories	Indicators	Nº of respondents
Attitudes towards TPE	A. Obstacles identified for its implementation	Not found necessary conditions (non-specific) for the implementation of the TPE	E7
		Resistance by teachers due to their stay in the institution only on a part-time basis	E1
		Conception of teacher as mere transmitter of information	E4; E5
		I am afraid to be evaluated because of the lack of knowledge of the methods and not recognizing the competence of the potential evaluators	E4; E7; E16
		Existence of few graduates, namely doctorates	E4; E11
		Use of the Portuguese Language	E4; E6; E7; E16
		Lack of interest to learn, to deepen knowledge, even when there are initiatives, do not seize the opportunity	E4
		Evaluation based on opinions rather than evidence	E16
		Strategic Plan of Institution Management	E9
		Lack of a culture of scientific writing	E4; E16
		Lack of specialists in the field of performance appraisal	E10; E12
	B. Recommendations and suggestions	TPE held annually in order to create spaces for the formation of good qualifications and not to dismiss teachers for their poor performance.	E2
		Creation of an independent team to evaluate in order to avoid bias at trial (external evaluation)	E3
		Self-assessment system within the faculties, which also allows self-assessment.	E4; E6
		TPE should be conducted by experts in the field. "Need for an international expert in the area of TPE to work with us to establish the CPAI and get the job done. We have a small commission as part of ANAAA composed of three people with the task of looking at the issue "	E5; E6; E12; E13; E6
		Encourage the creation of a collaborative and learning culture at the level of teachers.	E4; E5
		Incentive to research	E5; E11
		Assessment of the teacher by the students at the end of the school year.	E10; E14

Category	Subcategories	Indicators	Nº of respondents
		Teachers need support / training to understand how to make good TPE.	E16
		Teachers must obey to the rules and regulations. Those who do not comply with the law must be sanctioned	E8; E15
		TPE based on the institution's official dispatch from the higher management level.	E13
		Periodic assessment with observation of classes without previous communication.	E9
		Team / commission responsible for evaluation at UNTL	E9; E10; E11; E14
		Evaluation team in close collaboration with the Vice Rector for academic affairs.	E9
		Elaboration of form / grid by each department, to measure the performance of teachers, containing indicators related to pedagogical, professional, personal and social skills	E11; E15
		Close collaboration between the Ministry of Education and the University in order to clarify the procedures to be followed.	E7; E16

Asked about obstacles to the implementation of the TPE (See Table 57) at the University, the respondents, in addition to considering that the necessary conditions have not yet been met, without specifying (E7), refer to the fear of the teaching class to be evaluated, which is on the one hand, the lack of knowledge of the methods to be used (E4) and, on the other hand, the lack of specialists in the field of performance evaluation (E10; E12) and non-recognition of the competence of potential evaluators (E7; E16). The use of the Portuguese language is also presented as an obstacle (E4, E6, E7, E16), because although it is mandatory, there are teachers who do not yet use it in the context of the classroom. Some respondents consider that there is no investment in this area by top leadership / management (E9) and a scientific writing culture (E4; E16).

Other obstacles are presented, among which we highlight: traditional teacher conception (focus on the teacher rather than the student) (E4; E5); Shortage of graduates, namely Doctorates (E4; E11) and the lack of interest shown by some teachers to deepen knowledge in the teaching and learning process (E4). Thus, in order to address some of these obstacles, respondents suggest that the TPE should be conducted by experts in the field (E5, E12, E13), another option is to set up a responsible evaluation team / commission (E9, E10, E14), with the possibility of resorting to external evaluation (E3), in order to avoid bias at judging, and the specialized training of evaluators (E16) (internal and external).

The need for guidance documents (E8; E15) emanating from the upper management level (E13), as well as the rigorous selection / design of data collection instruments (E11; E15) to be used (such as observation grids, lists of verification, classification scales, record maps) were also accentuated. The encouragement of self-assessment (E6), observation of classes (E9), research (E5, E11), student involvement (E10, E14), and collaboration (E4, E5, E16) between the educational agents / agents (the Ministry of Education itself, the University as a whole, the teaching class), was also mentioned.

4.3. Category 3: Practices for improving teacher performance

Subcategories

- A. Valuing the formative/training component
- B. Examples of practices that enhance the teaching and learning process / academic success

Table 58. Practices for improving teacher performance

Category	Subcategories	Indicators	Nº of respondents
Practices for improving teacher performance	Valuing the formative/training component	Ongoing training in deficit areas, especially in the pedagogical field	E2; E3; E4; E6 E7; E8; E9; E10; E13; E14; E15
		ISO Formation	E1
		Training of Trainers (TOT) developed by United Nations Organization (UNO)	E4
		Training in the pedagogical area carried out in collaboration with the Calouste Gulbenkian Foundation.	E14
		Obtaining a certificate IV before starting the teaching profession.	E7
		Creation of discussion groups on pedagogical issues with the presence of experts in the area.	E4
	B. Examples of practices that enhance the process of teaching and learning / academic success	Use of diversified / mixed teaching methodologies	E2
		Development of critical / reflective thinking of students.	E5
		Involvement of students in the teaching and learning process.	E5
		Balance between learning in classroom context and practical / field work (in real context).	E9
		Knowledge sharing / collaborative teacher work.	E5
		Informal observation of daily tasks / activities.	E2
		Reward teachers for their good performance.	E2

Concerning the practices that promote the improvement of teaching performance (see Table 58), the respondents were observed to refer mainly to participation in training programs in the

pedagogical field (E2; E3; E4; E6 E7; E8; E9; E10; E13; E14; E15) and the stimulation of discussion groups with the presence of specialists (E4), in order to train teachers with appropriate methodologies and strategies to intervene in the classroom context in order to enhance students' learning and, consequently, their academic success. Respondents cite programs / projects where they participated (ISO - E1 Training, Training for Trainers (TOT) developed by the UN; Training in the pedagogical area carried out in collaboration with the Calouste Gulbenkian Foundation - E14), which constitute proposals for training programs and partnerships to be implemented. It was also emphasized by E7 the importance of obtaining certificate IV before joining as a professor at the University.

As examples of teaching and learning / academic success teaching practices, the following were presented: diversification of teaching and learning methodologies (E2) that promote greater student involvement (E5) and the development of critical and reflective thinking (E5). The balance between theoretical knowledge / know-why and know-how (E9) and knowledge sharing among teachers (E5), through the informal observation of daily tasks (E2) by peers and / or hierarchical superiors, are also references of good habits. Finally, the importance of recognizing / rewarding teachers for their good performance, giving them more time for training and research work, should be stressed. In summary, according to the analysis made to the interviewees' opinions, the TPE is an appropriate strategy for the diagnosis, monitoring, reflection and continuous improvement of pedagogical practices, following strict and systematic procedures based on guided documents built for this purpose. In this sense, it should be conceived as an instrument for the benefit of teachers, their performance and their professional development. However, it should be emphasized that its implementation will not always be peaceful and consensual, and may in some cases lead to insecurity, skepticism and rejection.

In the analysis of content made to the interviewees' discourses, specific cases were also identified (punctual, not included in the tables previously presented), which will be mentioned in the following: Interviewee 11 (E11) refers to the existence of the SNIC, which evaluates the scientific production, and an Advanced Training Center for the development of actions within the teaching and learning process.

The same interviewee (E11) considers that the greatest challenge in the TPE is compliance with the law and presents in detail the TPE process:

Each teacher will be evaluated through 10 surveys, including self-assessment (self-assessment), so being an organization with more than 400 teachers we are talking about at least 4000 inquiries,

which will not be an easy task. The first gap identified is the lack of a system, an online platform that allows students and teachers to be related and enables an effective and immediate implementation of the surveys. The Faculty of Engineering, through the Department of Informatics, will probably already implement the surveys through an academic information system, with later possibility of application in the remaining faculties of UNTL. Moreover, the TPE, according to the law, is carried out through two systems, that is, each teacher is evaluated in two ways. The first that is called perception evaluation, in this evaluation of perception the teacher is evaluated, say that empirically, there are no grids, there is no rigor, students fill out the questionnaire through an idea that they have about 5 students. There are 3 peer colleagues from the teacher who will also evaluate the teacher: a boss, a department director and the teacher himself. The teacher will be assessed on a scale of 1 to 5. The teacher will need to get a minimum grade of 3.5 to pass. The other form or modality of evaluation, the second, is through the so-called Personal Description Instrument that is based on the three pillars of higher education: teaching, research and service to the community. Basically, the teacher will make a personal description of his / her activities on the Ministry's own form, which indicates aspects such as his / her creative effort to improve the quality of learning and the impact of change, i.e. scientific works, work goals for the development of science, institutional support, impact on change, and support for society. After this "personal discussion," Self-rating, the evaluation will also be made by the superior, probably by the department director, who will assign the teacher a rating according to a grid. This report should have a minimum score of 3 on a scale of 1 to 5 to be approved. Once these procedures have been completed, the consistency of the two evaluations will be evaluated and a final result will be obtained by comparing them. Thus, the TPE presupposes the allocation of the number of credits, in which the points awarded will contribute to the certification of the teacher for the purposes of his career profession (or career development).

E11 also presents a pilot project under the TPE:

The Office of the Vice Rector for Academic Affairs implemented for the first time in the UNTL a pilot project only for the first years of the courses. It consists in the study of the students' perceptions not only in relation to the performance of the teachers, but also to the curricular units, to the faculty and to the institution in general, it was a more general survey, a more comprehensive evaluation. A report was then prepared and submitted. This evaluation would take place every year, however, UNTL is faced with technical difficulties, because the University does not yet have an adequate information system. The University needs to register all teachers, students, courses, use data in

the information system, which allows students to evaluate teachers through an information system, in order to get immediate results.

Thus, two types of evaluation were carried out, one by a closed questionnaire addressed to the students, which evaluated the aspects of the TPE, the validity of the course, the faculty and the university, and then implemented a questionnaire addressed to the teachers to give examples of activities at the pedagogical, scientific level, at the level of their teaching practice, examples of evaluation tests, of the work done by the students. Less experienced teachers, who taught only in the first years of the UNTL courses, are not part of the cadres who are hired, teachers who have been licensed only 2 or 3 years, and teachers with little experience. In this assessment of teachers, 40 aspects were identified that teachers should improve, presented at a workshop held with faculty members, departmental directors and faculty deans. It was also stressed the need to implement a higher quality in teaching, especially in evaluation, since the most critical point where teachers fail more and more is in the assessment made to the students.

According to E16, within the scope of this pilot project, there is a need to review the instruments / questionnaires used in order to make them more adequate to the context of the teaching practice.

CHAPTER V – DATA ANALYSIS AND INTERPRETATION

The study was carried out using both quantitative and qualitative analysis to determine: (1) how students evaluate the UNTL teachers' pedagogical, professional, personal and social competences and how students' background influence their evaluation; (2) how teachers self-evaluated their own competences and how their background characteristics influence their self-evaluation; (3) to what extent does students' evaluation of UNTL teachers and the UNTL teachers' self-evaluation competences differ or converge; and (4) how is the Teaching Performance Assessment (TPE) being implemented in the UNTL. Therefore, researcher would like to present the analysis based on the determined research questions, objectives and hypotheses.

5.1. Research questions and objectives

5.1.1. How do students evaluate the UNTL teachers' competences?

With this question – How do students evaluate the UNTL teachers' competences? – we attempt three objectives: 1) to describe the students' evaluation of UNTL teachers' competences, namely pedagogical, professional, personal and social competences; 2) to identify the highest teacher's competences evaluated by students in each set of competences and 3) to identify the lowest UNTL teacher's pedagogical, professional, personal and social competences according students and so need improvement.

Objectives:

- To describe the students' evaluation of UNTL teachers' competences;
- To identify the highest teacher's competences evaluated by students in each set of competences;
- To identify the lowest teacher's competences evaluated by students in each set of competences.

The UNTL teachers' competences were evaluated by a sample of 342 students with a 28 items questionnaire grouped under four categories of competences: pedagogical competences (9 items); professional competences (8 items); personal competences (6 items); social competences (5 items). The 28 items were measured with a five-item scale ranged from "1-very bad" to "5-very good". All categories of competences had satisfying Cronbach's alphas (between 0,77 and 0,81) and a strong positive and significant correlation coefficients (between 0,55 and 0,69). The significant correlations between the four competences exhibit the multifaceted nature of teacher

competences and confirm the overlap and connections among the teaching competences (Alhija, 2016). Mean score of each category was computed from the corresponding items and had an average score around “4-good”.

Pedagogical competences has a mean score of 3,56 (and a standard deviation of 0,67) and was the group of teachers' competences that students evaluate with lowest score. The coefficient variation of 19% suggests a relative homogeneity of students' evaluation on this set of competences. The majority of students, 53% evaluated UNTL teachers' pedagogical competences as sufficient and 29% as good. The highest evaluated teacher's pedagogical competences by students were “Adaptation of materials of the exam/ working towards the goal of the subject” (Mean=3,83) and “Give materials and clear answers to questions” (Mean=3,67). “Assignment of feedback on the work / evaluation” (Mean=3,33), “Use of media and educational technology” (Mean=3,36) and “Ability to create a living environment in the classroom” (Mean=3,49) were competences with lowest students' evaluation so need teacher's improvement.

Professional competences has a mean score of 3,59 (and a standard deviation of 0,66). The coefficient variation of 18% suggests students' evaluation on teacher's professional competences was homogenous and this is the set of competences with more agreement among students. Also the majority of students, 51% evaluated UNTL teachers' professional competences as sufficient and 34% as good – this proportion of students that evaluated as good professional competences were higher than the proportion evaluated as good the pedagogical competences. “Ability to properly explain the content/ approach/ topic/ concept” (Mean=3,80) and “Ability to give relevant examples of concepts” (Mean=3,66) were the highest evaluated teacher's professional competences by students. “Ability of using various communication technologies” (Mean=3,38) and “Ability to interdisciplinary explanations of the current topic with others” (Mean=3,48) were the were professional competences with lowest students' evaluation so need teacher's improvement. Students attribute greatest importance to teacher's “ability to properly explain the content / approach / topic / concept” was not unexpected (Alhija, 2017). Teaching in a clear and understandable manner is critical for transmitting learning material effectively as well as for communicating to students the teaching goals, expectations, and anticipated outcomes.

Personal competences were the UNTL teacher's competences students evaluated highest with a mean score of 3,66 (and a standard deviation of 0,71). The coefficient variation of 19% suggests a relative uniformity of students' evaluation on this set of teaching competences. 45% Of students considered the UNTL teacher's personal competences sufficient and 38% as good – personal

competences had the highest proportion of students that evaluated them as good. The highest evaluated personal competences were “Dignity and personality as a teacher” (Mean=3,91) and “Self-Confidence in any situation” (Mean=3,72) while teacher’s “Watchword and measures” (Mean=3,39) and have received a lowest mean score by students.

The students recognize a highest level of dignity, honesty, justice, fairness and self-confidence of UNTL teachers, but point a lack of authority: “Modelling by the position and behavior” (Mean=3,63). So, being helpful to students, easily accessible, being respectful and understand students were the highest personal and social competences valued by students, as had seen before Al-Mohaimed and Khan (2014) in their cross-cultural research and Alhija (2017).

Social competences had the second position in the most evaluated UNTL teacher’s competences according students with a mean score of 3,63 (and a standard deviation of 0,80). The social competences were the set of competences, among all, that students less converge in their teacher’s evaluation since had a highest coefficient variation of 22% (but already low). This suggests that students’ evaluation is more despair on social competences and less on professional competences. 44% Of students evaluated the UNTL teacher’s social competences as sufficient and 38% as good. Personal and social UNTL teachers’ competences had the most positive evaluation from students. The highest evaluated social competences refer to “Tolerance in religious groups of students” (Mean=3,81), “Ability to accept criticism and suggestions from students” (Mean=3,62) and “Knowing well the students of their subject” (Mean=3,62). “Ability of expression” (Mean=3,50) and “Easily get along with peers, staff and students” (Mean=3,57) were the social competences with lowest students’ evaluation so need teacher’s improvement.

5.1.2. How students’ background characteristics (like gender, age and faculty) influence their evaluation?

The purpose of this research question – How students’ background characteristics (like gender, age and faculty) influence their evaluation? – is to analyze if the student’s background characteristics like gender, age and faculty influence evaluation they made of UNTL teachers’ competences in order to identity profiles of students perceptions about teaching competences.

Objective:

- To analyze if the student’s background characteristics like gender, age and faculty influence evaluation they made of UNTL teachers’ competences.

Gender and faculty were the students' background characteristics with influence on students' evaluation of UNTL teachers' competences, while student's age has no influence. The students sample has 52% female students and 48% male students. The most frequent age group was 18 to 24 years old (88%). The Faculty of Medicine was the most frequent (30%) followed by the Faculty of Education, Arts and Humanities (29%), the Faculty of Philosophy (25%) and, finally, the Faculty of Political Science (16%).

Gender of the students has a significant influence on student's evaluation of UNTL teachers' pedagogical, professional and social competences. Male student's evaluated UNTL teachers' as bad on pedagogical (61%) and social competences (62%), while female students has a significant evaluation as good on professional (63%) and social competences (62%). The female students attribute a more positive evaluation to UNTL teachers' competences than male students, in relation to pedagogical, professional and social competences. These findings concur with results reported in previous studies (e.g. Anderson et al., 2012; Korte et al., 2013; Lavin et al., 2012; Alhija, 2017). The student's evaluations of teacher's personal competences didn't revealed significant differences between male and female students as the findings of Al-Mohaimed and Khan (2014).

The student's age had no significant influence on teacher's evaluation. The students' evaluation on UNTL teachers' competences were similar among the three age ranges. There were no significant differences across the different age groups of UNTL students on their ratings for the pedagogical, professional, personal, and social competence ($p > 0,05$). 88% Of students have between 18 and 24 years and their evaluation is concentrated on the evaluation level of "sufficient" and "good". Student's evaluation of UNTL teachers' is similar in all four sets of competences, independent from students' age. These results is according to previous studies from Al-Mohaimed & Khan (2014) and Lay (2016) who demonstrates teachers evaluation not vary between the juniors and the seniors students. Otherwise, Alhija (2017) found the older and more mature students perceived student long-term development as significantly more important, so they tend to attribute a higher valuate the pedagogical and professional competences of teachers. This assumption wasn't confirmed in our study. The age of the evaluator is not a factor that influences the students' evaluations provided to teachers.

Student's faculty also influences students' evaluation on all teachers' competences ($p < 0,05$). This analysis considered four faculties: faculty of philosophy; faculty of education, arts and humanities; faculty of medicine and faculty of political science. Philosophy integrates the humanities' field but in this study we analyzed four faculties separately.

Pedagogical competences has significant bad evaluation by students from Faculty of Philosophy (45%) while students from Faculty of Medicine attributed significant evaluation as good (44%) and very good (75%) to this competence.

Professional competences have significant bad evaluation also by students from Faculty of Philosophy (49%) while students from Faculty of Medicine attributed significant good evaluation (47%).

Personal competences were the most different among four faculties. Once again students from Faculty of Philosophy were the most critics on UNTL teachers' personal competences since 39% evaluated as bad as well 28% students from Faculty of Political Science. Students from Faculty of Medicine attributed significant evaluation as good (42%) and students from Faculty of Education, Arts and Humanities had significant evaluation as very good (75%).

Social competences has more positive evaluation again from Faculty of Medicine's students (43%) while the students from Faculty of Political Science were the most critics, attributing significant evaluation as very bad (50%) and bad (31%).

These results are according the previous study carried by Wiener et al. (2015) and Alhija (2017) that students in the social sciences and humanities (soft) attributed to instructor-student relation (personal competences) and teaching methods (pedagogical competences) greater importance compared to students in natural and exact sciences (hard). The results from our research were according to the Alhija (2017) conclusions in relation to personal competences because 75% of "very good" student's evaluation was from humanities' students.

Students' from philosophy were very critics about pedagogical, professional and social competences; students from humanities highest evaluated personal competences; students from political science were critics in relation to teachers' personal competences. Students from natural and exact sciences (medicine) did most positive evaluation in all competences' set. These findings also concur with results pertaining to medical students (Al-Mohaimed & Khan, 2014) who valued characteristics related to pedagogical performance more than their personality traits.

5.1.3. How do UNTL teachers self-evaluate their competences?

With this research question – how do UNTL teachers self-evaluate their competences? – we attempt three objectives: 1) to describe the self-evaluation of UNTL's teachers' pedagogical, professional, personal and social competences; 2) to identify the highest self-evaluated UNTL teacher's competences in each set of competences and 3) to identify the lowest self-evaluated UNTL teachers' competences in each set that need improvement.

Objectives:

- To describe the self-evaluation of UNTL's teachers' pedagogical, professional, personal and social competences;
- To identify the highest self-evaluated UNTL teacher's competences in each set of competences;
- To identify the lowest self-evaluated UNTL teachers' competences in each set.

The UNTL teachers' competences self-evaluation was carried in a sample of 192 teachers and with a 28 items questionnaire grouped under four categories of competences: pedagogical competences (9 items); professional competences (8 items); personal competences (6 items); social competences (5 items). The 28 items were measured with a five-item scale ranged from "1-very bad" to "5-very good". All categories of competences had satisfying Cronbach's alphas (between 0,79 and 0,88) and a strong positive and significant correlation coefficients (between 0,51 and 0,71). The significant correlations between the four competences exhibit the multifaceted nature of teacher competences and confirm the overlap and connections among the teaching competences (Alhija, 2016). Mean score of each category was computed from the corresponding items and had an average score around "4-good" to pedagogical and professional competences, and "5-very good" to personal and social competences.

Pedagogical competences has a mean score of 4,44 (and a standard deviation of 0,49) corresponding to good and a very positive teachers' self-evaluation. The coefficient variation of 11% suggests a considerable homogeneity self-evaluation. The great majority of teachers, 73% self-evaluated their pedagogical competences as "good". The highest self-evaluated teacher's pedagogical competences were "Objectivity of student assessment" (Mean=4,51) and "Seriousness in the preparation of classes (programs, quality plan and its implementation)" (Mean=4,48). "Mastering in the use of media and learning technologies" (Mean=4,39) and "Use of varied modes of assessment of student study behavior" (Mean=4,38) were competences with lowest teachers' self-evaluation so need teacher's improvement.

Professional competences has a mean score of 4,20 (and a standard deviation of 0,47) and is the group of competences that teachers' self-evaluate with lowest mean score. The coefficient variation of 11% suggests a considerable homogeneity self-evaluation. The great majority of teachers, 69% self-evaluated their professional competences as "good". The highest self-evaluated teacher's professional competences were "Mastering the disciplinary areas of its main function" (Mean=4,43) and "Ability of interdisciplinary explanations of the current topic with others" (Mean=4,38). "Ability to achieve the advancement of science and technology to enhance the

learning process" (Mean=3,99) and "Involvement in scientific work and professional organizations" (Mean=3,64) were professional competences with lowest teachers' self-evaluation so need teacher's improvement.

Personal competences had the second position in the most self-evaluated UNTL teacher's competences with a mean score of 4,50 (and a standard deviation of 0,41). The coefficient variation of 10% suggests a considerable homogeneity teachers' self-evaluation. The great majority of teachers, 76% self-evaluated their personal competences as "good". The highest self-evaluated teacher's personal competences were "Dignity and personality as teacher" (Mean=4,58) and "Honesty and fairness in decision making" (Mean=4,55). "Be a role model" (Mean=4,39) and "Firm position and application of measures" (Mean=4,44) were personal competences with lowest teachers' self-evaluation so need teacher's improvement.

Social competences were the highest self-evaluated competences by UNTL teacher's with a mean score of 4,54 (and a standard deviation of 0,43). The social competences were the set of competences, among all, that teacher's more converge in their self-evaluation since had the lowest coefficient variation of 9%. This suggests that UNTL teachers' self-evaluation is a little more despair on pedagogical and social competences (even though low) than in relation to personal (10% of variation) and social competences (9% of variation). 71% of UNTL teacher's self-evaluated their social competences as "good" and 22% as "very good". The highest self-evaluated social competences refer to "Tolerance for different religious denominations" (Mean=4,60) and "Ability to accept criticism and suggestions from students" (Mean=4,56). "Expression capacity" (Mean=4,49) and "Easily social coexistence" (Mean=4,47) were the social competences with lowest teachers' self-evaluation so need teacher's improvement.

Teachers emphasized the personal and humanistic characteristics and give importance to the ethical and social aspects of teaching, as demonstrated Kozikoğlu (2017). Teachers' authority, innovation capacity, expression and the use of information and communication technologies were competences that teachers self-evaluated lowest and should be improved. Kozikoğlu (2017) defends that there should be selective courses in teacher training programs in order to develop prospective teachers' personality, personal values and teaching pedagogical skills; a warm, lively and interactive environment should be created in the classroom so that prospective teachers can develop their sense of understanding, humor and communication skills. Teacher's personality, humanistic approach towards students, professional knowledge and skills are key elements in being perceived as ideal teacher by the students. As Kahraman (2014) reviewed the model of teacher change as

suggested by Ross and Bruce (2007) mentioned an important aspect that teacher's self-assessment is located at the core of the model, where teachers observe their own instructional practices, evaluate their effectiveness, and make judgments about their performance. This self-assessment procedure is susceptible to other people's opinions. In the same manner, according to the model, teachers' self-assessment can affect their self-efficacy beliefs which have direct correlations to teachers' goal setting and effort. Model of Teacher Self-assessment as a Mechanism for Teacher Change (Ross & Bruce, 2007). Furthermore, this finding can also be explained by Fullan, Gallardo & Hargreaves (2015) the issue of teacher accountability is that the teacher is taking responsibility his own actions, and aware that at the core of accountability in educational systems is student learning. Constantly improving and refining instructional practice so that students can engage in deep learning tasks is perhaps the single most important responsibility of the teaching profession and educational systems as a whole. In addition to this (Fullan, Gallardo & Hargreaves, 2015) thus emphasized that internal accountability occurs when individuals and groups willingly take on personal, professional and collective responsibility for continuous improvement and success for all students.

5.1.4. How do teachers' background characteristics influence their self-evaluation?

The purpose of this research question – How do teachers' background characteristics influence their self-evaluation? – is to analyze if gender, age, length of service, academic degree and department of UNTL teachers influence their self-evaluation and with this identify teachers' profiles of teachers' competences.

Objective:

- To analyze if teachers' background characteristics like gender, age, length of service, academic degree and teacher's department influence their self-evaluation.

The majority of UNTL teachers participated in study were male (79%); 11% have less than 35 years, 37% have [35-45] years and 49% have 45 years or more. 47% of UNTL teachers' have between 5 and 10 years of experience (included) and the majority have a master degree (72%), 39% were from Faculty of Political Science and 20% from Faculty of Agriculture.

In relation to gender, there were no differences on teachers' pedagogical, professional, personal and social competences between male and female teachers ($p > 0,05$). The majority of both male and female teachers considered their competences good: 71% of male and 80% of female in

pedagogical competences; 71% of male and 63% of female in professional competences; 76% of male and 73% of female in personal competences; 72% of male and 68% of female in social competences.

Regarding to UNTL teacher's age there were also no significant influence on teachers' self-evaluation ($p > 0,05$). The majority of both age groups considered their teaching competences good, so teachers' age don't influence teachers' self-evaluation. UNTL teacher's self-evaluation was similar between younger and older teachers. The majority of both age groups considered their competences good: 81% of <35 years, 71% of [35-45[years and 72% of ≥ 45 years in pedagogical competences; 76% of <35 years, 63% of [35-45[years and 72% of ≥ 45 years in professional competences; 71% of <35 years, 71% of [35-45[years and 79% of ≥ 45 years in personal competences; 76% of <35 years, 67% of [35-45[years and 73% of ≥ 45 years in social competences.

The length of service has a significant influence only on teachers' self-evaluation social competences ($p < 0,05$). So there were differences of social competences self-evaluation between the three groups of time of service. Teachers with less than five years of service self-evaluated their social competences lower (Mean=4,31) than teachers with a length of service between 11-20 years (Mean=4,55) and teachers with an experience in service between 5-10 years (Mean=4,58). Teachers with more time service have more confidence and a more positive self-perception of their social competences than teachers with less time of experience. In relation to pedagogical, professional and personal competences teachers' self-evaluation don't differ significantly and is similar between the teachers with less or more time of service ($p > 0,05$). Despite of the inexistence of significant differences we note teachers with less than five years of experience self-evaluated their pedagogical (Mean=4,25) and mostly their professional competences (Mean=4,01) lower than teachers with more experience that points 4,4 and 4,2 on pedagogical and professional competences, respectively.

Teachers' academic degree had an influence only on pedagogical competences self-evaluation: teachers with master (Mean=4,50) and PhD degree (Mean=4,59) had higher self-evaluation scores than teachers with a postgraduate (Mean=3,85) or bachelor degree (Mean=4,24) ($p < 0,05$). Otherwise, teacher's academic degree had no influence on professional, social and personal competences of UNTL teachers, since the average scores of self-evaluation were similar and with no significant differences between teacher's with different academic degree ($p > 0,05$). Despite of the inexistence of significant differences we note teachers with a post-graduate degree self-

evaluated their professional competences (Mean=3,79) lowest in relation to teachers with a bachelor (Mean=4,12), a master (Mean=4,22) or a PhD degree (Mean=4,19).

Regarding the teacher's department of UNTL and their field of study there were significant differences on pedagogical, professional, personal and social teacher's competences self-evaluation ($p < 0,05$). So teacher's department influences their self-evaluation competences. Teachers' self-evaluation in relation to pedagogical and professional competences was more heterogeneous than in relation to pedagogical and professional skills, since there were a highest number of significative differences.

Pedagogical competences has highest mean score among teachers from Political Science department (Mean=4,66) followed by teachers from Economy department (Mean=4,55), Medicine department (Mean=4,23) and Education, Arts and Humanities department (Mean=4,05).

Professional competences has highest mean score among teachers from Political Science department (Mean=4,40) followed by teachers from Economy department (Mean=4,38), Medicine department (Mean=4,03), Agriculture department (Mean=3,98) and Education, Arts and Humanities department (Mean=3,98).

Personal competences has highest mean score among teachers from Political Science department too (Mean=4,61) comparing with Education, Arts and Humanities department (Mean=4,32).

Social competences has also highest mean score among teachers from Political Science department (Mean=4,67) in relation to Agriculture department (Mean=4,38).

UNTL teacher's from Political Science department has the highest self-evaluated scores in all set of competences, by this order: social, pedagogical, personal and social competences.

The teacher had to think about self-evaluation to be successful every day and this would help increase their thinking competences more. A continuous and consistent practice it would help teachers to achieve their performance which they had to do qualitatively (Suknaisitha, Suwimon & Piromsombatc, 2014). The self-evaluation on teachers competences strategy could change behaviors to be positive as concerns the required goals and could be applied both in academic and social aspects. The crucial principles were about observation and consistently recording the behavior. Ideal teachers should have qualifications such as humaneness, joviality and personal values as well as professional knowledge (content knowledge and pedagogical skills). Teacher's personality, humanistic approach towards students, professional knowledge and skills are key elements in being perceived as ideal teacher by the students.

Our study concludes that self-perception of UNTL teachers is influenced by length of service (only

social competences), academic degree (only pedagogical competences) and UNTL department.

5.1.5. To what extent do students' evaluation of UNTL teachers and the teachers' self-evaluation competences differ or converge?

The purpose of this research question is to highlight – To what extent do students' evaluation of UNTL teachers and the teachers' self-evaluation competences differ or converge? – is compare student's evaluation of UNTL teachers' competences and the teachers' self-evaluation, in order to identify divergences and / or convergences. Then, to identify which teachers' competences are different between student's evaluation and teachers' self-evaluation and which teachers' competences are similar according to student's evaluation and teachers' self-evaluation.

Objectives:

- To compare student's evaluation of UNTL teachers' competences and the teachers' self-evaluation, in order to identify divergences and / or convergences;
- To identify which teachers' competences are different between student's evaluation and teachers' self-evaluation;
- To identify which teachers' competences are similar according to student's evaluation and teachers' self-evaluation.

To answer these objectives we review the main results describing students' evaluation of UNTL teachers' pedagogical, professional, personal and social competences with the self-evaluation that teachers did about their own pedagogical, professional, personal and social competences.

Personal (Mean=3,66) and social competences (Mean=3,63) were the highest scores in students' evaluation and in teachers' self-assessment. Pedagogical (Mean=3,56) and professional competences (Mean=3,59) were the lowest evaluated by students as well by UNTL teachers. In general students evaluated their teachers with a "sufficient" level: 53% in pedagogical, 51% in professional, 45% in personal and 44% in social competences. Personal and social competences were the most positive teachers' competences according to students since they classified it with a "very good" level (38% each competence).

Among UNTL teachers pedagogical (Mean=4,44) and professional (Mean=4,20) competences were the lowest evaluated competences too comparing with personal (Mean=4,50) and social (Mean=4,54). The majority of UNTL teachers self-evaluated them as "good": 73% in pedagogical

competences, 69% in professional competences, 76% in personal competences and 71% in social competences.

Students' evaluation were influenced by gender in relation to pedagogical, professional and social competences ($p < 0,05$) with more positive evaluations among female students. Otherwise the self-evaluation is independent from teacher's gender.

Students' assessment as well teachers' self-evaluation has no influence by age. Students' evaluation and self-evaluation teachers were equal between the younger and older participants ($p > 0,05$).

Students' faculty has impact on evaluation they do in all competences, as well the department of UNTL teachers belong impacts on teachers' self-evaluation. Students from humanities evaluated with the highest score teachers' personal competences while students from exact sciences (medicine) highest evaluated pedagogical competences. Teacher's from Political Science department reveals a significant highest self-evaluation in pedagogical, professional, personal and social competences ($p < 0,05$). UNTL teacher's from Political Science department has the highest self-evaluated scores in all set of competences, by this order: social, pedagogical, personal and social competences. UNTL teachers' social competences were significant higher self-evaluated by teacher with 5-10 years of length of service (Mean=4,58) than teachers with less than five years of experience (Mean=4,31). UNTL teachers' pedagogical competences were significant higher self-evaluated by teachers with master (Mean=4,50) and a PhD degree (Mean=4,59) than teachers with bachelors (Mean=4,24) and postgraduate (Mean=4,85).

According to Kozikoğlu (2017) teachers points at first teaching pedagogical skills, humaneness/joviality and personal-professional values, by this order. Teachers points at least three cognitive construct categories like innovativeness, sensitivity and leadership/ guidance, by this order.

Jónsson, Smith and Geirsdóttir (2018) addressed the issue of variability of perception of teachers and students regarding feedback and assessment. Feedback refers both to the information about students' work and their engagement with the feedback information, a formative assessment approach promoting the dialogue between students and teachers. Smith, Gamlem and Engelsen (2017) used the term "responsive pedagogy" for "the learning dialogue" taking place between the student and the teacher about goals, competences in achieving those goals and strategies for getting there, as well to enhancing further learning. Between UNTL teachers and students is necessary to implement feedback as a routine and to improve it pedagogical competence in a responsive and reflexive approach as referred by these authors.

Teachers self-evaluated their competences more positive than students. Students' evaluation differs by gender and faculty. UNTL teachers' self-evaluation depends by length of service (social competences) and degree (pedagogical competences) and department (all competences).

Pedagogical competences similarities: students evaluated higher "Adaptation of materials of the exam/ working towards the goal of the subject" (Mean score=3,83) and "Give materials and clear answers to questions" (Mean score=3,67), as well teachers self-evaluation "Seriousness in the preparation of classes (programs, quality plan and its implementation)" (Mean=4,48). The "Use of media and educational technology" (Mean score=3,36) according to students and use of media and learning technologies" (Mean=4,39) according to teachers' self-evaluation need improvements.

Pedagogical competences differences: UNTL students refer the "Assignment of feedback on the work / evaluation" (Mean score=3,33) as the competence need improvements, while teachers identified the "Use of varied modes of assessment of student study behavior" (Mean=4,38). UNTL teachers' consider their "Objectivity of student assessment" (Mean=4,51) the most positive pedagogical competence.

Professional competences similarities: teachers recognized that the need to use more the technologies to enhance the learning process – "Ability to achieve the advancement of science and technology to enhance the learning process" (Mean=3,99) – and UNTL students have the same perception, teachers had lowest competences in relation to the "Ability of using various communication technologies" (Mean=3,38).

Professional competences differences: to UNTL students the most positive teachers professional competences were the "Ability to properly explain the content/ approach/ topic/ concept" (Mean=3,80) and the "Ability to give relevant examples of concepts" (Mean=3,66). The "Ability to interdisciplinary explanations of the current topic with others" (Mean=3,48) is considered by students as the competence with the lowest evaluation. Otherwise, teachers considered that their best competences were "Mastering the disciplinary areas of its main function" (Mean=4,43) and "Ability of interdisciplinary explanations of the current topic with others" (Mean=4,38). UNTL teachers want to have a more "Involvement in scientific work and professional organizations" (Mean=3,64).

Personal competences similarities: "Dignity and personality as a teacher" were the personal competence most valued by students (Mean=3,91), as well by teacher's self-evaluation (Mean=4,58). According to students teachers need to improve personal competences like

“Watchword and measures” (Mean=3,39) and “Modelling by the position and behavior” (Mean=3,63). In the same way, teachers considered that need to improve their ability to “Be a role model” (Mean=4,39) and have a “Firm position and application of measures” (Mean=4,44).

Personal competences differences: Students give a highest evaluation also to the competence of “Self-Confidence in any situation” (Mean=3,72), while teachers considered that their best competences were “Dignity and personality as teacher” and “Honesty and fairness in decision making” (Mean=4,55).

Social competences similarities: “Tolerance in religious groups of students” was a social competence most valued by students (Mean=3,81) as well by teacher’s self-evaluation (Mean=4,60); “Ability to accept criticism and suggestions from students” were evaluated by students (Mean=3,62) as by teacher’s self-evaluation (Mean=4,56). “Expression capacity” is the lowest teacher social competence by students (Mean=3,50) and by teachers self-evaluation (Mean=4,49).

Social competences differences: students highest evaluated the fact of teachers “Knowing well the students of their subject” (Mean=3,62) while teachers claims to more “Easily social coexistence” (Mean=4,47).

Teachers give more emphasis to assessment materials and process than students. Teachers self-considered they use to the objectivity but want to use more varied modes of assessment. Suknaisitha, Suwimon and Piromsombatc (2014) also demonstrated teachers use measurement and evaluation instruments are not varied. The authors showed that teachers intended to learn about learning educational measurements and evaluations through participation in activities. So a teacher training program should be practice and action planning, emphasize the teachers’ performances adaptation in terms of educational measurement and evaluation by using the self-monitoring strategy on the educational measurement and evaluation performance of teachers.

The professional competence to do interdisciplinary explanations is the lowest evaluated by students and the highest self-evaluated by teachers. Suknaisitha et al. (2014) emphasize the teachers’ performances adaptation in terms of educational measurement and evaluation by using the self-monitoring strategy on the educational measurement and evaluation performance of teachers.

According to Jónsson, Smith and Geirsdóttirc (2018) teachers seem to overestimate how much students are involved in the feedback dialogue, since students experience feedback as being useful for further learning, which explain it importance to students. So, it is necessary to implement

feedback as a routine and to improve it pedagogical competence in a responsive and reflexive approach because students identified the feedback like the pedagogical competence than need be improved.

Students considered the interdisciplinary competence the lowest in their teachers and teachers self-evaluated this capacity as very good. Students high evaluated the fact of teachers knowing well the students of their subject.

Teachers give more emphasis to assessment materials and process than students. The preparation of classes with appropriate and clear materials was similar and positive evaluated by students and by teachers. The use of media and educational technologies was similar and less positive evaluated by students as by teachers and both recognized that teachers need to use more the technologies to enhance the learning process. Students and teachers were in agreement in relation to teachers' authority and objective assessments, despite of both considered the teachers' dignity, personality, religious tolerance and the acceptance of criticism were the personal and social competences most evaluated by students and by teachers.

It was concluded, therefore, that at the level of professional competences, students' evaluation and teachers' self-evaluation present convergent points regarding scientific and technical research work in an organizational context. It was found that research is an area that both teachers and students would like to see more developed and empowered in UNTL. To this end, it is necessary to create physical, material and human conditions conducive to scientific and technological research. Policy measures within public higher education policies should therefore be geared towards promoting scientific research, scientific work and the positioning of East Timorese researchers among the international scientific community. Scientific work, and particularly that which develops in universities, is decisive in the technical and scientific progress of a country, its business fabric, its human resources and the country's capacity to respond to current challenges.

5.1.6. How is the Teaching Performance Evaluation (TPE) being implemented in the educational institution?

Looking at the case study of UNTL, the public university of East Timor, the qualitative study intends to answer the question *how the Teaching Performance Evaluation (TPE) is being implemented in the educational institution?* This research question intended to identify the attitudes towards the TPE in Timorese higher education system, like the obstacles and the contributions to performance and teaching development; to know the teacher's experiences in the implementation of TPE in Timorese higher education system; to identify the teaching and learning competences should be

developed by teachers in a professional development program; and draw suggestions and recommendations from UNTL teachers in relation to effective implementation of TPE in higher education system. These findings will allow us to identify the teaching and learning competences should be developed by teachers in a professional development program, and to list teachers' suggestions and recommendations to carry out a better teacher performance evaluation in Timorese higher education system.

Objectives:

- To identify the attitudes towards the TPE in Timorese higher education system, like the obstacles and the contributions to performance and teaching development;
- To know the teacher's experiences in the implementation of TPE in Timorese higher education system;
- To identify the teaching and learning competences should be developed by teachers in a professional development program;
- Draw suggestions and recommendations from UNTL teachers in relation to effective implementation of TPE in higher education system.

To achieve these objectives 16 data interviews were analyzed through qualitative approach. The participants were UNTL teachers who holding management positions at the institution at the time of the interview. From the 16 participants there were 14 male and two female teachers. The mean age was 50 years (Mean=49,625 ± 8,42) and had a coefficient variation of 17% revealing variability and heterogeneous ages. The mean length of service was 10 years (Mean=10,438 ± 4,63) and had a higher coefficient variation of 44%. None of the participants have previous experience in teacher performance evaluation. In terms of their educational background, 11 teachers from the 16 have a master degree and only five have a PhD/Doctorate degree.

In relation to the UNTL teachers' attitudes towards the TPE in Timorese higher education system, the obstacles are referring to the reduced use of the Portuguese Language (n=4) by teachers and institution; teacher's afraid to be evaluated and lack of knowledge about the methods and the competence of the potential evaluators (n=3); lack of a culture of scientific writing (n=2); lack of specialists in the field of performance appraisal (n=2); few number of graduate teachers (PhD degree) (n=2); and a conception of teacher as a mere transmitter of information and knowledges (n=2). These are the attitudes more frequently referred by teachers. Despite of difficulties, teachers identified also positive contributions and benefits from the TPE implementation in UNTL, like

improve of teacher performance (n=4); professional development through training in pedagogical skills (n=5); and identification of the strengths and weaknesses of teaching practice (n=2).

The obstacles identified by teachers, in accordance to Van Schalkwyk et al. (2015) does not happen in isolation but is impacted by the organizational and social worlds that faculty members habit. So, UNTL teacher's professional focus on how professional learning is embedded in workplace structures and cultures, describing professional learning as the enhancement or change of teaching and assessment practices of lecturers in order to ensure quality student learning takes place.

Teodorović, Stanković, Bodroža, Milin & Đerić (2016) identified several barriers in the implementation processes such as a general devaluation of the work of principals, political instability and frequent changes of policy direction, which can lead to reform fatigue and a lack of support. The authors argue that is necessary a participatory implementation of accountability practices, in order to establish a comprehensive, understanding and positive assessment by stakeholders. UNTL teachers wanted an external and objective evaluation commission, as well claims to a professional teaching performance assessment.

These findings support the previous study of Suknaisitha, Suwimon and Piromsombatc (2014) according to which teachers had to think about self-evaluation to be successful every day and this would help increase their thinking skills more. If there was continuous and consistent practice, it would help teachers achieve their performance which they had to do qualitatively. The self-monitoring strategy could change behaviors to be positive as concerns the required goals and could be applied both in academic and social aspects.

The experience in the implementation of TPE in Timorese higher education system is quite different among UNTL teacher's participants. To 14 of the 16 UNTL teachers TPE is not implemented. Otherwise, other teachers recognized the TPE implementation promote self-assessment carried out by teachers in order to identify the existing gaps (n=2). Process' objectivity is also recognized – the performance evaluation of UNTL teachers and civil servants is based on the standard form of the Public Service Commission of Timor-Leste, however it is not the most appropriate (n=2) – but teachers defends a program of assessment more teachers' close and more individualized, not only implemented by the CPAI (Institutional Standing Evaluation Committee) approved by Law No. 21 dealing with the TPE (n=2).

Kozikoğlu (2017) defends that prospective teachers' cognitive constructs are influenced by their individual experiences and their learning/knowledge resulting from their interaction with the external environment, since they attend at different grade levels, differ in educational experiences

and types of teachers they met during their educational life. So we have to look to UNTL teachers' experience in TPE framing in a particular professional and personal context. Teacher's experience in the implementation of TPE in Timorese higher education system depends from their experience, professional and personal context.

In relation to the teaching and learning competences should be developed by teachers in a professional development program, there were mentioned the research work (n=1), teachers' competences, especially in pedagogical field (n=11); pedagogical competences training in collaboration with the Calouste Gulbenkian Foundation (n=1), the use of diverse teaching materials and methodologies (n=1), a critical and reflexive learning centered on students (n=1) and a collaborative work (n=1).

As showed by Suknaisitha, Suwimon and Piromsombatc (2014) teachers intended to learn about educational measurements, evaluations through participation in activities and improving by self-monitoring. By applying the self-monitoring process, teachers gained further knowledge in learning measurements and evaluations, participating in action planning, as well as greater realization and awareness of the importance of learning measurements and evaluations. As well had seen Kozikoğlu (2017) teachers want to improve and develop their qualification in relation to teaching pedagogical skills, competences of leadership and guidance, professional content knowledge and professional competences. UNTL teachers recognized they need to develop and improve pedagogical competences: the use of diverse teaching materials and methodologies; to obtain further knowledge about measurements and evaluations; a critical and reflexive teaching/learning. Teachers want to improve and develop their competences and qualifications.

The suggestions and recommendations mentioned by UNTL teachers in relation to effective implementation of TPE in higher education system are referring to assessment process, teachers' attitude, university culture and the management of human resources.

TPE process should be conducted by experts in the field (n=5) and should be created a team or commission responsible for UNTL teachers' evaluation (n=4). Teachers' assessment should be continues and should be done every year by the students, at the end of the school year (n=2). Study's participants also recommends the elaboration of a form/grid by each department to measure teacher's performance, this form must contain precisely indicators related to pedagogical, professional, personal and social skills (n=2). At least, the TPE process should be developed through a close collaboration between the Ministry of Education and the University in order to clarify the procedures to be followed (n=2). In academic context of the university should be encourage the

creation of a collaborative and learning culture at the level of teachers (n=2). In terms of human resources politics of management, teachers must obey to the rules and regulations and those who do not comply with the law must be sanctioned (n=2); reward teachers for their good performance (n=1).

According to Nevgi and Lofstrom's (2015) professional learning is impacted not only by faculty development initiatives but also by the organizational structures and practices within which faculty members operate, like experience, tenure, salary, rewards and promotion practices. So is necessary human resources politics oriented to and consistent with TPE. Providers of faculty development should first understand the role of the workplace (including tenure, salary, rewards and promotion practices) and then design their initiatives accordingly. Faculty development initiatives should, at best, incorporate reflection on teaching (beliefs) in order to stimulate an actual impact on teaching practice.

These findings are supported by Teodorović et al. (2016) according to which professional judgement is needed with respect to interpreting and mediating new policies and accountability expectations. The authors argue that policy and accountability need to be based on professional knowledge and expertise (practice and research).

Accountability need to be based on professional and expert knowledge and expertise (practice and research). Teachers want to collaborate in order to create a collaborative and learning culture. Teachers want to implement an evaluation and accountability culture in UNTL. It is necessary human resources politics oriented to and consistent with TPE.

5.2. Hypotheses

5.2.1. The personal and the social competences were the highest evaluated by students, as compared to the pedagogical and professional competences

It was initially hypothesized that the personal and the social competences were the highest evaluated by students, as compared to the pedagogical and professional competences. This hypothesis is fully supported by the results. The results show that personal competence has a mean of 3.66 which is higher than pedagogical (with a mean score of 3,56) and professional (with a mean score of 3,59). The mean for the social competences is 3.63 which is also higher than pedagogical and professional competences.

These results are supported by the findings of Magno and Sembrano (2007) that when students assess their teachers, they emphasize mostly on the personal and social factors. The mean levels

on students' ratings were higher than pedagogical related factors. High evaluations on social and personal factors are evaluated by students as characteristics of effective teaching. Students see personal factors such as personality, honesty and self-confidence and social factor such as expression, accepting criticisms, knowing students well, and being tolerant on individual differences are seen as effective qualities of students. Furthermore, the same results are likewise supported by the findings of Alhija (2017) that the predominant characteristics evaluated by students included the personal and social competences, which seems to show that students share their perceptions about good teaching and good instructors differently.

5.2.2. There is a relationship between students' evaluation on teacher's competences and students' background characteristics

5.2.2.1. The female students attribute a more positive evaluation to teachers' competences than male students

Second hypothesis points there is a relationship between students' evaluation on teacher's competences and students' background characteristics. At first it was hypothesized that the female students attribute a more positive evaluation to teachers' competences than male students. This hypothesis is partially supported in the study because there were observed gender differences only in relation to pedagogical, professional and social competences ($p < 0,05$). UNTL teachers' personal competences has equal evaluation among male and female students, prevailing the "sufficient" (49% by male and 51% by female) and "good" evaluation (42% by male and 58% by female).

Lavin, Korte and Davies (2012) as well Korte, Lavin and Davies (2013) found that female students tend to assign a higher rank to traits related to effective (personal) teaching than did male students, but this assumption was not verified in current study.

Pedagogical competences male students has a significant "bad" evaluation (61%) while majority of the female students rated the teachers' pedagogical competence in the "good" level (56%) with frequencies significantly different from males.

The same pattern is observed for professional competences where majority of the female students rated the professional teacher's competences in the "good" level (63%) and the frequencies were also significantly different from males.

Social competences vary significantly between male and female students. Male students are more critics about UNTL teachers' social competences and evaluated them as "bad" (62%) while 62% female students made a "good" evaluation.

These findings support the study of Morgan (2001) that female students receive more positive feedback from teacher, so making them higher rated than do male students. Female students do a positive (“good”) and significant evaluation on professional (63%) and social competences (62%). The results are likewise supported by Alhija (2017) that the gender have a significant effect on students’ perceptions of good teaching. Gender differences were found regarding all four set of competences of good teaching, whereby female students assign to all of them greater importance than did male students. Furthermore, the results were also supported by Anderson et al. (2012), in their study among doctoral students’ women were more likely to endorse the traits of professional, expert and student centered as characteristics of good instructors.

5.2.2.2. The older students give a more positive evaluation to teachers’ pedagogical and professional competences than younger students

It was also hypothesized that the older students give a more positive evaluation to teachers’ pedagogical and professional competences than younger students. This hypothesis was not supported in the study. The results showed that there were no significant differences across the age groups of the students on their ratings for the pedagogical, professional, personal, and social competence. The results show that the age groups did not influence their rating for teachers. These findings support the findings of Lay (2016) that junior and senior high school students do not differ on how they rate their teachers’ disposition. Younger and older groups of students can equally provide the same levels of teacher evaluations. The age of the evaluator is not a factor that influences the accuracy of evaluations provided to teachers. In the opposite position Alhija (2017) found the older and more mature students perceived student long-term development as significantly more important, so they tend to attribute a higher value to the pedagogical and professional competences of teachers. This assumption wasn’t confirmed in our study because no differences between younger and older students were reported.

5.2.2.3. Students from humanities provided higher evaluation to personal competences, while students in natural and exact sciences provided higher evaluation to pedagogical competences of the teachers

It was hypothesized that students from humanities provided higher evaluation to personal and social competences, while students in natural and exact sciences provided higher evaluation to pedagogical and professional competences of the teachers. At first it was confirmed significant differences on UNTL teachers’ evaluation pedagogical, professional, personal, and social

competences ($p < 0,05$). All evaluated competences were dependent from students' faculty and field of study. The analysis of this hypothesis considered four faculties: faculty of philosophy; faculty of education, arts and humanities; faculty of medicine and faculty of political science. Philosophy integrates the humanities' field but in this study we analyzed four faculties separately.

The results demonstrate students from philosophy were very critics about pedagogical, professional and personal teachers' competences, evaluated them as bad or very bad: pedagogical competences were "bad" to 45% students from Faculty of Philosophy; professional competences were "bad" to 49% students from Faculty of Philosophy too; personal competences were "very bad" only for students from Faculty of Philosophy and "bad" to 39% of students from the same faculty plus 28% of students from political science. Social competences have a significant "very bad" and a "bad" evaluation among students from Faculty of Political Science (50% and 31% respectively). Students from humanities (philosophy) were very critics about pedagogical, professional and personal teachers' competences, and students from political science were very critics about personal and social UNTL teachers' competences.

Students from faculty of education, arts and humanities give the most positive and significative evaluation to personal competences. 75% Of students evaluated personal competences as "very good" were from humanities. At this point it was confirmed students from humanities provided higher evaluation to personal competences.

Students from medicine did the most positive evaluation in pedagogical, professional, social and personal competences. Pedagogical competences are "good" and "very good" to 44% and 75% of medicine students, respectively; professional competences are "good" to 47% to students from Faculty of Medicine; personal and social competences are also "good" to 42% and 43% of medicine students, respectively. So students in natural and exact sciences (medicine) do a very positive UNTL teachers' evaluation in all four competences set. Students from natural and exact sciences (medicine) were those who give highest positive evaluation to pedagogical competences, as good (44%) and very good (75%). "Very good" evaluation was only significant in pedagogical competences (75%) confirmed that students in natural and exact sciences provided higher evaluation to teachers' pedagogical competences. These results are allowed to confirm the hypothesis that students from humanities provided higher evaluation to personal competences, while students from natural and exact sciences (medicine) provided higher evaluation to UNTL teachers' pedagogical competences. This supports the findings of Talbot (2011) that mathematics and science teachers are rated highly on their strategic knowledge that pertains to pedagogical and professional competence. The study

of Talbot (2011) further explains that science teachers are rated higher on pedagogical and professional competency because it matches their characteristics of applying and modifying their instructional repertoire in a given teaching context. These results are also according to studied developed by Al-Mohaimed and Khan (2014): medical students valued characteristics related to pedagogical performance more than their personality traits.

More recently, studies developed by Wiener et al. (2015) and Alhija (2017) demonstrate that students from the social sciences and humanities (soft) attributed to instructor-student relation (personal competences) and teaching methods (pedagogical competences) greater importance compared to students in natural and exact sciences (hard) – these previous findings were confirmed in current study in relation to personal competences. Students' in humanities (philosophy) who very critics about pedagogical, professional and social competences; students from humanities highest evaluated personal competences; students from political science were critics in relation to teachers' personal competences. Students from natural and exact sciences (medicine) did most positive evaluation in all competences' set. Alhija (2017) shows that the field of study has a significant effect on students' perceptions of good teaching, as well demonstrate the confirmation of this hypothesis.

5.2.3. Teachers self-evaluate their pedagogical and professional competences lower than their personal and social competencies

It was hypothesized that teachers self-evaluate their pedagogical and professional competences lower than their personal and social competencies. This hypothesis is fully supported by the results. UNTL teacher's pedagogical (Mean=4,44) and professional competences (Mean=4,20) had a lower self-evaluation score than personal (Mean=4,50) and social competences (Mean=4,54). These results allowed us to confirm the hypothesis that teachers self-evaluate their pedagogical and professional competences lower than their personal and social competencies. According to UNTL teachers' self-assessment they are needed such improvements in terms of their pedagogical and professional competences, despite of 73% and 69% respectively have considered their competences "good". UNTL teachers' feel much more confident in relation to their personal and social competences: 76% of teachers considered their personal competences "good" and 14% "very good", while 71% of teachers considered their social competences "good" and 22% "very good". These results are supported by the findings of Magno and Sembrano (2007) according to which personal and social skills were valued in the teaching and learning process. This is especially true in the Asian setting where the teaching building rapport and relationship among students is

important. The results were also supported by the findings' Kozikoglu (2017) teachers emphasized the personal and humanistic characteristic and give importance to the ethical and social aspects of teaching. The teacher's authority and innovation capacity were competences that teachers consider that should be improved. These results were also supported by the findings of Skunaisitha, Swimon & Piromsombatc (2014) that the teacher had to think about self-evaluation to be successful every day and this would help increase their thinking competences more. If there was continuous and consistent practice, it would help the teacher achieve their performance which they have to do qualitatively.

5.2.4. UNTL teachers' with more age and years of experience had higher self-evaluation than younger and least experienced teachers

It was hypothesized that UNTL teachers' with more age and years of experience had higher self-evaluation than younger and least experienced teachers. In relation to teachers' age it was observed no differences on self-evaluation scores among the three age groups – the teachers' with less than 35 years, [35-45] years and 45 or more years ($p>0,05$) so UNTL teacher's self-valuation is similar between younger and older teachers in relation to pedagogical, professional, personal and social competences. For each set of competences and in each age group the evaluation of "good" prevailed majority. So the hypothesis that UNTL teachers' with more age self-evaluated their competences higher than younger teachers' is not confirmed and teacher's age is a background characteristic with no influence on teachers' self-evaluation.

These findings are supported by the previous study of Roe (2010) according to which teacher's age have a little to a non-impact on teachers' self-evaluation. Roe's results did not support the assumption that teaching performance increases with teacher's age, since it is not evident more experience gained by the teacher better they become in the teaching activity. This poses a perspective that good teaching is not determined by teachers' age but their ability to provide better instruction.

In relation to length of service, the UNTL teacher's self-valuation of pedagogical, professional and personal competences were equal between teachers with less than five years of experience, with five to 10 years and with 11-20 years of experience, since no significative differences were found ($p>0,05$). Otherwise, the UNTL teacher's self-valuation of social competences differ significantly ($p<0,05$) and teachers with less than five years of experience self-evaluated them lower (Mean=4,31) than UNTL teachers' with an experience around five and 10 years (Mean=4,58). The UNTL teachers' length of service is a background characteristic with influence only on social

competences. So the hypothesis that UNTL teachers' with more time of service self-evaluated their competences higher than younger teachers' is confirmed in relation to social competences only. Hypothesis 4 is partially support by these results, only in relation to teacher's length of service and the social competences.

These findings are supported by the previous study of Roe (2010) according to which teacher's experience have a little impact teacher's rapport. The results did not support the assumption that teaching performance increases with years of practice (time of service and experience). It is not evident that the more teachers' experience influences positively the pedagogical and professional competences of teaching.

5.2.5. UNTL teachers with a higher academic degree have a higher self-evaluation on their competences

It was also hypothesized that the teachers with a higher academic degree (as master and PhD) have higher self-evaluation scores on their competencies than teachers' with lower degree (bachelors and postgraduate). This hypothesis was partially supported by the findings of the study only in relation to pedagogical competences. UNTL teachers' pedagogical competences had different self-evaluation scores between three academic degrees ($p < 0,05$). The highest self-evaluation on pedagogical competences belongs to UNTL teachers with PhD degree (Mean=4,59) followed by teachers with master (Mean=4,50) and then bachelor degree (Mean=4,24). UNTL teachers' with postgraduate (Mean=3,85) has the lowest score of self-evaluation on pedagogical competences. These results demonstrate UNTL teachers' academic degree influence teachers' self-perception about their pedagogical competences and more high academic degree teachers feel more capable in their pedagogical functions. Otherwise, teacher's academic degree had no influence on professional, social and personal competences of UNTL teachers self-evaluation, since the average scores were similar and don't have significative differences between teacher's with different academic degree ($p > 0,05$). Only attending to UNTL teachers' pedagogical competences the hypothesis that UNTL teachers with a higher academic degree have a higher self-evaluation on their competences is confirmed. These findings were supported by the study of Santelices, Valencia, Gonzales and Taut (2017) according to which teacher's degree of training accounts much on their teaching performance. The results point that one's educational degree and training only explains pedagogical competence and no other aspects of the teaching competency. It is a common assumption that part of teacher qualifications is the degree that they earn when have higher degrees, such as possessing graduate advance degrees are rewarded in educational institutions.

However, the results of the current research is helpful to explain only teacher's pedagogical skills and not on all other areas.

5.2.6. UNTL teachers' from political science department highest self-evaluated their competences

It was also hypothesized that the UNTL teachers' from political science department self-evaluated their competences highest. This hypothesis was supported by the study findings. There were observed significant differences among UNTL departments in pedagogical, professional, personal and social competences ($p < 0,05$). Teachers from Political Science department had the highest mean score for their self-evaluation on all four competencies than others.

In relation to pedagogical competences was found teachers the Political Science department have a significative higher self-evaluation (Mean=4,66) than the Education, Arts and Humanities department (Mean=4,05) and the Medicine department (Mean=4,23). Teachers from Economy department have also a significative higher self-evaluation (Mean=4,55) than teachers from the Education, Arts and Humanities department (Mean=4,05). Teachers from Political Science and Economy departments are those who have significative and higher self-evaluation score on pedagogical competences, comparing to department of Humanities and the Medicine department. In relation to professional competences was found teachers from the Political Science department have a significative higher self-evaluation (Mean=4,40) than teachers from three other departments: the Economy department (Mean=4,38), the Medicine department (Mean=4,03), the Agriculture department (Mean=3,98) and the Education, Arts and Humanities department (Mean=3,98). Teachers from Economy department have also a significative higher self-evaluation (Mean=4,38) than the Education, Arts and Humanities department (Mean=3,98) and the Agriculture department (Mean=3,98).

Personal competences differ only between the Political Science (Mean=4,61) department and the Education, Arts and Humanities department (Mean=4,32) with a significative higher self-evaluation among teachers from the first department.

Social competences differ between the Political Science (Mean=4,67) and the Agriculture department (Mean=4,38). Teachers from Political Science department have a higher self-evaluation and a self-perception about their social competences than teachers from Agriculture department.

UNTL teacher's from Political Science department has the highest self-evaluated scores in all set of competences, by this order: social, pedagogical, personal and social competences.

These findings support the studies of Brintnall and Mealy (2014) where political scientists' orientation that inclined in the arts makes them effective in the teaching profession. Political scientists' characteristics such as being argumentative, persuasive and the ability to lead are perceived to be good characteristics of teaching.

The students' evaluation and faculties' self-evaluation were compared. The percentage of agreement on the self-evaluation and students' evaluation were significantly different for pedagogical, professional, personal, and social competence. Higher percentage of agreement was found for teacher's self-evaluation than students' evaluation on pedagogical, professional, personal and social competency. These findings did support the findings in the study of Matsuno (2009) where self-raters assessed their own performance lower than predicted. The higher self-rating result accounts for bias patterns among raters. The students' ratings shows to be a more accurate way of providing assessments.

Qualitative analysis was conducted to determine the (1) recognition and importance given to teacher performance evaluation, (2) attitudes towards teacher performance evaluation and (3) practices to improve teacher performance.

It was found that the importance given to teacher performance evaluation is manifested to the contribution it gives to teachers, the need to have legislations on the system of evaluation, and specific practice and implementation of the teacher performance evaluation. Teachers see the importance of teacher performance assessment because they see the value that the information it brings help them improve their teaching practice. This is especially true when coupled with an effective supervision. When teachers are aware on the areas that they need to improve on, they are able to devise ways on improving their performance. These value in determining the faculty member's appraisal is explained by Braskmap (1994) that teachers improvement becomes good indicators of institutional success. The results show that when the teacher performance becomes instrumental in contributing to teacher improvement, classroom and schools becomes effective. These are the same explanations pointed by Pullman (2018). Results of teacher appraisal are used to sustain the improvement in educational institutions.

It was found that the attitudes of teachers on their performance evaluation points to obstacles identified in the conduct of the evaluation and the recommendation for the practice of evaluation. Teachers' attitudes are shaped by the difficulty and hindrances it gives to their work. This is consistent with the Kupermintz (2003) that when the system of teaching appraisal becomes favorable to teachers when there is opportunity for growth and advantage for the teachers.

It was found that the teacher performance evaluation improves teacher performance in two ways. First is the value given to the formative assessment and training provided to the teachers. Second, are the specific practices that lead to the process of teaching and learning. System on teacher performance becomes useful for teachers and in the perspective of human resource when the weak areas are used to provide further assistance and training for teachers (Darling-Hammond, 2010; Ehren & Shackleton, 2016; Teodorović, Stanković, Bodroža, Milin & Đerić (2016).

There is convergence found in the results of the quantitative and qualitative analysis. Both results in the analysis points on factors that shape teacher's performance. The evaluations using students and self-evaluation explains that students gender and faculty's specialization and degree influences how ratings are provided. The qualitative analysis explains how the benefits, regulations, and attitude of teachers shape the evaluation process.

The findings in the evaluation show that when the rating is done by students or self-assessment, there is a degree to which the results would vary. The qualitative results explains that these variations would account for how much value and importance is given to evaluation when it is conducted. The findings in the evaluation show how the teacher becomes better is explained by the experience and degree gained by the teacher. The qualitative analysis shows that teacher performance improved when the evaluation becomes instrumental when the appraisal system helps teachers through formative and trainings.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The following conclusions are made given the results of the study:

1. Personal and social competency is highly valued by students and teachers in the teaching practice. This is evident in the high mean scores in the student and teacher evaluation.
2. Students background when evaluating the teacher performance matters. More specifically, females give higher levels of evaluation, and teachers belonging in the sciences were rated higher.
3. Self-evaluations made by the faculty is influence by their years of experience, higher academic degree, and belonging in a specific area of specialization.
4. Some form of bias exists when teachers evaluate themselves. The results of the evaluation was higher when the teacher evaluated themselves.
5. Teacher recognize importance of the TPE when it contributes to their performance, when regulations are present and when implemented well.
6. The attitude of teachers on the TPE is shaped by the obstacles and recommendations the evaluation provides.
7. TPE improves teacher performance when it is used as formative assessment, for training, and when enhancing the teaching and learning process.

The students of the UNTL (N = 342) assigned a very positive and favorable evaluation to the performance of the teachers' competences. The average of the evaluations of pedagogical, professional, social and teacher competences ranged from (3) "sufficient" and (4) "good". In analyzing how students measured and evaluated the competences of public higher education university teachers, it was possible to conclude about the quality of the performance of this teaching system in East Timor, although it is worth pointing out some improvements resulting from the students' evaluation of the teacher's competences.

In pedagogical terms, lower scores were observed (3) "sufficient" in relation to the teacher's ability to create a living environment in classes (M = 3.49, SD = 1.05), to the use of pedagogical means and technologies (M = 3.36, SD = 1.20) and the assignment of feedback on the students' work and evaluations (M = 3.33, SD = 1, 10). These results, therefore, support the need to introduce improvements in empathy and proximity to students, both through the promotion of a living environment of learning in the classroom and through the assignment of feedback to students. This

position was reinforced by the fact that the students classified as "sufficient" ($M = 3.39$; $SD = 1.00$) the ability of the teacher to implement a classroom order environment (in terms of personality competencies). It was also highlighted the need of teachers to use more innovative and inclusive pedagogical practices, and to take greater advantage of available educational pedagogy resources, as well as introduce more diversified pedagogical resources.

Regarding professional competences, it was concluded that there is a need to reinforce the use of ICT in the classroom, since this was one of the items with the lowest classification ($M = 3.38$, $SD = 1.13$), as well as the capacity for interdisciplinary explanations of the given subject ($M = 3.48$, $SD = 1.01$). The gender of the students showed to have a significant relation with the evaluation attributed to the pedagogical, professional and social competences of the teachers ($p < .05$). In the pedagogical and social competences, the male students presented statistically lower scores ("bad"), while the female gender had a significant association with the higher "good" scores for occupational and social competences. The age of the students did not present a significant association with the evaluation attributed to the teachers' performance ($p > .05$). The Faculty from which the students came had a statistically significant association with the students' evaluation of the four sets of competences ($p < .01$). The students of the Faculty of Philosophy have appeared associated to the most negative classifications: from "bad" to pedagogical, professional and personality competences, as well as to "very bad" classification for personality competences. Students of the Faculty of Political Science also emerged statistically associated with the evaluation of "bad" and "very bad" regarding the performance of teachers' social competences and the evaluation of "bad" personality competences. Associated with the "good" classifications were the students of the Faculty of Medicine for the four groups of competences and "very good" for the pedagogical competences. The students of the Faculty of Education, Arts and Humanities presented a significant association with the evaluation of "very good" attributed to the competencies of the teachers' personality.

The results also pointed to the need to deepen the interdisciplinary between the different areas of study of the curriculum, reinforcing the transversal learning. To this end, it is considered necessary the continuous and professional formation of teachers, an educational project that the UNTL has already tried to answer through the creation of the Center for Advance Teaching and Learning (CATL) - which aims to Improvement of the quality of UNTL teachers, based on the conviction that there is a close relationship between the didactic performance of the teacher and the performance of the student (UNTL, 2015).

The social competences of the teachers were the most valued by the students, as they received superior and close classifications of (4) "good" and showed a strong correlation with the personality traits ($r = ,682$) Those who had a lower weight on the pedagogical competences ($r = ,551$). For students, social competences and personality traits of the teacher correlate strongly, but social competences have a reduced influence on the evaluation of the pedagogical competences by the students.

In practical terms, the study's conclusions support the effective improvement of teachers' competences, insofar as they inform students about the pedagogical and professional performance, as well as the social and personality competences of the teachers. The classification attributed by the students allowed identifying the competences that require improvements, and these will be considered in the elaboration of the professional development plan of the teachers of the UNTL. A positive evaluation of public higher education in East Timor today prevails, despite the recognition of the need to introduce improvements in teachers' performance, especially in terms of pedagogical practices and pedagogical resources used in the classroom.

These results also highlight the need to consolidate the different types of competencies of university professors in the country, in order to improve their performance in a continuous way, and, through this, to promote a quality university education throughout the country, both in institutions Public institutions, or in private institutions, according to the Strategic Plan for National Education 2011-2030 (METL, 2011). The policy measures introduced in the meantime have led to an improvement in the higher education system, but there is still a need to make progress in improving teacher performance through professional development programs (UNESCO, 2013). One of the key purposes of the public higher education evaluation policy in East Timor is to improve the quality of higher education, to inform students (and other stakeholders) about the real quality of institutions and to encourage students to participate in this process, as they are the main beneficiaries of quality assessment. We consider that, at this level, the study developed makes a relevant contribution since, for the first time, it presents the results of the evaluation of the students about the competences of the teachers and, by this way, it gives the students the voice to express themselves about the quality of the Public higher education of East Timor.

Since we are living in the era of student centered education, the qualities and attributes that students need to see in an ideal medical teacher should be acquired by them. The faculty development programs, both pre- and in-service, will have to be tailored to meet these new realities. Going a step further, the performance monitoring and evaluation of medical teachers will have to

be synchronized with the training programs as well. The authors of this empirical study defend a teacher development program focus on defining the top attributes of performance and personality identified by the students, like communication skills, updated knowledge and expertise in their own subject and specialty. These should be blended with the knowledge, skills and competencies based on what the students “should and need to know” as safe practicing physicians in the future, than simply imparting the knowledge teachers are expert in without considering students’ needs. For example, one of the top performance attributes was ‘understand the role of teacher’, so that is important to know what students actually understand by that construct. Knowledge of that information would help teachers understand the areas where their roles are adequate and the areas where improvement is necessary (Al-Mohaimed & Khan, 2014).

The results obtained through the students evaluation allowed to outline a general analysis framework on the evaluation of the performance of UNTL teachers in terms of their pedagogical, professional, social and personal competences. The influence of personal variables such as gender and age, on the one hand, and the context variable as the college attended by the students were studied.

The issue of assessing teacher performance and improving the quality of higher education institutions is important for East Timor. After independence, many teachers who were foreigners left the country, leaving a strong gap in the education and training of young people. Today the country faces skills gaps and a reduced ability to respond to the needs of the labor market. Reconstruction of national institutions and infrastructures requires qualified training for young people and a superior quality system with the same standards as other universities in Asia and Europe.

In terms of the public policies to be implemented in the area of education, the study supports the need for greater investment in higher education, with policies and measures oriented to the reality of young Timorese. This investment in higher education should begin with the training and qualification of teachers, with active methodologies and supervised teaching practices that promote the improvement of skills and techniques undertaken in the classroom. In order to promote continuous teacher training, performance evaluation should be implemented throughout the higher education system, encompassing all teachers and pedagogical supervisors. It is intended that the evaluation of teachers' performance be consistent, that is to say, it will result in a professional training plan aiming at the continuous improvement of the competences and capacities of the teachers, in their varied dimensions. Finally, in order for the implementation of the performance

evaluation cycle to be successful, that is, to provide for the effective improvement of the entire education system, the evaluation procedures, the role of evaluators and the evaluation criteria to be considered. The results of teacher assessment, which should also address students' feedback, should be disseminated according to the performance evaluation schedule in order to contribute to the accountability of the public service with the regularity and transparency that is sought.

As the quality of higher education depends heavily on the quality of teachers' competences, and therefore improvements in teaching also mean improvements in student learning (Minelli, Reborá and Turri, 2015, Brewer, Knoeppel and Lindle, 2015).

Although educational policies aim to establish an increasingly strong causal relationship between evaluation, professional development and the improvement of the quality of teaching, this relationship has not been studied in the higher education system in East Timor, a broader and for which we sought to make a contribution throughout the study.

In future research, empirical studies on the impact of the quality evaluation system on public higher education are planned, based on the evaluation and feedback of the various stakeholders in this process: besides students, teachers, course directors and the directors of colleges. Through in-depth interviews, it will be possible to evaluate teachers' receptivity, as well as the difficulties experienced and the added value identified in the evaluation process.

The self-evaluation study of the teachers' competences allowed to analyze the self-assessment scores of the teachers of the UNTL (N = 192) regarding their pedagogical, professional, personality and social performance. There was a marked positive self-evaluation, where the mean scores ranged from (3) "sufficient" to (4) "good". Most of the teachers who participated in the study rated the "good" level as the four skill sets. In a second phase of the analysis, the parallelism with the results of the evaluation to the teachers of the same university, but made by the students, allowed to observe a differential between the evaluation of the students and the self-evaluation of the teachers.

The self-assessment of the competences of UNTL teachers (N = 192) revealed a very positive and satisfactory appreciation of the pedagogical competences, as well as the competences of personality and social competences, since the average values (of each item, as well as of the set of competencies) were between "good" and "very good" (4) level, with items with an average rating approaching "very good" (5) being more frequent. The self-evaluation of the professional competences, in addition to registering lower average values in relation to the others, presented a reduced variability of the data, with the average of the classifications being located in the "good"

level (4). In order to better understand the self-assessment of UNTL teachers, it was sought to verify if the demographic and socio-professional characteristics influenced the self-assessment of competences. There was no statistically significant association between the self-evaluation of teachers' competences and the variables gender and age group ($p < .05$).

In the higher education system of East Timor, the teacher emerges as the professional educator and scientist, whose mission is to transform, develop and disseminate science, technology and art through education, teaching and social service. By professional educator is meant the one that has as professional activity the production of knowledge. The competence of the teacher is defined as a set of characteristics, knowledge and qualifications that they must possess and demonstrate in the exercise of their activity, pedagogical, social and professional level. The competence of the teacher is decisive for the quality and implementation of the three pillars of higher education, namely teaching / pedagogy; the investigation; and community service.

The university professor is one of the key players in the Higher Education System, since its functions, duties and responsibilities are crucial for the achievement of national objectives in terms of educational policy and the guarantor of a qualification of the population (in the various domains Technical, scientific, technological, artistic, civic, religious, among others) with a view to ensuring development based on the values of justice and citizenship. In this perspective, professional, competent and quality university teachers are needed.

The evaluation of the competences of the teachers should be used to focus on the competences and to achieve improvements, based on the commitment to want to improve. The evaluation of teachers' competences should be based on rigorous principles and criteria such as transparency, requirement and objectivity, with a view to promoting the professional development of teachers on the basis of recognition of merit, effort and excellence. In addition to these principles, the evaluation must be viewed in an integrated and contextualized way, associated to the evaluation of the university institution itself and based on a prior and clearly defined professional profile, adequate to the functions carried out by the teacher.

The quality of teachers' skills and the quality of their teaching depend on the development of effective and meaningful teaching practices for students. Therefore, it is necessary for teachers in all disciplines to know the best teaching practices, as well as the strategies that allow them to develop effective teaching behaviors and capacities. Although faculty members of higher education are considered experts in their field of study, many may not have been trained in effective teaching practices, such as sharing their scientific knowledge, promoting research with students, and

engaging them in the work of scientific production. Therefore, the establishment and evaluation of a faculty development program can assist in the training and adoption of best instructional practices as well as enhancing the skills of teachers.

These results allow us to support a teacher evaluation model oriented towards the improvement of schools, since these correlations reflect a "synergistic relationship that will surely favor the systemic possibility of the various actors and educational agents to achieve their objectives" (Alves and Machado, 2010, p. 91).

Recommendations

Given the results of the study, the following future studies are recommended:

1. Investigate further the validity and reliability of the instrument used. At present, the background the students are used as a form of contrast validity. Confirmatory factor analysis and item analysis needs to be conducted.
2. Investigate the accuracy of the appraisal by correlating it with other measures of teacher performance.
3. Determine which items are easy and difficult to attain by using an Item Response Theory Model.
4. Use the instrument to relate to other factors with dependent variables such as student performance in order to validate the instrumental value shown in the qualitative analysis.
5. It is suggested to address the influence of the qualities of the teachers in the students' results. This analysis, together with the results of teachers' self-assessment and student assessment, allows the development of teacher training and professional development programs at a later stage. In designing and evaluating a new faculty development program, we hope to gain a better understanding of the impact of development programs on faculty competencies and student outcomes.

Given the results, the following policy recommendations are:

6. Use the items in the instrument to set teaching standards in the university. These standards need to be oriented among the faculty members so that they would be aware on the requirements on their teaching.
7. The results of the evaluation can be used to decide on faculty promotion, reclassification, and retention.
8. A system of faculty evaluation needs to be designed such as observation practices, classroom administration, and timing of using the scale within the semester.

REFERENCES

- Abma, T. A., & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, 7–22.
- Afonso, A. J. (2009). Nem tudo o que conta em Educação é mensurável ou comparável. Crítica à accountability baseada em testes estandardizados e rankings escolares. *Revista Lusófona de Educação*, 13(13), 13–29.
- Afonso, A. J. (2011). Questões polémicas no debate sobre políticas educativas contemporâneas; o caso da accountability baseada em testes estandardizados e rankings escolares. In M. P. Alves & J.-M. De Ketele (Eds.), *Do currículo à avaliação, da avaliação ao currículo* (pp. 83–101). Porto: Porto Editora.
- Alberello, L., & W, A. (1997). Pesquisa e crítica das fontes de documentação nos domínios económico, social e político. In *Práticas e Métodos de Investigação em Ciências Sociais* (pp. 15–47). Lisboa: Gradiva.
- Albergaria-Almeida, P., Martinho, M., & Cabrita, I. (2014). Evaluating the impact of restructuring Secondary Education in East Timor. *Procedia-Social and Behavioral Sciences*, 141, 665–669.
- Albergaria Almeida, P., Martinho, M., & Lopes, B. (2013). Teacher Professional Development in the Context of International Cooperation: opportunities and challenges. *Professional Development: Perspectives, Strategies and Practices*, 147–164.
- Alhija, F. N.-A. (2017). Teaching in higher education: Good teaching through students' lens. *Studies in Educational Evaluation*, 54, 4–12.
- Alkin, M. C. (2013). *Evaluation roots: A wider perspective on theorists' views and influences*. Thousand Oaks, CA: SAGE Publications.
- Alkin, M. C., & Christie, C. A. (2004). An Evaluation Theory Tree. In M. C. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences* (pp. 12–65). London: Sage Publications.
- Alkin, M. C., & King, J. A. (2016). The Historical Development of Evaluation Use. *American Journal of Evaluation*, 34, 568–576.
- Al-Mohaimed, A. A., & Khan, N. Z. (2014). Perceptions of Saudi medical students on the qualities of effective teachers: A cross-sectional study. *Saudi Medical Journal*, 2, 183–188.
- Alves, M. (2010). Aprendizagem ao longo da vida: entre a novidade e a reprodução de velhas desigualdades. *Revista Portuguesa de Educação*, 23, 7–28.
- Alves, M. P., & Aguiar, J. (2013). O sistema de avaliação de desempenho docente: tensões e desafios nas escolas e nos professores. In E. A. Machado, N. Costa, & M. P. Alves (Eds.), *Avaliação do desempenho docente: compreender a complexidade, sustentar a decisão* (pp. 125–166). Santo Tirso: De Facto Editores.

- Alves, M. P., Flores, M. A., & Machado, E. A. (2011). *Quanto vale o que fazemos? Práticas de avaliação de desempenho*. Santo Tirso: De Facto Editores.
- Alves, M. P., & Machado, E. A. (2011). O sentido do currículo e o sentido da avaliação. In M. P. Alves & J.-M. De Ketele (Eds.), *Do currículo à avaliação, da avaliação ao currículo* (pp. 59–70). Porto: Porto Editora.
- Amado, J. S. (2000). A técnica da análise de conteúdo. *Referência*, 5, 53–63.
- Anderson, M., Ingram, J. M., Buford, B. J., Rosli, R., Bledsoe, M. L., & Onwuegbuzie, A. J. (2012). Doctoral students' perceptions of characteristics of effective college teachers: A mixed analysis. *International Journal of Doctoral Studies*, 7, 279–309.
- Araújo, P. (2013). Avaliação de desempenho docente na perspetiva dos professores avaliadores. In E. A. Machado, N. Costa, & M. P. Alves (Eds.), *Avaliação do desempenho docente: compreender a complexidade, sustentar a decisão* (pp. 65–81). Santo Tirso: De Facto Editores.
- Audi, R. (2015). *The Cambridge Dictionary of Philosophy* (3rd ed.). Cambridge: Cambridge University Press.
- Bakar, K. A., Safavi, S. A., Tarmizi, R. A., & Alwi, N. H. (2011). Student Ratings of Instruction from Perspectives of Higher Education Administrators. *Journal of Academic Administration in Higher Education*, 7, 61–69.
- Ball, S. J., & Goodson, I. F. (1985). *Teachers' lives and careers*. London: Falmer Press.
- Bardin, L. (2013). *Análise de Conteúdo*. Lisbon: Edições 70.
- Baron, M. A., & Boschee, F. (1995). *Authentic assessment: The key to unlocking student success*. Lanham: Rowman & Littlefield.
- Barroso, J. (2005). O Estado, a educação e a regulação das políticas públicas. *Educação & Sociedade. Revista de Ciências Da Educação. Campinas (Brasil): Centro de Estudos Educação E Sociedade (CEDES)*, 26, 725–751.
- Baynham, M., & Stevens, M. (2014). Are We Planning Effectively for Climate Change? An Evaluation of Official Community Plans in British Columbia. *Journal of Environmental Planning and Management*, 57, 557–587.
- Berelson, B. (1984). *Content analysis in communication research*. New York: Hafner.
- Berke, P., Smith, G., & Lyles, W. (2012). Planning for Resiliency: Evaluation of State Hazard Mitigation Plans under the Disaster Mitigation Act. *Natural Hazards Review*, 13, 139–149.
- Best, J. W. (1982). *Como investigar en educación*. Madrid: Ediciones Morata.
- Bisquerra, R. (1989). *Métodos de Investigación Educativa: Guía Práctica*. Barcelona: Ediciones CEAC.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay.

- Bogdan, R., & Biklen, S. (1994). *Investigação qualitativa em educação: Uma introdução à teoria e aos métodos*. Porto: Porto Editora.
- Bolivar, A. (2007). Um olhar actual sobre a mudança educativa: onde situar os esforços da melhoria? In C. Leite & A. Lopes (Eds.), *Escola, currículo e formação de identidades: Estudos de investigação* (pp. 12–50). Porto: Edições Asa.
- Bouckaert, G., & Halligan, J. (2008). *Managing Performance: International Comparisons*. London: Routledge/ Taylor and Francis.
- Boulmetis, J., & Dutwin, P. (2011). *The ABCs of Evaluation: Timeless Techniques for Program and Project Managers* (2nd ed.). San Francisco: Jossey-Bass.
- Boyne, G. A., Kenneth, J. M., O'Toole Jr., L. J., & Walker, R. M. (2006). *Public Service Performance: Perspectives on Measurement and Management*. Cambridge: Cambridge University Press.
- Bradley, L. H. (1985). *Curriculum Leadership and Development Handbook*. Englewood Cliffs: Prentice Hall.
- Brandt, R. S. (1981). *Applied strategies for curriculum evaluation*. Alexandria, VA: ASCD.
- Brintnall, M., & Mealy, K. A. (2014). Advancing teaching and learning in political science: The role of the American political science association. *European Political Science*, 13, 159–170.
- Bryman, A. (1998). *Quality and Quality in Social Research*. London: Unwin Hyman.
- Candau, V. (2013). Currículo, didática e formação de professores: uma teia de ideias-força e perspectivas de futuro. In M. R. Oliveira & J. A. Pacheco (Eds.), *Currículo, didática e formação de professores* (pp. 7–19). São Paulo: Papirus Editora.
- Cardoso, A. (2012). *A. Avaliação do desempenho docente e o professor titular: um estudo de caso numa perspetiva organizacional. Tese de Doutoramento* (Dissertação de Doutoramento). Aveiro. Universidade de Aveiro.
- Carmo, H., & Ferreira, M. (1998). *Metodologia da Investigação: Guia para auto-aprendizagem*. Lisboa: Universidade Aberta.
- Carneiro, M. (2011). *Avaliação de desempenho e clima de escola* (Dissertação de Mestrado). Porto. Instituto Superior de Educação e Trabalho.
- Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890-1930*. New York: New York University Press.
- Chelimsky, E., & Shadish, W. R. (1997). *Evaluation for the 21st Century: A handbook*. Thousand Oaks: Sage.
- Chelimsky, E. (2012). Balancing evaluation theory and practice in the real world. *American Journal of Evaluation*, 34, 91–98.
- Chen, H. T., Donaldson, S. I., & Mark, M. M. (2011). Validity frameworks for outcome evaluation. *New Directions for Evaluation*, (130) 5–16.

- Christensen, K. S. (2015). Both Process and Outcome Are Essential to Planning. *Journal of Planning Education and Research*, 35(2), 188–198.
- Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation*, 34, 131–135.
- Christie, C. A., & Alkin, M. C. (2013). An Evaluation Theory Tree. In M. C. Alkin (Ed.), *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences* (pp. 12–65). Los Angeles: University of California.
- Churton, M., & Brown, A. (2010). *Theory and method (2^o ed.)*. New York: Palgrave Macmillan.
- Chyung, S. Y. (2015). Foundational Concepts for Conducting Program Evaluations. *Performance Improvement Quarterly*, 27, 77–96.
- Clarke, D. J., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18, 947–967.
- Coe, C. (2003). A Report Card on Report Cards. *Public Performance Management Review*, 27, 53–76.
- Coelho, A. C. A. (2011). *Avaliação do desempenho docente: a realidade no 1.º Ciclo* (Dissertação de Mestrado). Coimbra: Universidade de Coimbra.
- Coelho, M. J. C. (2013). *Avaliação de Desempenho Docente: Efeitos no Desenvolvimento Profissional* (Dissertação de Mestrado). Escola Superior de Educação de Lisboa, Lisboa.
- Contandriopoulos, D., & Brousselle, A. (2012). Evaluation model and evaluation use. *Evaluation*, 18, 61–77.
- Cookingham, F. G. (2012). Model as a Concept in Evaluation. *Discourse World Vision International*, 25, 1–11.
- Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. *New Directions for Evaluation*, 87, 27–34.
- Cousins, J. B. (2004). Crossing the Bridge: Toward Understanding through Systematic Inquiry. In M. C. Alkin (Ed.), *Evaluation Roots* (pp. 293–318). Thousand Oaks, CA: Sage.
- Cousins, J. B., Goh, S. C., Elliott, C., & Aubry, T. (2013). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, 44, 1–13.
- Coutinho, C. M. (2005). *Percursos da investigação em Tecnologia Educativa em Portugal*. Braga: Universidade do Minho, Centro de Investigação em Educação.
- Coutts, L. M., & Schneider, F. W. (2004). Police officer performance appraisal systems: How good are they? *Policing: An International Journal of Police Strategies & Management*, 27, 67–81.
- Craft, A. (2000). *Continuing Professional Development: A practical guide for teachers and schools*. London: Routledge Falmer.

- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: Sage Publications.
- Cronbach, L. J. (1957). Beyond the two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.
- Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist*, *30*, 1–14.
- Cronbach, L. J. (1980). *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L. J. (1991). Functional Evaluation Design. In W. R. Shadish, T. D. Cook, & Laura C. Leviton (Eds.), *In Foundations of Program Evaluation: Theories of Practice* (pp. 323–376). Newbury Park, CA: Sage.
- Cronbach, L. J., & Associates. (1980). *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L. J., & Suppes, P. (1969). *Research for tomorrow's schools: Disciplined inquiry for education*. New York: MacMillan.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Virginia: Educational Testing Service Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Center for American Progress.
- Day, C. (1999). *Continuing Professional Development*. London: Falmer Press.
- Day, C. (2001). *Desenvolvimento profissional de professores. Os desafios da aprendizagem permanente*. Porto: Porto Editora.
- Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research* *37*, 677–692.
- Day, C. (2004). Evaluer la formation des enseignants dans le contexte d'une école en changement en vue de favoriser le développement professionnel des acteurs et d'améliorer leur efficacité? In L. Paquay (Ed.), *L'évaluation des enseignants* (pp. 94–119). Paris: L'Harmattan.
- Day, C. (2010). Teachers in the Twenty-first Century: Time to renew the vision. *Teachers and Teaching: Theory and Practice*, *6* (1), 101–115.
- Day, C., Elliot, B., & Alison, K. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, *21* (5), 563–577.

- Day, C., Flores, M. A., & Viana, I. (2007). Effects of national policies on teachers' sense of professionalism: findings from an empirical study in Portugal and in England. *European Journal of Teacher Education, 3*, 249–265.
- Day, C., & Gu, Q. (2007). Variations in the conditions for teachers' professional learning and development: sustaining commitment and effectiveness over a career. *Journal Oxford Review of Education, 33* (4), 423–443.
- Day, C., Kington, A., Stobart, G., & Sammons, P. (2006). The personal and professional selves of teachers: stable and unstable identities. *British Educational Research Journal, 32* (4) 601–616.
- Day, C., & Sachs, J. (2004). Professionalism, Performativity and Empowerment: Discourses in the Politics, Policies and Purposes of Continuing Professional Development. In C. Day & J. Sachs (Eds.), *International Handbook on the Continuing Professional Development of Teachers* (pp. 3–32). Maidenhead: Open University Press.
- Day, C., Sammons, P., Stobart, G., Kington, A., & Gu, Q. (2007). *Teachers Matter: Connecting Lives, Work and Effectiveness*.
- De Ketele, J. M. (2010). A avaliação do desenvolvimento profissional dos professores: postura de controlo ou postura de reconhecimento? In M. P. Alves & E. A. Machado (Eds.), *O pólo de excelência. Caminhos para avaliação do desempenho docente* (pp. 13–31). Porto: Edições Asa.
- De Ketele, J., & Roegiers, X. (1999). *Metodologia da Recolha de Dados*. Lisboa: Instituto Piaget.
- Demirkasimoğlul, N. (2010). Defining “Teacher Professionalism” from different perspectives. *Procedia Social and Behavioral Sciences, 9*, 2047–2051.
- Donaldson, S. I., Christie, C. A., & Mark, M. (2008). *What Counts as Credible Evidence in Evaluation and Evidence-based Practice?* Thousand Oaks: Sage.
- D.Tucker, P., & Stronge, J. H. (2005). *Linking Teacher Evaluation and Student Learning*. Alexandria: Association for Supervision and Curriculum Development.
- Edmonds, S., & Lee, B. (2002). Teacher feelings about continuing professional development. *Education Journal, 61*, 28–29.
- Ehren, M. C. M., & Shackleton, S. N. (2016). Risk-based school inspections: impact of targeted inspection approaches on Dutch secondary schools. *Educational Assessment Evaluation and Accountability, 28*, 299–321.
- Eisner, E. W. (1979). *The educational imagination: On the design and evaluation of school programs*. New York: Macmillan.
- Englund, T. (1996). Are Professional Teachers a Good Thing? In I. F. Goodson & A. Hargreaves (Eds.), *Teachers' Professional Lives*, p. 27. London: Falmer Press.
- Erasmie, T., & Lima, L. (1989). *Investigação e projectos de desenvolvimento em educação*. Braga: Universidade do Minho.

- Esteves, J. M. (2000). The transformation of the teachers' role at the end of the twentieth century: new challenges for the future. *Educational Review*, 2, 197–207.
- Esteves, M. (2006). Análise de conteúdo. In J. A. Lima & J. A. Pacheco (2006). *Fazer investigação contributos para a elaboração de dissertações e teses*. Porto: Porto Editora.
- Estrela, A., & Ferreira, J. (2001). *Investigação em educação: Métodos e Técnicas*. Lisboa: Educa Lisboa.
- Estrela, M. T. (2011). *Profissão docente. Dimensões afectivas e éticas*. Porto: Edições Asa.
- Fernandes, S. (2010). *A aprendizagem baseada em projectos no Contexto do Ensino Superior. A avaliação de um dispositivo pedagógico no Ensino de Engenharia* (Dissertação de Doutoramento). Braga. Universidade do Minho.
- Fernet, C., Gagné, M., & Austin, S. (2010). When does quality of relationships with coworkers predict burnout over time? The moderating role of work motivation. *Journal of Organizational Behavior*, 31, 1163–1180.
- Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education*, 28, 514–525.
- Ferrero, D. J. (2006). Having it all. *Educational Leadership*, 63 (8), 8–15.
- Fetterman, D. M. (2004). Branching Out or Standing on a Limb: Looking to Our Roots for Insight. In C. M. Alkin (Ed.), *Evaluation Roots* (pp. 304–318). Thousand Oaks, CA: Sage.
- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage Publications.
- Filipe, J. M. (2011). *O relator enquanto supervisor na avaliação de desempenho docente: uma relação entre a teoria e a prática* (Dissertação de Mestrado). Leiria. Escola Superior de Educação.
- Fitzpatrick, J. L. (n.d.). An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, 2012(135), 7–24. <https://doi.org/10.1002/ev.20024>
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines*. New Jersey: Pearson Education.
- Flores, M. (2002). *Learning, development and change in the early years of teaching: A two-year empirical study* (Doctoral Dissertation). University of Nottingham: School of Education.
- Flores, M. A. (2010). *Avaliação de professores numa perspetiva internacional. Sentidos e implicações*. Porto: Porto Editora.
- Fonseca, M.J., Gonçalves, M.F., Mouraz, A., & Ramalho, H. M. P. (2004). *Avaliação: rotas e viajantes*. Viseu: ISPV.
- Fortin, M. F. (2000). *O Processo de Investigação. Da Conceção à realização*. Loures: Lusodidacta.
- Freixo, M. (2009). *Metodologia Científica. Fundamentos Métodos e Técnicas*. Lisboa: Instituto Piaget.

- Fullan, M. (2011). *Change leader : learning to do what matters most*. San Francisco: Jossey-Bass.
- Fullan, M., Rincón-Gallardo, S., & Hargreaves, A. (2015). Professional capital as accountability. *Education policy analysis archives*, 23, 15.
- Fullan, M. (2015). *The new meaning of educational change*. Routledge.
- Gamlem, S. M., Smith, K., & Engelsen, K. S. (2017). *Responsive Pedagogy and Student Learning: Feedback practice and student learning in Mathematics*. San Antonio, Texas: American Educational Research Association.
- Garcia, M. M. R. (2011). *Avaliação do desempenho docente: as lógicas profissionais na voz crítica dos professores* (Dissertação de Mestrado). Universidade de Lisboa, Lisboa.
- Goddard, I., & Emerson, C. (1995). *Appraisal and your school*. Oxford: Heinemann.
- Gomes, M. (2010). *Avaliação do desempenho docente – objetivos e controvérsias* (Dissertação de Mestrado). Universidade Lusófona de Humanidades e Tecnologias, Lisboa.
- Goodson, I. (2000). The principled professional. *Prospects*, 20, 181–188.
- Goodson, I. F. (2014). Investigating the life and work of teachers. *Eesti Haridusteaduste Ajakiri*, 2, 28–47.
- Goodson, I. F. (2015). *Narrativas em educação: a vida e a voz dos professores*. Porto: Porto Editora.
- Graça, A., Duarte, A. P., Lagartixa, C., Tching, D., Tomás, I., Almeida, J., ... Santos, R. (2011). *Avaliação do Desempenho Docente. Um Guia para a Acção*. Lisboa: Lisboa Editora.
- Greene, J. C. (1994). The Art of Interpretation, Evaluation, and Presentation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Qualitative Program Evaluation: Practice and Promise*. Thousand Oaks, CA: Sage.
- Greene, J. C. (2011). The construct(ion) of validity as argument. *New Directions for Evaluation*, 130, 81–91.
- Green, J. C., & Abma, T. A. (2001). Editors' notes. In J. C. Greene & T. A. Abma (Eds.), *Responsive evaluation. New Directions for Evaluation* (pp. 1–6).
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage Publications.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Guba, E., & Lincoln, Y. S. (2005). Paradigmatic Controversies, Contradictions, and Emerging. In N. Denezin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research Methods* (pp. 191–206). Thousand Oaks: Sage.
- Guyadeen, D., & Seasons, M. (2018). Evaluation Theory and Practice: Comparing Program Evaluation and Evaluation in Planning. *Journal of Planning Education and Research*, 38 (1), 98–110.

- Hadji, C. (1994). *A avaliação, regras do jogo. Das intenções aos instrumentos*. Porto: Porto Editora.
- Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: development, reliability and validity. *Procedia Social and Behavioral Sciences*, 15, 592–599.
- Hammersley, M. (2013). *The Myth of Research-Based Policy & Practice*. London: Sage.
- Hargreaves, A. (2000). Four ages of professionalism and professional. *Learning Teachers and Teaching: Theory and Practice*, 6 (2), 151–182.
- Harris, A., & Chapman, C. (2002). Leadership in schools facing challenging circumstances. *Management in Education*, 16(1), 10-13.
- Harris, A. (2002). *Leadership in Schools Facing Challenging Circumstances*. Copenhagen.
- Hedges, L., Pigott, T., Polanin, J., Ryan, A., Tocci, C., & Williams, R. (2016). The question of school resources and student achievement: A history and reconsideration. *Review of Research in Education Centennial Edition*, 40, 143–168.
- Heller, K. A., Holtzman, W. H., & Messick, S. (1982). *Placing children in special education: A strategy for equity*. (K. A. Heller, W. H. Holtzman, & S. Messick, Eds.). Washington, DC: National Academy Press.
- Hendriks, M., Luyten, H., Scheerens, J., Slegers, P., & Steen, R. (2010). *Teachers' professional development: Europe in international comparison: an analysis of teachers' professional development based on the OECD's Teaching and Learning International Survey (TALIS)*. Luxembourg: Office for Official Publications of the European Union.
- Herdeiro, R. (2013). *Avaliação do Desempenho Docente*. Lisboa: Chiado Editora.
- Hood, C. (1996). Exploring Variations in Public Management Reform of the 1980s'. In J. Perry H. Bekke & T. Toonen (Eds.), *In Civil Service Systems in Comparative Perspective* (pp. 268–317). Indianapolis: Indiana University Press.
- House, E. R. (2012). Introduction to Section 1: Evaluation Theory. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International Handbook of Educational Evaluation. Part One: Perspectives* (pp. 9–14). Dordrecht: Kluwer Academic Publishers.
- Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying Public Policy: Policy Cycles and Policy Subsystems*. Toronto: Oxford University Press.
- Hustler, D., McNamara, O., Jarvis, J., Londra, M., Campbell, A., & Howson, J. (2003). *Teachers' Perspectives of Continuing Professional Development: DfES Research Report No. 429*. London: DfES.
- Ivaldi, S., Scaratti, G., & Nuti, G. (2015). The practice of evaluation as an evaluation of practices. *Evaluation*, 21, 497–512.
- Jacob, S., & Boisvert, Y. (2010). To be or not to be a profession: pros, cons and challenges for evaluation. *Evaluation*, 16, 349–369.

- Jaquith, A., Mindich, D., Wei, R.C., & Darling-Hammond, L. (2010). *Teacher professional learning in the United States: Case studies of state policies and strategies*. Oxford: Learning Forward.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *American Educational Research Association, 7*, 14–26.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects and materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personal evaluation standards*. Thousand Oaks: Sage Publications.
- Joint Committee on Standards for Educational Evaluation. (1994). *Program evaluation standards*. Thousand Oaks: Sage Publications.
- Jónsson, I. R., Smith, K., & Geirsdóttir, G. (2018). Shared language of feedback and assessment. Perception of teachers and students in three Icelandic secondary schools. *Studies in Educational Evaluation, 56*, 52–58.
- Ka-ho, M. (2003). Similar Trends, Diverse Agendas: Higher education reforms in East Asia. *Globalisation, Societies and Education, 1*, 201–221.
- Kellaghan, T., Stufflebeam, D. L., & Wingate, L. A. (2003). Introduction. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International Handbook of Educational Evaluation. Part One: Perspectives* (pp. 1–6). Dordrecht: Kluwer Academic Publishers.
- Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. New York: Routledge Falmer.
- Korte, L., Lavin, A., & Davies, T. (2013). Does gender impact business students' perceptions of teaching effectiveness? *Journal of College Teaching and Learning, 10*, 167–178.
- Kozikoğlu, İ. (2017). Prospective Teachers' Cognitive Constructs Concerning Ideal Teacher Qualifications: A Phenomenological analysis Based on Repertory Grid Technique. *International Journal of Instruction, 10*, 63–78.
- Krathwohl, D. R. (1964). *Taxonomy of Educational Objectives: Handbook II, Affective Domain*. New York: Longman.
- Kupermintz, H. (2003). Lee J. Cronbach's contributions to educational psychology. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology: A century of contributions* (pp. 289–302). Mahwah, NJ, US: Erlbaum.
- K. Yin, R. (1998). The abridged version of case study research: Design and method.
- Laurian, L., Crawford, J., Day, M., Kouwenhoven, P., Mason, G., Ericksen, N., & Beattie, L. (2010). Evaluating the Outcomes of Plans: Theory, Practice and Methodology. *Environment and Planning B: Planning and Design, 37*, 740–757.

- Lavin, A., Korte, L., & Davies, T. (2012). Student Gender and Perceptions of Teaching Effectiveness. *Research in Higher Education Journal, 18*, 1–16.
- Lay, C. S. (2016). *The effectiveness of assessing teacher dispositions in the teacher selection processes of Taiwan's junior high schools* (Doctoral Dissertation). USA. Morehead State University.
- Ledermann, S. (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation, 33*(2), 159–178.
- Leibowitz, B., Bozalek, V., Schalkwyk, S. V., & Winberg, C. (2015). Institutional context matters: the professional development of academics as teachers in South African higher education. *Higher Education, 69*, 315–330.
- Lima, A. J. (2006). Ética na investigação. In A. J. Lima & A. J. Pacheco (Org.). *Fazer investigação contributos para a elaboração de dissertações e teses*. Porto: Porto Editora.
- Lima, L. (2011). Avaliação, competitividade e hiperburocracia. In M. P. Alves & J.-M. De Keteke (Eds.), *Do currículo à avaliação, da avaliação ao currículo* (pp. 71–82). Porto: Porto Editora.
- Lüdke, M., & Andrew, M. (1986). *Pesquisa em educação: Abordagens qualitativas*. São Paulo: Editora Pedagógica e Universitária.
- Luo, H. (2010). The Role for an Evaluator: A Fundamental Issue for Evaluation of Education and Social Programs International. *Education Studies, 3*, 42–50.
- Machado, E. A. (2013). *Avaliar é ser sujeito ou sujeitar-se?* Mangualde: Edições Pedagogo.
- Machado, E. A., Abelha, M., Barreira, C., & Salgueiro, A. (2012). Avaliação pelos pares: percurso normativo da avaliação do desempenho docente em Portugal. *Revista Portuguesa de Pedagogia, 46*, 73–93.
- Machado, E. J., & Abelha, M. (2014). A avaliação de professores: que lições do caso português? *Olh@res, 2*, 55–80.
- Magno, C., & Sembrano, J. (2007). The role of teacher efficacy and characteristics on teaching effectiveness, performance, and use of learner centered practices. *The Asia-Pacific Education Research, 16*, 73–90.
- Marcos, A. (2013). *Lógicas de supervisão pedagógica em contexto de avaliação de desempenho docente* (Dissertação de Doutoramento). Porto. Universidade Portucalense.
- Mathinson, S. (2009). Public Good and Private Interest: A History of Educational Evaluation. In W. C. Ayers, T. Quinn, & D. O. Stovall (Eds.), *The Handbook of Social Justice in Education* (pp. 5–14). London: Routledge.
- Matsuno, A. (2008). The UN Transitional Administration and democracy building in Timor-Leste. In D. J. Mearns & S. Farram (Eds.), *Democratic governance in Timor-Leste: Reconciling the local and the national*. Darwin: Charles Darwin University Press.

- McLaughlin, M. W., & Mitra, D. (2001). Theory-based Change and Change-based Theory: Going Deeper, Going Broader. *Journal of Educational*, 2, 301–323.
- Mertens, D. M. (2015). Philosophical Assumptions and Program Evaluation. *Spazio Filosofico*, 75–85.
- Mertens, D. M., & Wilson, A. T. (2012). *Program Evaluation Theory and Practice*. New York: Guilford.
- Mesquita, E., Formosinho, J., & Machado, J. (2012). Supervisão da prática pedagógica e colegialidade docente. A perspetiva dos candidatos a professores. *Revista Portuguesa de Investigação Educacional*, 12, 59–77.
- Miller, R. L., King, J. A., Mark, M. M., & Caracelli, V. (2016). The Oral History of Evaluation: The Professional Development of Robert Stake. *American Journal of Evaluation*, 37, 287–294.
- Millo, Y., & Barnett, J. (2004). Educational development in East Timor. *International Journal of Educational Development*, 24, 721–737.
- Monyatsi, P. P. (2006). An investigation of the effectiveness of the schoolbased workshops approach to staff development in secondary schools in Botswana. *Educational Research and Reviews*, 1, 150–155.
- Monyatsi, P., Steyn, T., & Kamper, G. (2006). Teacher perceptions of the effectiveness of teacher appraisal in Botswana South African. *Journal of Education*, 26, 427–441.
- Morais, F., & Medeiros, T. (2007). *Desenvolvimento profissional do Professor. A chave do problema?* Ponta Delgada: Universidade dos Açores.
- Morgado, J. C. (2011). Identidade e profissionalidade docente: sentidos e (im)possibilidades. *Ensaio*, 19, 793–812.
- Morgado, J. C. (2014). Avaliação e qualidade do desenvolvimento profissional docente: que relação? *Avaliação*, 19, 345–361.
- Morgado, J. C. (2016). *O estudo de caso na investigação em educação*. Santo Tirso: De Facto Editores.
- Morgado, J. C., & Carvalho, E. (2012). Políticas educativas e avaliação do desempenho docente: as teias de uma problemática. In M. Alvez (Ed.), *Avaliação em Educação: políticas, processos e práticas* (pp. 75–102). Santo Tirso: De Facto Editores.
- Morgan, C. (2001). The Effects of Negative Managerial Feedback on Student Motivation: Implications for Gender Differences in Teacher–Student Relations. *Sex Roles*, 44, 513–535.
- Muijs, D., Day, C., Harris, A., & Lindsay, G. (2004). Evaluating Continuing Professional Development: An Overview. In C. Day & J. Sachs (Eds.), *International Handbook of the Continuing Professional Development of Teachers* (pp. 291–310). Maidenhead: Open University Press.

- Murdock, G. (2000). Introducing a teacher supportive evaluation system. *ETL Journal: An International Journal for Teachers of English Speakers of Other Languages*, 54, 54–64.
- Neuman, L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches*. Edinburgh: Pearson Education Limited.
- Nevgi, A., & Lofstrom, E. (2015). The development of academics' teacher identity: Enhancing reflection and task perception through a university teacher development program. *Studies in Educational Evaluation*, 46, 53–60.
- Newman, I., & Ridenour, C. (1998). Qualitative-Quantitative Research Methodology: Exploring the Interactive Continuum. *Educational Leadership Faculty Publications*, (122).
- Nie, Y., Chua, B. L., Yeung, A. S., Ryan, R. M., & Chan, W. Y. (2015). The importance of autonomy support and the mediating role of work motivation for well-being: Testing self-determination theory in a chinese work organisation. *International Journal of Psychology*, 50, 245–255.
- Nóvoa, A. (1999). Os professores na virada do milénio: do excesso dos discursos à pobreza das práticas. *Educ. Pesqui. [online]*, 25, 11–20.
- Oliveira-Formosinho, J. (2009). Desenvolvimento profissional dos professores. In J. Formosinho (Ed.), *Formação de Professores. Aprendizagem profissional e acção docente* (pp. 221–284). Porto: Porto Editora.
- Oliveira, M. E. S. (2012). *O caminho labiríntico da avaliação do desempenho docente: um estudo com professores de línguas estrangeiras* (Dissertação de Mestrado). Universidade do Minho, Braga.
- Oliveira, M. R. O. (2013). *Professor: formação, saberes e problemas*. Porto: Porto Editora.
- Oliveira, V., & Pinho, P. (2011). Bridging the Gap between Planning Evaluation and Programme Evaluation: The Contribution of the PPR Methodology. *Evaluation*, 17, 293–307.
- Owston, R. (2008). Models and Methods for Evaluation. In J. M. Spector, M. D. Merrill, J. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology (Chapter 45)* (pp. 605–617). Toronto: York University.
- Pacheco, J. A. (1998). Avaliação da aprendizagem. In *Conhecer, aprender e avaliar*. (Leandro Almeida; José Tavares (Org.), pp. 111–132). Porto: Porto Editora.
- Pacheco, J. A. (2011). Currículo, Aprendizagem e Avaliação. Uma abordagem face à agenda globalizada. *Revista Lusófona de Educação*, 75–90.
- Pacheco, J. A., & Flores, M. A. (1999). *Formação e avaliação dos professores*. Porto: Porto Editora.
- Pal, L. A. (2014). *Beyond Policy Analysis: Public Issue Management in Turbulent Times*. Toronto: Nelson.
- Pamela, D. T., & Stronge, J. H. (2005). *Linking Teacher Evaluation and Student Learning*. Alexandria: Association for Supervision and Curriculum Development (ASCD).

- Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford.
- Patton, M. Q. (2012). Contextual pragmatics of valuing. *New Directions for Evaluation*, 133, 97–108.
- Patton, M. Q. (2012). *Essentials of Utilization-Focused Evaluation*. London: SAGE Publications.
- Perry, J. L. (2010). A strategic agenda for public human resource management research. *Review of Public Personnel Administration*, 30(1), 20-43.
- Picciotto, R. (2011). The logic of evaluation professionalism. *Evaluation*, 17, 165–180.
- Pollit, C., & Bouckaert, G. (2011). *Public Management Reform. A Comparative Analysis—New Public Management, Governance, and the Neo-Weberian State, 3rd Ed.* New York: Oxford University Press Inc.
- Popham, J. W. (2001). *The truth about testing: An educator's call to action*. Alexandria: ASCD.
- Prestes, E. M. da T. (2012). A avaliação de políticas sociais: questões teóricas, metodológicas e políticas. *Revista Lusófona de Educação*, 22, 85–101.
- Provus, M. M. (1969). *The Discrepancy Evaluation Model: An Approach to Local Program Improvement and Development*. Berkley: McCutchan.
- Pullman, E. (2018). Applying the Framework as a reflective tool: A teacher learner perspective. *College & Research Libraries News*, 79, 416–419.
- Quiyv, R., & Campenhoudt, L. V. (2008). *Manual de Investigação em Ciências Sociais*. Lisbon: Gradiva.
- Ramalho, H. (2014). Contributos Teóricos e Metodológicos Para a Referencialização da Avaliação do Desempenho do Docente. *Revista Eletrónica de Educação E Psicologia*, 1, 12–21.
- Reis, F. (2010). *Como elaborar uma dissertação de mestrado* (2nd ed.). Lisboa: Pactor Editora.
- Ribeiro, F. M. (2011). *Supervisão entre pares: contributos para a melhoria das práticas de ensino* (Dissertação de Mestrado). Porto. Universidade Portucalense.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 905–947). Washington D.C.: American Educational Research Association.
- Rodrigues, A., & Peralta, H. (2008). *Algumas considerações a propósito da avaliação do desempenho dos professores*. Ministério da Educação de Portugal (DGRHE).
- Rodrigues, P. (1999). A Avaliação Curricular. In A. Estrela & A. Nóvoa (Eds.), *Avaliação em Educação: Novas Perspectivas* (pp. 15–57). Porto: Porto Editora.
- Rodríguez-Campos, L. (2012). Advances in collaborative evaluation. *Evaluation and program planning*, 35(4), 523-528.

- Roe, J. W. (2010). *Student self-assessment and student ratings of teacher rapport in secondary student course ratings*. Utah: Utah State University, ProQuest Dissertations Publishing.
- Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsí, T. A. (2000). Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation*, 87, 5–13.
- Roldão, M. C. (2013). Considerações finais – questões emergentes. In A. C. Salgueiro, A. Marcos, C. Barreira, E. A. Machado, F. L. Da Silva, F. Mesquita Alves, & N. Costa (Eds.), *Avaliação do desempenho docente: compreender a complexidade, sustentar a decisão* (pp. 167–174). Santo Tirso: De Facto Editores.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach* (7th ed.). Thousand Oaks: Sage.
- Salgueiro, A., & Costa, J. A. (2013). Lógicas de ação organizacional na avaliação de desempenho docente: dados de um estudo de caso sobre o primeiro ciclo avaliativo. In E. A. Machado, N. Costa, & M. P. Alves (Eds.), *Avaliação do desempenho docente: compreender a complexidade, sustentar a decisão* (pp. 91–107). Santo Tirso: De Facto Editores.
- Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2017). Two teacher quality measures and the role of context: evidence from Chile. *Educational Assessment, Evaluation and Accountability*, 29, 111–146.
- Saunders, M. (2012). The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421–436.
- Schalkwyk, S. V., Herma, N., Leibowitz, B., & Farmer, J. (2015). Reflections on professional learning: Choices, context and culture. *Studies in Educational Evaluation*, 46, 4–10.
- Scheerens, J., Glas, C., & Thomas, S. M. (2005). *Education, Evaluation, Assessment, and Monitoring. A Systemic Approach*. The Netherlands. Swets & Zeitlinger Publishers.
- Scott, D. (2016). Introduction: Curriculum, Learning and Assessment. In *New Perspectives on Curriculum, Learning and Assessment*. (R. Sharon & M. Peters, (Eds.)). London: Springer.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Scriven, M. (1974). *Evaluation: A study guide for educational administrators*. Davie: Nova University.
- Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation*, 4, 393–404.
- Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary view. In I. Donaldson & S. Scriven (Eds.), *Evaluating Social Programs and Problems: Vision for the New Millennium* (pp. 19–42). Mahwah: Erlbaum.
- Scriven, M. (2007). The logic of evaluation. In H. V. Hansen (Ed.), *Dissensus and the Search for Common Ground* (pp. 1–16). Windsor: OSSA.

- Scriven, M. (2012). Evaluation Theory and Metatheory. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation. Part One: Perspectives* (pp. 15–30). Dordrecht Kluwer: Academic Publishers.
- Shadish, W. (1998). Evaluation Theory is Who We Are. *American Journal of Evaluation*, 19(1), 1–19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). Good Theory for Social Program Evaluation. In W. R. Shadish, T. D. Cook, & L. C. Leviton (Eds.), *Foundations of Program Evaluation: Theories of Practice* (pp. 36–67). Newbury Park: Sage.
- Shadish, W. R., & Luellen, J. K. (2005). History of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 183–186). Thousand Oaks, CA: Sage.
- Shaw, I. F., Greene, J. C., & Mark, M. M. (2005). *The Sage handbook of evaluation*. London: Sage.
- Shepard, L. A. (1993). Evaluating Test Validity. *Review of Research in Education*, 19, 405–450.
- Shepard, L. A. (2016). Testing and Assessment for the Good of Education: Contributions of AERA Presidents, 1915–2015. *Educational Researcher*, 45(2), 112-121.
- Simões, G. (2000). *A avaliação do desempenho docente: contributos para uma análise crítica*. Lisboa: Texto Editora.
- Sousa, J. M. (1997). Investigação em educação: novos desafios. In A. Estrela & J. Ferreira (Orgs). *Métodos e técnicas de investigação científica em educação*. (PP. 661-672). Lisboa: AFIRSE Portuguesa. FPCE – Universidade de Lisboa.
- Stake, R. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523–540.
- Stake, R. (1974). Program evaluation: Particularly responsive evaluation. *New Trends in Evaluation, Report*, 6, 1–20.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stake, R. (1999). *Investigación con estudio de casos*. Madrid: Morata.
- Stake, R. E. (1971). *Measuring what learners learn*. Urbana: Center for Instructional Research and Curriculum Evaluation.
- Stake, R. E. (1975). To Evaluate an Arts Program. In R. E. Stake (Ed.), *Evaluating the Arts in Education: A Responsive Approach*. p.122. Columbus: Merrill.
- Stake, R. E. (1980). Program Evaluation, Particularly Responsive Evaluation. *Rethinking Educational Research*, 72–87.
- Steiner, G. (2005). *As lições dos Mestres*. Lisboa: Gradiva.
- Steinert, Y., Cruess, S., Cruess, R., & Snell, L. (2005). Faculty development for teaching and evaluating professionalism: From programme design to curriculum change. *Medical Education*, 39, 127–136.

- Stevens, M. (2013). Evaluating the Quality of Official Community Plans in Southern British Columbia. *Journal of Planning Education and Research*, 33, 471–490.
- Stronge, J. H. (2010). O que funciona, de facto, na avaliação de professores: breves considerações. In M. A. Flores (Ed.), *Avaliação de professores numa perspetiva internacional. Sentidos e implicações* (pp. 22–43). Porto: Porto Editora.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. New York: Eye on Education.
- Stufflebeam, D. (2007). *CIPP Evaluation Model Checklist. A tool for applying the CIPP Model to assess long-term enterprises*. Evaluation Checklists Project.
- Stufflebeam, D. L. (1971). The use of experimental design in educational evaluation. *Journal of Educational Measurement*, 8, 267–274.
- Stufflebeam, D. L. (1980). Interview: An EEPA interview with Daniel L. Stufflebeam. *Educational Evaluation and Policy Analysis*, 2 (4), 85–90.
- Stufflebeam, D. L. (1983). The CIPP Model for Program Evaluation. In G. F. Madaus (Ed.), *Evaluation Models*. pp. 117.141. Boston: Kluwer.
- Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 89, 7–98.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan (Ed.), *The international handbook of educational evaluation (Chapter 2)*. Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L., & Madaus, G. F. (1983). The Standards for Evaluation of Educational Programs, Projects, and Materials. In *Evaluation Models. Evaluation in Education and Human Services*, 6. (pp. 395-404). Springer, Dordrecht.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models and applications*. San Francisco: Jossey-Bass.
- Suknaisitha, A., Wongwanichb, S., & Piromsombatc, C. (2014). Development of teacher performance in educational measurements and evaluation through self-monitoring strategies. *Procedia - Social and Behavioral Sciences*, 116, 1683–1688.
- Talbot III, R. M. (2011). *Validity issues in the evaluation of a measure of science and mathematics teacher knowledge*. Colorado: University of Colorado at Boulder, ProQuest Dissertations Publishing.
- Tardif, J., & Foucher, C. (2010). Um conjunto de balizas para avaliação da profissionalidade docente. In M. P. Alves & E. A. Machado (Eds.), *O pólo de excelência. Caminhos para avaliação do desempenho docente* (pp. 32–53). Porto: Edições Asa.
- Teodorović, J., Stanković, D., Bodroža, B., Milin, V., & Đerić, I. (2016). Education policymaking in Serbia through the eyes of teachers, counselors, and principals. *Educational Assessment, Evaluation and Accountability*, 28, 347–375.
- Terman, L. M. (1919). *The intelligence of school children*. Boston: Houghton Mifflin.

- Terman, L. M. (1922). *Intelligence tests and school reorganization*. Yonkers-on-Hudson: NY: World Book.
- Tiantong, M., & Tongchin, P. (2013). A Multiple Intelligences Supported Web-based Collaborative Learning Model Using Stufflebeam's CIPP Evaluation Model International. *Journal of Humanities and Social Science*, 3, 157–165.
- Tokmak, H. S., Baturay, H. M., & Fadde, P. (2013). Applying the Context, Input, Process, Product Evaluation Model for Evaluation, Research, and Redesign of an Online Master's Program. *The International Review of Research in Open and Distributed Learning*, 14, 273–293.
- Tuckman, B. (2002). *Manual de Investigação em Educação*. Lisboa: Fundação Calouste Gulbenkian.
- Tuinamuana, K. (2011). Teacher professional standards, accountability and ideology: Alternative discourses. *Australian Journal of Teacher Education*, 36 (12), 72–82.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35, 492–501.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Tyler, R. W. (1969). *Education Evaluation: new Roles Means*. Chicago: University of Chicago.
- Valencia, R. R. (1997). *The evolution of deficit thinking: Educational thought and practice*. London: Routledge Falmer.
- Varela-Petido, G. (2011). Evaluating public higher education in Mexico. *Higher Education Management and Policy*, 23(2), 1–20.
- Vedung, E. (2010). Four Waves of Evaluation Diffusion. *Evaluation*, 3, 263–277.
- Vilelas, J. (2009). *Investigação. O Processo de Construção do Conhecimento*. Lisboa: Edições Sílabo.
- Villar Angulo, L. M. (1987). *Evaluación diagnóstica de los procesos mentales de los profesores*. Sevilla: Publicaciones de La Universidad de Sevilla.
- Weiss, C. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs: Prentice-Hall.
- Weiss, C. (1998). *Evaluation Research*. New Jersey: Prentice Hall.
- Weiss, C. H. (1993). Where Politics and Evaluation Research Meet. *Evaluation Practice*, 14 (1), 93-106.
- Wiener, D. G. (2015). What students think about good teaching and who is the excellent teacher? *Teaching in the Academia*, 1, 31–40.
- Yin, R. K. (2009). *The abridged version of case study research: Design and method* (2nd ed.). Sage publications.

- Young, M. (2006). Education, knowledge and the role of the state: The “nationalisation” of educational knowledge. In A. Moore (Ed.), *Schooling, Society and Curriculum*. (pp. 33-44). London: Routledge Falmer.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Product Evaluation Model (CIPP) as a Comprehensive Framework to Guide the Planning, Implementation, and Assessment of Service-learning Programs. *Journal of Higher Education Outreach and Engagement*, 15, 57–84.

Normative Documents

- Ministry of Education Timor Leste (METL) (2011). National Education Strategic Plan 2011-30. Díli, Timor Leste.
- República Democrática de Timor Leste (RDTL) (2002). East Timor 2020: Our Nation, Our Future, 2002-2020. Díli, Timor-Leste.
- República Democrática de Timor Leste (RDTL) (2011). Timor-Leste Strategic Development Plan 2011-2030. Version Submitted to the National Parliament. Díli, Timor-Leste.
- República Democrática de Timor Leste (RDTL) (2014). Decreto-Lei N.º 3 /2014 de 15 de Janeiro (Primeira alteração ao Decreto-Lei n.º 7/2012, de 15 de Fevereiro, que estabelece o Estatuto de Carreira Docente Universitária.
- República Democrática de Timor Leste (RDTL) (2014). Diploma Ministerial N.º 33 /ME/2014 de 10 de Setembro (Procede à aprovação do Manual de Certificação do Docente Universitário - Manual CEDU).
- UNESCO (2013). Review of national policies on learning and teaching. Paper commissioned for the EFA Global Monitoring Report 2013/4, Teaching and learning: Achieving quality for all.
- Universidade Nacional de Timor Leste (UNTL) (2015). Relatório do Inquérito Pedagógico da UNTL - Implementação do Currículo de 2014. Díli, UNTL.

APPENDICES

Appendice 1. Questionário sobre avaliação de desempenho de professores

PARTE 1. DADOS PESSOAIS / DADOS DE IDENTIFICAÇÃO

Faculdade : _____ Sexo: M ___ F ___ Idade: _____

Local : _____ Data: _____

PART 2. QUESTIONÁRIO SOBRE AVALIAÇÃO DE DESEMPENHO DE PROFESSORES

Guião

De acordo com seu conhecimento, faça uma avaliação justa, objetiva e responsável do seu ensino. Suas informações são estritamente confidenciais e limitadas apenas ao processo de avaliação. A avaliação é feita sobre os aspectos listados na tabela a seguir, simplesmente marque um círculo em torno de um único número de 1 a 5, a coluna de pontuação. Cada número corresponde à seguinte classificação:

1 = Muito ruim 2 = Ruim 3 = Suficiente 4 = Bom 5 = Muito bom

Nº	Aspetos da avaliação	Pontuação
A. Competência Pedagógica		
1	Preparação das aulas e práticas	1 2 3 4 5
2	Organização e disciplina nas aulas	1 2 3 4 5
3	Capacidade para criar um ambiente vivo nas aulas	1 2 3 4 5
4	Dar matérias e respostas claras às perguntas nas aulas	1 2 3 4 5
5	Aproveitamento dos meios e tecnologia pedagógica	1 2 3 4 5
6	Variedade nas formas de avaliação dos resultados de aprendizagem	1 2 3 4 5
7	Atribuição do feedbacks sobre os trabalhos/avaliação	1 2 3 4 5
8	Adaptação das matérias do exame/trabalho ao objetivo da disciplina	1 2 3 4 5
9	Harmonização das notas atribuídas na pauta com o resultado da aprendizagem	1 2 3 4 5
B. Competência Profissional		
10	Capacidade de explicar corretamente os conteúdo/abordagem/tópico/conceito	1 2 3 4 5
11	Capacidade de dar exemplos relevantes dos conceitos	1 2 3 4 5
12	Capacidade de explicações interdisciplinares do tópico atual com outros	1 2 3 4 5
13	Capacidade de explicações interdisciplinares do atual tópico com as realidades quotidianas	1 2 3 4 5
14	Domínio das questões mais avançadas da disciplina (matérias/referencias das aulas de aprendizagem).	1 2 3 4 5
15	Utilização dos resultados da investigação para promover a qualidade da aprendizagem nas aulas	1 2 3 4 5
16	Envolver estudantes na investigação/ análises ou desenvolver o desenho efetuado pelo professor	1 2 3 4 5
17	Capacidade do uso das várias tecnologias de comunicação	1 2 3 4 5
C. Personalidade		
18	Dignidade e personalidade como Docente	1 2 3 4 5
19	Honestidade e justiça na tomada da decisão	1 2 3 4 5
20	Ser modelo através da posição e comportamento	1 2 3 4 5
21	Palavra de ordem e medidas	1 2 3 4 5
22	Autoconfiança em qualquer situação	1 2 3 4 5
23	Justo e neutral com todos os estudantes	1 2 3 4 5
D. Competência Social		
24	Capacidade de expressão	1 2 3 4 5
25	Capacidade de aceitar críticas e sugestões dos estudantes	1 2 3 4 5
26	Conhecer bem os estudantes da sua disciplina	1 2 3 4 5
27	Conviver facilmente com os pares, funcionários e estudantes	1 2 3 4 5
28	Tolerância nas confissões religiosas dos estudantes	1 2 3 4 5

Appendice 2. Questionário sobre auto-avaliação dos professores

PARTE 1. DADOS PESSOAIS / DADOS DE IDENTIFICAÇÃO

Faculty: _____ Academic Background: _____ Sex: M ___ F _____

Age : _____ Professional situation _____ Title of the work _____

Years of Experience in the actual work _____ Venue: _____ Date: _____

PARTE 2. QUESTIONARIO SOBRE AUTO-AVALIAÇÃO DOS PROFESSORES

Guião

Faça uma avaliação justa, objetiva e responsável sobre sua própria pessoa. Suas informações são estritamente confidenciais e limitadas apenas ao processo de avaliação. A avaliação é feita nos aspectos listados na tabela a seguir, simplesmente marque um círculo em torno de um único número de 1 a 5, coluna de pontuação. Cada número corresponde à seguinte classificação:

1 = Muito ruim 2 = Ruim 3 = Suficiente 4 = Bom 5 = Muito bom

Nº	Aspetos da avaliação	Pontuação
A. Competência Pedagógica		
1	Seriedade da preparação das aulas (programas, plano de qualidade e sua implementação)	1 2 3 4 5
2	Organização e disciplina da efetuação das aulas (preencher e encontro graduais)	1 2 3 4 5
3	Adaptação da gestão das aulas com as metas de aprendizagem	1 2 3 4 5
4	Disciplina e cumprimento dos regulamentos académicos	1 2 3 4 5
5	Domínio do uso das médias e tecnologia de aprendizagem	1 2 3 4 5
6	Utilização de modos variados de avaliação do comportamento de estudo dos alunos	1 2 3 4 5
7	Objetividade de avaliação aos estudantes	1 2 3 4 5
8	Capacidade de orientar os estudantes	1 2 3 4 5
9	Perceção positiva sobre a capacidade dos alunos	1 2 3 4 5
B. Competência Profissional		
10	Domínio das áreas disciplinares relativas à sua função principal	1 2 3 4 5
11	Capacidade de explicações interdisciplinares do tópico atual com os outros	1 2 3 4 5
12	Capacidade de ilustração das áreas interdisciplinares com a realidade quotidiana	1 2 3 4 5
13	Domínio das questões (referências) potenciais nas áreas de aprendizagem	1 2 3 4 5
14	Disponibilidade de refletir e discutir os problemas enfrentados com os pares	1 2 3 4 5
15	Envolver os alunos na investigação /análises e o projeto desenvolvido pelo docente	1 2 3 4 5
16	Capacidade de conseguir o avanço da ciência e tecnologia para potencializar o processo de aprendizagem	1 2 3 4 5
17	Envolvimento nos trabalhos científicos nas organizações profissionais	1 2 3 4 5
C. Personalidade		
18	Dignidade e personalidade docente	1 2 3 4 5
19	Honestidade e justiça na tomada da decisão	1 2 3 4 5
20	Ser modelo do comportamento	1 2 3 4 5
21	Palavra de ordem e medidas	1 2 3 4 5
22	Autoconfiança em quaisquer situações	1 2 3 4 5
23	Neutro e justo com os pares	1 2 3 4 5
D. Competência Social		
24	Capacidade de expressão	1 2 3 4 5
25	Capacidade de aceitar críticas e sugestões dos estudantes	1 2 3 4 5
26	Fácil Convivência com os seus pares, funcionários e estudantes	1 2 3 4 5
27	Fácil convivência social	1 2 3 4 5
28	Tolerância sobre as diferentes confissões religiosas da sociedade	1 2 3 4 5

Appendice 3. Guião da Entrevista

GUIÃO DE ENTREVISTA A PROFESSORES AVALIADOS:

Este inquérito por entrevista realiza-se no âmbito do projeto de doutoramento em Ciências de educação especialidade desenvolvimento curricular Universidade do Minho Braga Portugal. Com os dados recolhidos pretendemos esboçar um modelo de avaliação desempenho e supervisão pedagógica adequado ao desenvolvimento profissional do professor/docente. Está garantido o anonimato e confidencialidade de todas as respostas dadas pelos participantes neste estudo.

PARTE 1: Dados Pessoais, Dados Profissionais, Formação académica e Processo de Avaliação de desempenho

1. Dados pessoais

1.1. Género do inquirido – Feminino ___ Masculino ___

1.2. Idade : Menos de 35 anos ___ Entre os 35 e os 45 anos ___ Entre os 45 e os 55 anos ___ Mais de 55 anos ___

2. Dados profissionais

2.1. Tempo de serviço na docência: Menos de 5 anos _ Entre 5 e 10 anos _ Entre 11 e 20 anos _ Entre 21 e 30 anos ___ Mais de 30 anos _____

2.2. Experiência de aulas observadas antes deste processo de avaliação de desempenho:
Sim ___ Qual? _____ Não ___

3. Formação académica

3.1. Habilitações académicas de habilitação para a docência: Bacharelato _ Licenciatura _ Mestrado _ Doutoramento ___ Outro ___

3.2. Outras habilitações académicas : Bacharelato ___ Licenciatura ___ Pós graduação ___ Mestrado ___ Doutoramento ___ Outro ___ Especificação: _____

PARTE 2: Representação do processo de avaliação por parte do professor/docente

Questões Gerais

Q1 Existe alguma das melhores práticas relacionadas à avaliação de desempenho do professor implementadas em seu departamento ou a nível da UNTL (Universidade Nacional de Timor Lorosae)?

Q2 Você acha que este é um aspecto importante a ser implementado em seu departamento ou nível universidade?

Q3 Antes de se tornar um professor, você tem alguma experiência em adquirir conhecimentos relacionado a habilidades pedagógicas?

Q4 Você acha que o treinamento pedagógico no futuro é importante para você?

Q5 Existe alguma regra, regulamentação ou legislação que regule a avaliação do desempenho do professor na Universidade?

Q6 Por que você acha que o TPE (Teacher Performance Evaluation) parece não ser muito aspecto importante na Universidade?

Q7. Há alguma recomendação que você possa fornecer?

Q8 Existe alguma questão chave e problemas que inibem a implementação do TPE? (Avaliação de desempenho)?

Q9. Existe algum mecanismo que já tenha sido adotado em relação a esse tipo de atitude dos professores?

Appendice 4. Interview transcript

PROFILE OF INTERVIEWEES

INTERVIEWEE 1. – Interviewed on 27th of June 2016 Time: 09.00 – 09.45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	42
2. Professional Data	
2.1 Years of work as Teacher	10
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	PhD/Doctorate
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any best practices related to the teacher performance evaluation implemented at your department or at UNTL (Universidade Nacional de Timor Lorosae) level?	A1, Well, as far as I know, I am not aware of it and we haven't done it. Is a good thing to do but unfortunately rarely we implement.
Q2. Do u think that this is an important aspect to be implemented at your department or University level?	A2. Indeed it is very important, however it has not been implemented. It is a good thing because it promotes people and encourages people to perform better. Because It is a kind of good competition.
Q3. Before you are becoming a lecturer do you have any experience in terms of acquiring knowledge related to pedagogical skills?	A3. No. Unfortunately not. Well, when I did my first, second and third degree basically deepen my knowledge in the area I am interested in. But I never planned in my life to be a lecturer, when I do my first degree, I never thought become a lecturer, but of course after second and third degree I know that I am a lecturer but still we just learn more about our field. About pedagogical stuff skills we need na extra training. I dont think we ever learn that at school.
Q4. Do you think pedagogical training in the future is important for you?	A4. Oh.. Yes absolutely.
Q5. Is there any rules, regulations, or legislations that regulates the teacher performance evaluation at the university.	A5. Well, I am not aware of it. We have a Vice Rector for quality control that Vice Rector should issue or make a regulation to control us but so far I did not see it.
Q6. Why do you think that the TPE(Teacher Performance Evaluation) seems not to be very important aspect at the University?	A6. It is important, but you know, we are new to this system you know. Some important aspects that are important but we don't do it yet, it does not mean it is not important but probably we don't put it as

Questions	Answers
	first priority. You know last year we did a training on ISO and that was a good beginning to start tidy up our management. Unfortunately that was just limited to a training, after that training there was no follow up action on it.
Q7. Is there any recommendations that you could provide?	A7. Well that ISO training should be followed up by action. Many of us receive ISO training, it was a good training, unfortunately if only training no action, what is the point. OK.
Q8. But, Is there any key issues and problems that inhibiting for the implementation of TPE (Teacher Performance Evaluation)?	A8. Well, some lecturers they have many activities even more activities out there then inside here, I think when the evaluation is done some people will complain a lot, but not by those who diligently come here, only by a few people who are naughty who are making this university like their part time job and then go somewhere else for permanent job. Only for those naughty guys.
Q9. Is there any mechanism that has already been taken towards that kind of attitudes of lecturers?	A9. Well, we talk to verbally with them that they should be here more than somewhere else.
Q10. Any changes?	A10. No changes yet. From the University also do like a few weeks ago or months ago there is a new department established by our new Rector called Controla interna, they try to control the presence list, basically they come every morning at 8 o clock they will collect the presence list so whoever is not there if you are late that's it you are absent and then 4 o clock they bring it back and you sign it back and they collect it at 5 o clock. But many lecturers reject it.
Q11. Any instance were the lecturers been dismissed because of such activity out the University?.	A11. Unfortunately not yet. No one. No measures is taken. No. As long as the law is like a rubber you can fall apart the rule or regulation has to be firm. When you do something must be followed by action. Otherwise people do not scare of

INTERVIEWEE 2. – Interviewed on 27th of June 2016 Time: 10.00 – 10:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	47
2. Professional Data	
2.1 Years of work as Teacher	16
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. How long have you been working in the University (UNTL).	A1. I have been at the University since this university established. I am with University since 2000 up to now. I am also involved and avail in establishing Education department. As Dean already 2 years.
Q2. Is there any implementation on Teacher Performance Evaluation with rules and procedures?.	A2. TPE, is an evaluation that its own rules and procedures. After the University become an autonomy institution based on statute, there is a career regime, and there is a unit of certification at university. There is no clear rules and criterion from the Units, or University to the faculties to implement process of evaluation that follows the character as an Institution as academics. Until now no TPE yet. As Dean I haven't done any evaluation to my lecturers. Though I personally evaluate that is not affecting their careers. I only observe their tasks daily, not evaluation written with certain formats
Q3. Do you see that the lecturers at the University (UNTL) have competences such as pedagogical competence, professional competence, social competence and personal competence in its implementation in the classroom.?	A3. In the Department of Education Lecturers with different background of studies. Composition of Lecturers here at the Department some do have pure knowledge background without pedagogical knowledge and some others have pedagogical knowledge, there is no doubt. We need to do innovations to ameliorate our knowledge, because the globalization world demand us to innovate our knowledge. Those who are not having knowledge on pedagogical skills sometimes they may find difficulties in transferring science to the students or in teaching students and including and make a good plan to teach students. Lecturers are encouraged to use multimethod in their teaching. There is a need to provide training or continuous formation or training to lecturers for those who

Questions	Answers
	does not have pedagogical knowledge to know more about pedagogical aspects. Even those who have already knowledge about pedagogical knowledge they also still need a continuous formation to deepen more their knowledge and improve their quality of teaching.
Q4. You as Dean why is that University has already established for a long time but there is no such TPE is taken place?	A4. TPE, first question is that when University established since in the beginning was not an autonomous institution which do our own internal evaluation but we adopt the general regime coming from the public services (função publica), Firstly, The public service itself start doing TPE to the lecturers is around 2009-2010. Secondly, in 2010 we start doing University Statute, thirdly, in 2012 University come up with small rules that regulates all the processes at the university. We establish on how we come up with rules to regulate all the processes and proceedings within the university. University must have a comprehensive preparation. This is something that University pay attention to. We already have time to evaluate our lecturers and think to provide further training or continuous formation for lecturers when we find gaps in our lecturers. When the rules are already in place, we can intensify the TPE, I request that the unit of certification to provide award or certificate to lecturers to give them incentive to better perform. In case if there something wrong then the institution should provide continuous training to lecturers. The objective of TPE is to provide venues for formation of good qualification and not to dismiss lecturers for their poor performance is to improve what is not yet achieved and take an effort to improve it.
Q5. So far is there any regulations or special legislation that regulates TPE or not?	A5. I notice that there are already some items for evaluation in the Manual of Certification, I don't know why the Manual is already in existence but why no such TPE taking place. I am not aware whether there is already any regulation. I have not seen one.
Q6. What are your key recommendations to the University Managers to better implement TPE at University?	A6. I think we need TPE annually, we should start, so that we can do continuous formation for our lecturers, often there are issues on quality. Unit that is responsible for this type of work then they must plan to carry out TPE and then provide venues for further training to improve lecturers knowledge.

INTERVIEWEE 3. – Interviewed on 28th of June 2016 Time: 10:00 – 10:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	57
2. Professional Data	
2.1 Years of work as Teacher	13
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. In your experience Is there any instance where TPE implemented at University level especially to the Lecturers to identify competences that they need before teaching?	A1. There is none.
Q2. Do you think this is important or there is a need to do TPE?	A2. But we need to do TPE to improve the teaching quality of lecturers. Those who thinks that they are not capable they must decide to other jobs rather than insist himself for teaching.
Q3. Are there competences such as Pedagogical competence, Professional Competence, Social competence and Personality competence in the part of lecturers?	A3. There is no such evaluation from the University that the lecturers have those competences mentioned. Only those in Faculty of education they are trained in pedagogical knowledge. Many of us at the university we don't have those competences. Not all of us were formed to become teachers. We don't have those competences. University said that will send the lecturers for training and formation in the area of pedagogy but so far University has not taken any mechanisms.
Q4. Is there any legislation that regulates TPE?	A4. Regime Career, Rules to evaluate teachers, Manual of Certification, these are tools that can use to evaluate lecturers to see whether they better perform or not then University can take mechanisms to provide venues for lecturers to have those competences mentioned. University must identify the lecturers those who are with competence and allow them to teach and those who are not must have further training on pedagogical issues. Aline their functions with the law.

Questions	Answers
Q5. Why University has been established for long time as Public University why no such TPE?	A5. Because we base on the General Career Regime. Now that we already have special carrier regime for lecturers and the lecturers are insisted to comply to rules and regulations and better perform their functions properly. We are in the stage of creating mechanisms.
Q6. Key recommendations to University to better implement TPE?	A6. My suggestion University must be based on the law, and University cannot send away those lecturers, because they have taught for years. I hope University could create mechanisms for lecturers to improve their personal growth so that they can provide better services in teaching and learning.
Q7. I heard that some lecturers were resistant to direct and indirect observation in the classroom?	A7. Why should be resistant? Lecturers must be gentle to acknowledge their own limitations and be open for change and at the same time reflect to perform better in teaching.
Q8. What is the role of ANAAA? Relating to TPE?	A8. ANAAA role is to evaluate the criterion of establishment of courses. Do institutional evaluation and programmatic evaluation
Q9. Do you think Ministry of Education do inspection at the University.	A9. We need an independent team to evaluate to avoid bias in judgement. We need external evaluation.

INTERVIEWEE 4. – Interviewed on 28th of June 2016 Time: 09:00 -09:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	48
2. Professional Data	
2.1 Years of work as Teacher	10
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	PhD/Doctorate
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any instance where UNTL was able to do some kind of TPE to measure some of the services that has been delivered at UNTL?.	A1. It's a new question because so far I have been here for a decade exactly 12 years. To the best of my knowledge there has been no any form of teacher performance evaluation at UNTL. Since I came into University as Junior lecturer. Up to now. I did not know if there has been evaluation that I did not come across.
Q2. What do you think and why is it that the TPE has not been taken place but the UNTL has already been established for years operating in the country?	A2. I don't know. Maybe, people in a situation where, there are complexity of issues. Majority of professors or lecturers had been 2 years ago undergraduate graduates and teaching undergraduates, Masters are emerging in the past 2 years. There are few PhDs now, I don't know statistics, although the number are increasing, there are issues like managerial issues, international partnerships, curriculum, pedagogy and so on. I don't know why University has not done any kind of TPE for lecturers, perhaps they know where the capacities are. Or whether this is considered to be something of need, a priority or not. Maybe for various reasons there has been no implementation.
Q3. But do you think that such TPE for the lecturers is important at the University level in the future or maybe along the process?	Q3. I think it's important, at least for transitions. I mean a lot of people been teaching or lecturing here they become lecturers by job. I mean by job opportunity offered by the UNTL and people probably came across their friends on the street, or family lines and so on that are recruited to the university, and then sometimes with low not intellectual capacity but low interest in teaching because they maybe suitable for other areas for work but not necessarily teaching.
Q4. Do you think that all lecturers at UNTL either those who are with pure science background and applied science background they have this experience in the area pedagogical skills or competence	Q4. There are problems. One is pedagogical competences, if you see the people who studied in educational science, they are probably more prepared, because they were trained in a way to become lecturers or teacher, and for those that are non-educational faculties, they don't have that pedagogical skills and majority of UNTL lectures

Questions	Answers
or professional competence in delivering teaching to students at university?	comes from that background.
Q5. If that is the case then how would measure the quality of service of each lecturer?	A5. Exactly, In my opinion if you do a pedagogical evaluation, aside from that background of education, it is better also good to consider their background of work. For example a government officer maybe a former director or an employee in the government office, certain government officers their educational background that is maybe suitable for teaching. Somebody else from other areas maybe have less knowledge on issues of pedagogy, but if they do involve in some sort of trainings, informal on job trainings, or giving trainings, the so called the training for trainers (TOT) develop by UN (United Nations) Capacity Development, that is one layer of experience. If there are people that come from maybe NGOs (non-governmental organizations). These people are not train by education in terms of pedagogy but they practically involve in those kind of practices such they use PRA (Participatory Rural Appraisal) or action research, they need to know Paulo Freire Methods, adult education area, so often they use those methods to interact with communities and they have experiences in basic pedagogy, if they become teachers at University then they tend to perform better that I can trust them and they are better than teachers that are straight forward coming from University without any educational background in teaching. We have three people here in the University, I know former Dean of Social Science Faculty, one other colleague named Mateus Tilman studying Masters student in ASI (Asian Social Institute) in the Philippines, then I ask the students and the students said that they like the way teaches in the class. Compare to other lecturers that without teaching experience. Experiences can make difference, in social interaction in the class, communication skills.
Q6. Do you think that continuous formation for lecturers at UNTL is important or crucial?	A6. Exactly, this should have been taken place but not because we don't have good system at University. Continuous formation, or training, use adult education method while working they learn, they teach and reflect, why you need a system? That's keep a critical dialogue going on in the faculties. Maybe in Education Faculty I don't know, Brazilian started with Brazilian East Timor Study Group 2 years ago and nobody from the lecturers joined and only two of us were there and there were no other lecturers, maybe because of language but I think is beyond language is attitudes, the interest of people to learn, to deepen their knowledge, there is the absence of

Questions	Answers
	willingness, even if there are initiatives people do not take the chance and embrace it. I think it is because the pedagogy officials, I call it officials, they do exist but they don't know what to do with pedagogy. In my opinion, you are in the position you can discuss with lecturers maybe 6 months once, how are things going on pedagogically, invite a more experience lecturer delivers a lecturer on pedagogy to everybody once in a year and review in initiate discussions on pedagogy and allow to people to reflect and write about that. I don't think something like that happen, there is no system in place, even if teaching, learning, on the job training are rare initiatives.
<p>Q7. Do you think that in terms of the existence of the statute of the UNTL or academic regulations or any other legislations that are already in existence that could provide as a reference for each of the lecturers at the University to be aware that they do need those competences to be implemented in their process of teaching and learning?</p> <p>Q8. What do you foresee and think is the key challenges for the university in relation to this teacher performance evaluation?</p>	<p>A7. Exactly, in fact there have been discussion in the Ministry of Education few years ago, there are three competencies such as language, intellectual and pedagogical skills, how do people understand this and take it seriously into everyday life as lecturer because my definition academics are people whose job is to learn all the time, teaching is one but learning is another.</p> <p>A8. I think one is mentality, that evaluations sometimes is seen as patronizing, people would tend to react, who are you to evaluate? I have been a teacher for ten years, there is a resistance and defense mechanism, but that is my assumption. But different ways we can do evaluation, depend on the methods, there are participatory methods that lecturer do not know that people are evaluating them, they tend to come and lecture you as lecturer but they are not aware that you are evaluating them and it is important I think also to make it into an attention that consider as a policy at the University so that we keep the control of the process of teaching by having direct and indirect observation in the classroom which is important. Some kind of policy that allows the university to have a group of people to put attention to decisions by regularly updating them.</p>
<p>Q9. Do you think this regime career or academic regulations consider it is enough to provide some kind of norms or regulations or frameworks?</p>	<p>A9. There is not been one. The career regime is for all lecturers it is new that is dealing more with salaries and certificates. There are so many complex issues related to evaluation.</p>
<p>Q10. Any key recommendations that you could suggest to the University to at least put this in place so that TPE will take place?</p>	<p>A10. First must make it a policy, a policy that can be applicable, can be implemented to faculty not just top down type of evaluation, that allows self-evaluation, autonomous, internal evaluation system within faculties so that people would think as their own and not patronizing, frameworks are developed, of course motivations issues, provide them certificate, or increase salaries or penalties</p>

Questions	Answers
	for those who are not performing well, but essentially you evaluate people and encourage people to love their jobs, that is my understanding.
Q11. Do you think one point of recommendation could be continuous training or formation?	A11. Exactly, frameworks that allows them to evaluate but also to do develop strategies to deal with their own findings. Mechanisms is needed to exist. Encourage culture of learning, our job at University is to learn, not only students who come to learn but we lecturers has to learn all the time to break down what Freire called the Banking Education System, where lecturers tends to know everything, but today is no longer, anybody can go to Google and then download and read them and go to the classroom and discuss with lecturers, if you don't read it you don't access information to it then you have less knowledge.
Q12. Do you think that research and writing up articles are important for the lecturers or not?	A12. There are not many lecturers write books, they mostly write thesis, Licentiate thesis, Master Thesis or PhD thesis but they do not have this culture of writing scientific articles, books, journals, daily newspapers, university reviews. Veritas Magazine of University only students are writing but not many lecturers. Maybe it is because no culture of writing. We hope change would come in the future.

INTERVIEWEE 5. – Interviewed on 28th of July 2016 Time:11:00 – 11:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	54
2. Professional Data	
2.1 Years of work as Teacher	16
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Q1. Does teacher performance evaluation takes place at your university such as UNTL?	Answers
	A1. Thank you very much. When we talk about TPE, the evaluation is can be done from teacher himself, or evaluation from others, evaluators or evaluation to students, where refer to the preparation of subjects, ie methodology of teaching, instruments used, and expression of teachers that are dynamic so that can bring the students to attain the objectives that has already determined in the plan of teaching, this whether like or not must reflect several things that theoretically the experts inculcate in us, in order to be a good teacher he or she has to do three things. 1) How the teacher can transmit the information, transmit knowledge related to cognitive aspect, and also a teacher is not enough only transfer science, but also how transform people or students in the area of 2) affective aspect so that he or she will be familiar to have love in order to understand the science. Another aspect 3) Psico-motoric aspect is to familiarize the students in relation to the pedagogical practice and didactical aspects through the teacher on how carry out their functions such as teacher, educator and trainer. This is a very important aspect but in reality in Timor Leste majority there are some people could attain 1/3 or 2/3 that are aware of and apply the teaching system, but some people are still only in the domain of transferring knowledge but not yet form the character of people or students so that would bring the students as a subject involve in the teaching and learning processes. However, when the teachers still act like a king,

Questions	Answers
	<p>father, or teacher to inject information to the students but do not teach the students to do thing by themselves, to find out by themselves what is valuable, and capable of exploring more their secret and the root of his or her knowledge. As we reflect that in many other places in every corner of the announcement board stated statement such as: YOU TELL ME AND I FORGET, SOMETIMES AS HUMAN BEING I FORGET, BUT IF YOU SHOW ME AND I REMEMBER. SHOW ME THROUGH SYMBOLIC SIGNS HELP ME REMEMBER AND MAKE ME UNDERSTAND. Not only that but it is also important to involve students in the teaching and learning processes so that they will be ACTIVE AND BE able to deepen their own knowledge always. Psychologically, ONE OF the psychologist said that if we ONLY HAVE INFORMATION FROM THE TEACHER WITHOUT MAKING daily review we may be lessened our knowledge by 15 %. When there no continuous review our knowledge may be lessen. Many teachers, as teachers of University have the awareness to change the students attitudes or skills but there are some factors that are not advantageous.</p>
<p>Q2. In terms of mutual evaluation like for example is there any forms of evaluation available where the students can evaluate the teachers and their performance? The content of evaluation covers the fundamental competences that a teacher/lecturer should have in order to help students so that they would progress better in the process f teaching and learning in the class room?</p>	<p>A2. As I see this type of evaluation with the presence of JICA (Japan NGO) teachers, they always do evaluation, because they submitted to me the form of evaluation, in education department some do evaluation but I notice that some of them in the meeting, majority of them do not apply such evaluation. There is a need to have mechanism through experts or through research to help on how to do teacher performance evaluation or teacher self-assessment.</p>
<p>Q3. In terms the presence of Career Regime that is already been implemented at your University UNTL, is that in the career regime there is no space in order to demand the teachers to whether like or not, perform their teaching in the classroom, more or less have fundamental competences not only dealing with salaries payment but demand teachers should have basic competences for him or her before becoming teacher in the classroom?</p>	<p>A3. Well. as we know after the Regime implemented many teachers still in the level of assistant, many still in probationary level, level C1 still less people and C2 enough. That means true teachers who are still assistant many times only assisting. Their availability is learning more and before teaching. In the regime for example If the C1 become orientador teacher, I want to say that there is limitation in the area of competences, even</p>

Questions	Answers
	<p>though in reality many teachers have experienced teaching for 5 to 10 years. Competences itself make some people morally do their functions but when look at the legislation may feel inadequate but the teacher also aware that the evaluation is not only composed of one person, managers, but also comes from other parties, in an integrated manner. In order to reach the level of a professional teacher or senior or C1 level try to do something. It is not yet a demand for the teachers. In the career regime relate to the level of salary, teachers already have their rights and duties to be performed. Sometimes, reflecting to the formation during Indonesia time, before becoming a teacher a person needs to have a certificate or a license in order for him or her to teach. He knows the situation and know how to prepare lessons, He or she already know how to teach and that helps him or her to understand on how to deliver information in the classroom, form and train and how to make a dynamic class.</p>
<p>Q4. Many teachers still do not have basic competences for teaching, so what are the mechanisms from the University and must provide to help the teachers to obtain basic competences before the teaching in the classroom.</p>	<p>A4. University has a positive regard to this issue by sending the teachers to have further studies to improve their professional growth. But the weaknesses is that the issue that the University has not aware of, that is the right man on the right place. Means that some subject that should be taught, case in a department that reflect on the all subjects included in the curriculum where the teacher who are already sent for studies should be properly replaced so that there would be no emptiness and sometimes it does not happen. Continuous formation to Doctoral or Masters studies, also not yet reflect the definition of whether send people to prepare themselves as administrators or teachers or as researcher. Looking at the three principles of the university, sometimes not yet really reflect on how to train teacher or continuous formation mostly in the area of didactics and pedagogy.</p>
<p>Q5. Based on your experience, aside from the career regime, is there any other legislation that already exist that would appear as a policy so that would also implemented the TPE?</p>	<p>A5. I think Government already have educational policy through Basic Laws on Education and Ministerial Diploma, Decree Law to have different types of formation</p>

Questions	Answers
	<p>teachers, but I have not seen continuous formation in all times either in trimestral or semester, always, or a need to have a place where teachers can always discuss important issues in their área in their department. From the government in the year 2000-2009 just started with institutional evaluation but not programmatic evaluation. In here, I am referring to an evaluation on higher education in general. In order to be seen as an Institution with quality must follow basic criterion or requirements established by the government the so called ANAAA.</p>
<p>Q6. What are the key recommendations that you suggest to UNTL so that all teachers must be given the chance to have basic competences before teaching in the classroom as I mentioned?</p>	<p>A6. There is a paradigm shift. Shift in the objectives of the process of teaching and learning where in previous time make people mastered and reproduce knowledge. But forward is not only mastered but discover and produce knowledge. Previously everything is focused in teacher (teacher-oriented), but nowadays must be student-oriented. Previously questions where oriented to 4 W but nowadays must focus on 5 W 1 H. This process of thinking previously make students memorize, know and understand, how the teachers should make students be creative, innovator, imaginative, in this process of developing their thinking, if to teach use inductive, if describe anything use deductive, or may be mix between inductive and deductive so that the students will have concrete idea and have his or her own conclusion. In the evaluation, previous system just uses True and False, in reality in order the student think better, must use open ended or close questions, important to obtain variable answers. Reflecting to the people 's thinking imparted to us, in order to obtain an education with quality there are already information published by UNESCO depend on the 10 factors such as: 1) focus in the personal relevancy, social relevancy. 2) Conviction, esteem and self-esteem from involving parties, teacher as resource and facilitators and students as subject of learning 3) Lack of Ethical Force and professionalism, there is an awareness, become mirror for the students, not reflect yet 4) lack of ability of directors with</p>

Questions	Answers
	<p>paternalism leadership 5) work as a team in school but sometimes also have individualism. 6) there is no reflection of relationship between educational agents, school or higher education, sometimes we notice there is a spirit of egoism between the governors where students become victims. Educational agents must be coming from vertical line as well as horizontal line. In our curriculum. Our curriculum keeps changing from time to time, sometimes also we who are following the Bolonha System but in our neighboring countries adopt Anglo-Saxon system. We are alone in the middle, for example we are going to be part of ASEAN. Between ASEAN members, within ASEAN I don't know what is a good system that can bring a vision attain our goals or objectives. Another is quantity, quality and disponibility or availability educative materials, such as in library, we are truly already entering in the world of CPLP, our concentration is more focused on information international with English language, at the same time we use Portuguese language, as well as using Indonesian language because most of basic education in science was with Indonesia. With the sophisticated technology, some of us are still left behind, Lack of didactic materials, lack of support materials, no laboratory for language, language where can make students become master for himself or herself in their practice.</p> <p>For UNTL based on the career regime already improve the prosperity of people, though not yet in the right manner and some still feel as victim. But we must already have support materials, incentives and so on.</p>
<p>Q7. What is the role of Division of Certification for Teachers or Lecturers at your university such as UNTL?</p>	<p>A7. Bring the teachers or lecturers to understand their own position and motivate them to have more deep knowledge to increase their level. There is no certification yet based on the diploma of the educational background of teachers. Teachers is not only teaching but must be also become researcher.</p>
<p>Q8. I heard that some teachers are resistant to have their performance being evaluated? Especially direct and indirect observation?</p>	<p>A8. Resistant because of their habits that they have. Reluctant to innovation. Preoccupied with materials conditions, conditions in Timor</p>

Questions	Answers
	<p>Leste demands a lot. Teaching sometimes demanded 18 to 30 class but they are not capable of complying to it therefore many of them are reluctant to change or observations. I can say there has been no teacher performance evaluation conducted to each teachers at all levels in the department. In terms of legislation there is code of conduct but it has not been disseminated to teachers and not yet use as law. Yearly evaluation that we have right now is still using the one used by Pubic Service Commission which does not reflect proper items to measure the performance of teachers at public university such as UNTL</p>

INTERVIEWEE 6. – Interviewed on 29th of June 2016 Time: 10: 00 – 10:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	57
2. Professional Data	
2.1 Years of work as Teacher	6
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Backgorund	Masters
3.2 Others academic backgorund	None
Questions	
Answers	
Q1. Is there any evaluation such as TPE taken place at UNTL?	A1. As far as I know there is no such TPE taken place at UNTL. I am aware that TPE is very important where can help lecturer to improve and become better in facilitating the learning process in the classroom. And TPE is also one way where one could be able to identify their own weaknesses and strengths and open for improvement of capacities and knowledge of lecturers. I have done interview and there is no TPE yet. Based on the orientation from Rector, I am assigned to open CPAI (Comissão Permanente Avaliação Institucional). This will function to take care of TPE at the University. The law is already being approved such Law no 16/2010 and Law no 21 such as ANAAA, but there is no implementation yet. So the existence of CPAI is important and it will help the process and how we will be able to evaluate the lectures in TPE.
Q2. Do the lecturers have basic competencies such as pedagogical competences, professional comtences, personality traits, social competences in facilitating them to better facilitate teaching and learning in the classroom?	A2. There are some lecturers who came from the Faculty of Education and they do have pedagogical skills for teaching, however those lecturers who are not from the faculty of Education they are oriented from the University level to provide to have a continuous formation in the area of pedagogical skills for all lecturers who are having lack of knowledge in aforementioned competencies. We are in the process identifying all lecturers and we are doing SWOT analysis and having more evidences on those who lack of knowledge in pedagogical skills to attend continuous formation.

Questions	Answers
<p>Q3. Is there any experiences where stakeholders such as NGO or International Universities come to UNTL to make evaluation or do research with regard to TPE and identify the existing gaps?</p>	<p>A3. There is no such external organization to do such TPE. I am not aware of. Only there are Masters graduates who were evaluated by the Human resource Management Office and found out that there are lecturers who are not obeying the rules and regulations such as using Portuguese Language and Tetum Language to teach in the classroom as official language but most of the lecturers are still using Indonesian language as main instruction in the classroom. We are looking for mechanisms to address those lecturers that are unfaithful to the rules. Our constitution demand us all to teach in Portuguese Language.</p>
<p>Q4. Is there any legislation that specifically deals with TPE?</p>	<p>A4. There is no legislation yet. If within this year we can establish CPAI (Comissão Permanente Avaliação Institucional). Then we will look into this important aspect related to legislation on TPE. This is also link to ANAAA.</p>
<p>Q5. Which regime that UNTL use to measure TPE? Do UNTL have their own format for TPE or still using Public Service Commission Format to evaluate lecturers?</p>	<p>A5. There is none yet. UNTL still using the general regime used by Public Service Commission where the indicators used for evaluation are not consistent and not suitable to evaluate lecturers because the indicators are too general and it may be suitable for administration staffs. The general director of academics have mentioned about that and we discuss but we are awaited for the establishment of CPAI. We need an international expert in the area of TPE to work with us to establish the CPAI and do the work. We have a small committee as part of ANAAA consist of three people with the task to look at the issue. Specifically there is no such TPE. There were several discussion taken place at HUMAN RESOURCE MANAGEMENT GORUP in ministerial lines but until today nothing has happened yet.</p>
<p>Q6. Why UNTL has already established long time ago but no such TPE taken place.</p>	<p>A6. I do not know. Why? There are laws but no implementation yet. I have talked to the Deans to establish CPAI and look after the lecturers at the Department level. I will suggest more at the Management Meeting.</p>
<p>Q7. Do you think teachers self-assessment is important?</p>	<p>A7. It is very important, because moving toward the institutional capacity building the lecturers should identify their own weaknesses and</p>

Questions	Answers
	strength and seek to obtain basic competences in order to better facilitate the teaching and learning process in the classroom
Q8. What key recommendations would you suggest to UNTL to better carry out the TPE in the future?	A8. I, as part of internal chief for control of quality, I always propose in management meeting to have TPE. If it is possible ANAAA help out in the establishment of CPAI, so that can tackle the TPE.

INTERVIEWEE 7. – Interviewed on 29th of June 2016 Time: 14:00 – 14:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	45
2. Professional Data	
2.1 Years of work as Teacher	11
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any TPE being conducted at this institution such as University?	A1. At university level there is no such none on the TPE. But it is only happen at the Department level. I tis the creativity among the lecturers at the department. One instance only the foreign lecturers such as Australia Ambassador send volunteers to our department and do such evaluation.
Q2. What do you think would be the obstacle of not implementing the TPE?	A2. I think for the English department is flexible. But I don't think the other departments could do it. I think the many other lecturers from other department could not really accept the TPE in the class. I don't believe that everyone will accept the evaluation in the class. They are resistant. They said, "who are you to evaluate". Then if the foreigners come and do TPE in the class they will not cooperate, maybe, I just guess. I believe they will not accept such evaluation, but not everyone. The lecturers many times do not cooperate about such evaluation.
Q3. What do you think is the best solution to address this issue of TPE?	Q3. There must be a policy from the higher level. Ministry of education, UNTL or stakeholders.
Q4. Is there any policy or rules taht could regulate this TPE?	A4. I think all are there, only still not implemented yet. I think the curriculum is good one, facilities are a bit okay now. The conditions now are better than in the past and is improving, though still limited, the renovation of infrastructure implemented. The laws are there but no implementation yet with regard to TPE.
Q5. Do you think the lecturers will follow those rules and regulations in the future?	A5. I think they will follow the law step by step. Because the pressure, the time of globalization, if UNTL have more cooperation's with other universities and want to be recognized then there is no choice everyone should perform and adjust with the condition.

Questions	Answers
Q6. Now Timor Leste is already a independent country and have its own law and the constitution demanded that the official language such as Portuguese should be taught at all levels of Public University as well as private University? Any progress?	A6. UNTL is only pubic university. Is Ok. Because most of the lecturer's study Portuguese and doing Masters and PhD and graduated from Portugal and Brazil so they can teach in Portuguese, but we still have many lecturers with Bachelor's Degree and Masters but they are still not afford to teach in Portuguese because most of them were trained in Indonesia.
7 Q7. Why is the key problem of the UNTL for not implementing this TPE even the many lecturers are resistant but the law is there?	A7. The conditions are still not in place, Though there is no evaluation in the class. We have many qualified teachers at UNTL and with high quality and bring about good results. Everyone is feeling proud that they are teaching well already.
Q8. Do you think that the lecturers have this basic skill in Pedagogical competence and other competences?	A8. Those who graduated with Bachelors, Masters and PhD in Education there is no problem in terms of pedagogical competences, teaching, doing syllabus and lesson planning but the other many lecturers have no skills in pedagogy and other competences such as professional competence and personality trait and social competences. They must have continuous formation in the area of pedagogical skills before teaching. One should have a diploma of teaching such as Certificate IV.
Q9. Are there any willingness to have a collaboration with other universities, institutions, stakeholders to provide continuous formation on those competences aforementioned?	A9. The stakeholders do not decide on this yet. If there is any decision then all lecturers should obtain Certificate IV before teaching. It depends upon the decision of higher level.
Q10. What is the key role of Ministry of Education in the issue of TPE? Is there any direct observation about TPE?	A10. It's a complicated issue. Ministry of Education have lack knowledge about the pedagogical skills and other competences to conduct TPE. Many staffs are nominated because of political appointment but they have lack of knowledge on the pedagogical issues but they may have skills on administration issues but not pedagogical issues, and psychology of education.
Q11. How about the curriculum?	A11. The change the curriculum every two years. And there is no evaluation on the curriculum. I am doing my study on the curriculum reform. I interviewed many important people or key leaders but nobody mentions about the evaluation on the curriculum.
Q12. Key recommendations to UNTL on Teacher Evaluation? What is key recommendations to UNTL to address the problem of lecturers without fundamental competences in teaching?	A12. Key recommendations to UNTL <ul style="list-style-type: none"> • UNTL should prepare and disseminate more information on the new curriculum of ECTS so that all the lecturers are well-prepared to implemented the curriculum. • There must be a close evaluation to all changes within the curriculum

Questions	Answers
	<ul style="list-style-type: none">• Policy is very fundamental to guide all implementation actions at any level within the university• They must have continuous formation on the area of pedagogical competences, professional competences, personality trait, and social competences, so that they will be able to have a better preparation in the preparation syllabus, lesson planning and didactic methodic on how to deliver a good teaching. Otherwise the students will become victims.

INTERVIEWEE 8. – Interviewed on 30th of June 2016 Time: 11:00 – 11:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	52
2. Professional Data	
2.1 Years of work as Teacher	16
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. In your experience as lecturer is there any instance that TPE have taken place in UNTL either at institutional or departmental level?	A1. At Departmental Tetum level, not yet. Because the Tetum Department is new one.
Q2. Is there any format where UNTL use to measure the pedagogical competences of teachers.	A2 It has happened already at all departmental level but it has happened institutionally, each lecturer where given form of evaluation such as teacher self-assessment. It was once in the year 2012.
Q3. Is there any indicators show that some of the lecturers are demanded to leave their job as teacher because of deficient performance?	A3. There is no sanction taken place however as lecturers after the evaluation started to recognized their own weaknesses to improve their performance, method and services in teaching. UNTL is thinking to provide continuous formation to lecturers in order to improve their quality service in relation to the teaching and learning process in the classroom.
Q4. Why UNTL has already been established long time ago, but the TPE seems not taken much consideration yet? Any obstacles?	A4. I have no answer to that.
Q5- Is there any legislation already exist to ensure the implementation of TPE?	A5. In the Regime Carrier Law has already identify clearly the duties and role in their daily work in the classroom.
Q6. What are key recommendations for UNTL to better implement the TPE?	A6. Lecturers must obey the rules and regulations. Those who go against the law, should take any concrete sanction against them. All are obliged to obey the law'
Q7. Do you think that continuous formation for the lecturers is important for improving their quality of service?	A7. We need a continuous formation for the lecturers. Mostly in the area of pedagogical competences, professional competences, social competences and personality trait. How lecturers could do lesson planning etcetera.

INTERVIEWEE 9. Interviewed on 30th of June 2016 Time: 14:00 – 14:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	43
2. Professional Data	
2.1 Years of work as Teacher	10
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any instance where UNTL already implemented TPE?	A1. First of all, I would like to explain in two ways. One way must be coming from the institution, and the other is from the lecturers themselves. I have not seen any TPE taken place. I myself evaluate myself the way I teach in the classroom.
Q2. What is the obstacle of not implementing TPE?	A2. As far as I know our Rector has set up an office of quality of assurance but I have not seen any TPE taken place. The obstacle is very related to the management plan of the office. I do not know how they carry out their plans. But currently, the new governance (New Rector) has already plan to have this TPE. Not only for lecturers but as well as for all the administration staffs, things related to practical work. And figuring out the percentages between theoretical work in the class and practical work outside of the classroom or even at the field or at laboratory work. This will be useful if it will be implemented. I am sure it will be implemented.
Q3. Is there any legislation already in place to carry this out?	A3. I am not aware of this, I have not read any document on this, but Magnificent rector come up with a dispatch and has appointed Mr. Angelo dos Santos to set up a Commission of Permanent institutional Evaluation to in charge of the evaluation issue in the classroom, look after of schedules for lecturers in the class.
Q4. I heard that some lecturers are resistant to the establishment of CPAI handled by Mr. Angelo? Direct observation.	A4. I think it depends on what kind of lecturers group. I can see we have few groups of lectures, first group are those who hold position responsibility at the university such Rector, pro Rector and Deans, other than teaching they have huge responsibility, and the second group is only who are teaching and doing research, and other group lecturers with other jobs outside. They

Questions	Answers
	would probably adjust their schedules to the hours that they come to teach so that they can better facilitate the learning process.
Q5. Is there any mechanisms to provide sanctions to those who do dual jobs either inside or outside job?	A5. I think that it is clear that UNTL statute is there, we have the code of conduct. Those are legal document to comply with. Otherwise there will be consequences. At this moment some of the lecturers are still doing jobs outside, they are still free to do jobs outside, I do not know why. for those who focus in working at University it is easy for them to be in class on time. There must be measures taken against them.
Q6. Do you think that the existence of Regime Career for the lecturers is sufficient to regulate the dynamic movements of lecturers at the university?	A6. I think it has a few categories and those categories are being there appropriately, some lecturers they are already at this level, but they are still in the other level, some of them at the other level but they are already on top. This needs to be adjusted, and reevaluated. Based on any legal document in relation to the Regime Carrier we need to count the contribution of chums, I do not know how far it is being implemented. Is already implemented or not. How they are going to implement. Some of the lecturers working from the morning until night. In my experience, I work the all week, I only have 2 and 3 hours to rest. I do not know how they would be accumulated into cum. Because it is related to salary payment every month, it must be slowly and surely followed by institution. otherwise if they will not implement and things will still the same.
Q7. Any key recommendations to UNTL as an institution to better carry out the TPE?	A7. I am thinking It would be much better, the Vice Dean of Academics affairs of each faculty work with Mr. Angelo from CPAI who is responsible for evaluation and they rearrange the program no how they are going to do the evaluation. They can do evaluation periodically, and evaluation with sudden appearance in the classroom. this is good because some lecturers for no reason they did not appear in the class and teaching. Is good for Mr. Angelo (CPAI) to see this thing. It is good for CPAI working closely as well with Vice Rector for Academics and vice dean for academic affairs they set up a system which is structurally easy to be followed. It is good because we will have a permanent team in charge of evaluation at UNTL.

Questions	Answers
Q8. Do you think that continuous formation or on going formation for the lecturers is important to have appropriate competences in teaching such as pedagogical skills?	A8. I can say two things. First, for students to be successful, we have to see the division of the percentages of the classroom and practical work. This lecturers from education faculty they have learn methodology of teaching, lesson planning, syllabus, and easy for them to facilitate students, all depend on the percentages between theoretical aspects in the class and practicalities in the field. For those who are not from education faculty with lack of knowledge in pedagogical skills should go through a certain training on the teaching methodology to enhance their skills to better provide service in teaching.
Q9. How far the code of conduct has been conducted at UNTL.	A9. It has been implemented in the last year governance. We have disciplinary council will be in charge of code of conduct.
Q10. Is pedagogical council and scientific council already establish at the faculty level?	A10. It has established.
Q11. Do you know what is the role of ANAAA?	Q11. ANAAA will be in charge of the accreditation of Universities in the country.

INTERVIEWEE 10. – Interviewed on 30th of June 2016 Time: 16 :00 – 16:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Female
1.2 Age	66
2. Professional Data	
2.1 Years of work as Teacher	10
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Backgorund	PhD/Doctorate
3.2 Others academic backgorund	None
Questions	
Answers	
Q1. Is there any instance where TPE is implemented at UNTL?	A1. I am not aware of. But I only know is that there is a teacher self-assessment where each lecturer knows their weaknesses but not evaluation toward lecturers from the institution or from the students. It is good to have a TPE. Ultimately, I am not aware of TPE. I tis important to provide a evaluation format at the end of each school year and give out to the students so that the students would have feedback to lecturers in order to find out whether or not lecturers are performing well or not so that could provide input and venues for the improvement of teaching quality.
Q2. What are the gaps or obstacles for not implementing TPE?	A2. In my perspective is that we lack of human resources expert in the area of evaluation.
Q3. Do all the lecturers have basic skills on teaching competences?	A3. Not all. Some must go through training to obtain pedagogical skills so that could better perform in the class management.
Q4. Why University has already been established for many years why no such TPE taken place?	A4. I have no idea why?
Q5. What are the key recommendations that you would suggest to UNTL to better implement TPE?	Q5. There must be a commission with specific task to carry out this type of evaluation. Lecturers must perform more. Not only teaching but also do other things such research and writing up scientific articles and other activities.

INTERVIEWEE 11.– Interviewed on 4th of July 2016, Time:14:00 – 14:45 Venue: UNTL

1. Personal Data		
1.1 Gender	Male	
1.2 Age	42	
2. Professional Data		
2.1 Years of work as Teacher	2	
2.2 Experience of class observation prior to the process of performance evaluation	None	
3. Academic Formation		
3.1 Educational Background	Masters	
3.2 Others academic background	None	
Questions		
Q1. Is there any implementation of Teacher Performance Evaluation at UNTL	Answers	
	<p>Q1. Is there any Teacher Performance Evaluation implementation (TPE) in the UNTL</p> <p>A1 Teacher performance Evaluation is the biggest challenge we have is compliance with the law, besides the law we have to make multiple inquiries. Each teacher will be evaluated through 10 surveys. Including yourself, therefore being an organization with more than 400 teachers, so we are talking about at least 4000 inquiries, implement this on paper and analyze all this is not easy. Our first gap is the lack of a system, an online internet platform that allows students and teachers to be connected and an effective and immediate implementation of the surveys. That would be the best way. We have college engineering through the IT department that will probably implement the surveys already in college engineering through an academic information system, I hope it will be implemented also in the remaining colleges of UNTL. Moreover TPE according to the law is carried out through two systems. Each teacher is evaluated in two ways. A first that calls evaluation of perception, in this evaluation of perception the teacher is evaluated say that empirically there are no grids, there is no rigor, say students fill out questionnaire through an idea they have about 5 students. There are also 3 fellow peers of the teacher, who will also evaluate, a boss, department director and the teacher himself. The teacher will be assessed on a scale of 1 to 5. You must have a minimum grade of 3.5 to pass. The other form or</p>	

Questions	Answers
	<p>modality of evaluation, the second, is through the so-called Personal Description Tool that has as base the three looting of higher education that is teaching, research and service to the community. Basically the teacher will make the personal description of his activities in the Ministry's own form where he will indicate things like his creative effort, therefore to improve quality of learning, and impact for change, let's mention scientific works, work goals for science development , Institutional support, impact on change, and support for society, so let's say a number of issues that teachers will be evaluating. After this personal discussion, the teacher will self-classify himself, the evaluation will also be made by the hierarchical superior probably the director of the department also who will analyze this personal description of the teacher and will assign within a grid, therefore a classification. And then this report should have a minimum score of 3 values, to be approved on a scale of 1 to 5. At the end of it all, we will, say, evaluate the consistency of the two evaluations and compare these two assessments and have reached a final result. Therefore and there will be TPE. On the board after teacher certification. Therefore, TPE presupposes assigning the number of credits after which these points will contribute to the certification of the teacher for the purposes of his career profession.</p>
<p>Q2. Is there any existing legislation that regulates this Teacher Performance Evaluation?</p>	<p>A2. Yes, there is legislation Decree Law No 13/2014 of January 15. Subsequent to this decree law issued a dispatch that approves a new certification of the university professor, it instructs the law that each university has a cabinet, a cabinet of certification of the university professor both the GAB CEDU (Office of University Teaching Certification). I can mention that at this moment we have a small technical problem in the manual and probably this will be the discussion that and after evaluating the two modalities when it is going to be verified the consistency of the two evaluations we have here a technical error in the mention of approvals and reprobation's</p>

Questions	Answers
	that it will have to be fixed. We have at this time the problem that there will be some person in the Ministry with the technical capacity to legislate and correct this technical error.
Q3. Is there any direct observation either directly or indirectly from the side or part of the Ministry of Education to measure the skills of teachers namely pedagogical competences, professional competence, personal competence, and social competence applied in the classroom?	A3. No, it is not planned that the Ministry of Education, an external team to make evaluation to teachers. In the university, according to the law we have to have an internal regulation basically to have the institutionally approved manual, therefore the part of the TPE that appears in the CEDU Manual. Is there no direct evaluation of the performance of teachers in class, which makes it difficult for the department director to assess whether or not the faculty member has quality? The only ones who can measure services are the students empirically in a perception assessment. In technical and strict terms this is not foreseen.
Q4. What are the fundamental reasons that does not permit to implemented the TPE because the University has already been established 15 years ago?	A4. Well, that was the decision of the Ministry of Education. In my personal understanding, we are not yet in the implementation phase, and busy with many issues and we will still draft the internal regulations for TPE. Based on this, we can add some aspects that contribute to this, which will be approved in the academic senate, in the management council and in the general council. I will just put this into consideration, because we should have in each department the director, academic assistant director to have the opportunity to go to the classroom and according to an own grid implemented institutionally, say evaluate the compliance of the teacher in their activities. It is the only aspect that may be difficult to gauge. All other scientific production, service to the community have done the quality of work can also be evaluated, we have the Research Center, we can consider the research center. Now at the level of school activities we have no mechanism to allow someone to go to the room to assess the TPE. I will put this as a consideration whether it should be done or not, and then at the institutional or university level.
Q5. I heard that some teachers were very much resistant relatedly to TPE, the empirical questionnaire that has been distributed to	A5. We implemented before the implementation of questionnaires defined in the Manual, the office of the Academic Vice-

Questions	Answers
<p>teachers somekind of na evaluation that comes from the cabinet of Vice Rector of Academics to fill up?</p>	<p>Chancellor first implemented in the UNTL kind of pilot project only for the first year of the courses, assessing the students' perception regarding TPE, not only TPE, But also in relation to the curricular unit, and the faculty and institution in general, was a more general survey, it was a more comprehensive assessment. A report was prepared on this and presented, the idea was to evaluate every year again we are confronted with technical difficulties because the university does not yet have adequate information system, so basically, we want to implement a system of contemporary quality, but without existing Conditions. Therefore, the university is feeling that it will need to register all teachers, students, courses, use data in the information system that allows students to evaluate teachers and implement the surveys in a much simpler, practical way, through the system Information and we have the result on time.</p>
<p>Q6. What was the results of the pilot questionnaire about TPE which implemented by UNTL and given-out to the teachers?</p>	<p>A6. We did two types of evaluation, one by closed questionnaire, I said that I evaluated the aspects of the TPE, the validity of the course of the university and university and then implemented a questionnaire was directed to the teachers, to give examples of activities at the pedagogical level, at the level Students, we are talking less experienced, teachers who only teach in the first year of UNTL courses, teachers who are not part of the staff, teachers who are Contractors, teachers who are licensed, about 2 or 3 years ago, teachers with little experience. This evaluation made to the teachers identified 40 aspects that teachers should improve as a consequence of the organization of a workshop that was presented to the faculty, department directors and deans of the faculty where all the aspects that should improve were addressed, such as suggestions and then made an approach on the need to implement greater rigor in teaching quality and assessment. The most critical point that teachers increasingly fail in their assessment of students.</p>
<p>Q7. What is your observation that there are teachers that lack of experience in the</p>	<p>A7. In here we have advanced training center for learning teaching, have been implemented</p>

Questions	Answers
<p>pedagogical area, but have already worked as teacher and now teach in the classroom already? Is continuous formation become an important problem?</p>	<p>teacher inquiries identify what kind of course they want to attend, and want to continue their training, younger teachers is encouraged to improve ensure their quality, and provide opportunities Teachers to improve their skills.</p>
<p>Q8. What are your key recommendations to UNTL in the implementation of TPE in the future?</p>	<p>First factor is a strong entity with full support from the Rectorate that is concerned with academic quality, so a team should be a committee eventually, then with executive director who deals with quality aspects related to academic quality, include the TPE, and this is an organizational aspect. In the more technical aspect we need a contemporary information system to meet the quality standard that is established by law. We must be rigorous to have a good regulation that allows us to evaluate the quality and readiness of how teachers do, how they prepare the syllables, the quality of their tests, the tests and evaluation, the quality of their classes, besides what is in the law, And because, because this we would give the possibility to fill small obvious technical flaws, give the opportunity through the center advanced training for learning teaching to improve their pedagogical and didactic skills so that they can better perform their teaching activities. Teachers do not only teach, do research and also serve the community. At the level of research, we have to have patience, because at the moment UNTL is a growing institution, with human resources still very young, we are not very rich in doctorate professors, to do research, the quality of an investigation depends on the quality of teacher training, Teachers with doctorates naturally have higher quality in research compared to teachers with masters. You must have patience to admit it with time, and of course it also includes service to the academic. In any case I know that the university next year will have all the investigations centralized in the SNIC. What is going to happen is that teachers will present their proposals in order to start their research in the chosen area. Thank you very much.</p>

INTERVIEWEE 12.– Interviewed on 4th of July 2016 Time: 16:00 – 16:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	65
2. Professional Data	
2.1 Years of work as Teacher	5
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. What is your opinion and experience as the Dean of Faculty of Law would like to find out whether or not any TPE taken place in your University and using specific evaluation format to measure the teachers competences such as pedagogical, professional, personality and social competences in the classroom?	A1. There is no TPE because there is no rules, regulation that is being disseminated to us about the Teacher Performance Evaluation. There were no mechanism created by the institution. We have no structure to measure the TPE. One of the reason is also because we have lack of resources in the área of law and expertise in the área of performance evaluation we sometimes think to hire expertise from outside to share their experiences. Up to the present there is no mechanisms in place to ensure the teacher performance evaluation.
Q2. Is there any thought where University would bring experts from other countries in the area of TPE to share their experiences with teachers at UNTL?	A2. I think that is a good idea because we are relative an new institution
Q3. Key recommendations for UNTL to improve the TPE?	A3. Invite experts in the área to share with us on how to do TPE.

INTERVIEWEE 13. – Interviewed on 4th of July 2016 Time: 16:45 – 17:30 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	47
2. Professional Data	
2.1 Years of work as Teacher	15
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. Any instance where TPE (Teacher Performance Evaluation) implemented at UNTL?	A1. TPE whether feel or not it is starting to develop, however, its processes and its mechanisms is not well-managed or controlled. There was an instance where the institution set up a type of machine to regulate the teacher's punctuality and presence at the institution mostly be in the classroom on time, and I think that is one of the instrument to evaluate the TPE, however, it is not that effective. For example; the system opens up 8 am to 8.30 and I, one day I came early at 8 am and want to press the monitoring machine for my punctuality and my presence, however, the machine itself did not function well. And the monitoring system established within institution is not being followed by teachers and no sanction has been taken against them. The monitoring machine is only at the center of the University and it does not set in other faculties within the University the same type of monitoring machine set in their own faculties, so it raises a question where the monitoring machine is only apply to those teachers at the center of University or also other teachers in other faculties.
Q2. Is there any rules to regulate the punctuality of teachers in coming to class? Or not coming to class?	A2. Truly, when we talk about rules we talk of black white evidence such as written document. I am not aware that the institution provide any dispatch containing instructions such as when one of the teacher violate the rules, or does not sign the list of absence will have a deduction of salary of teachers or others, but this does not take place. There are rules but many of the teachers are still violating

Questions	Answers
	the rules and there is no sanction being taken against the teachers.
Q3- What are the key recommendations that you want to suggest to UNTL (Universidade Nacional Timor Lorosae) and to yourself to better implemented the TPE in the future?	Q3- We really need to have the TPE taken place and must be based on the official dispatch of the institution mostly coming from the upper level management. And hope that all the teachers should be aware of all the proceedings related to the Teacher Performance Evaluation is already stipulate in the Special Career Regime where all the teachers are obliged to follow it because it does affect their career and the promotion or dismissal from the job as a teacher. The TPE must be conducted by an expert in the area in order to avoid misleading in the implementation of the TPE.
Q4. Do you think continuous formation in the area of Teacher Performance Evaluation and training on the pedagogical skills, professional skills, personality and social competences is crucial for all the teachers at UNTL?	A4. Yes exactly, training is very much important because it helps to upgrade the skills of the teachers in the aforementioned area, and so that the teachers will be able to ensure the quality assurance and promote the professional learning and taking into consideration that the students are target beneficiaries of the teaching and learning.

INTERVIEWEE 14. – Interviewed on 13th of July 2016 Time: 16:00 – 16:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	47
2. Professional Data	
2.1 Years of work as Teacher	6
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	PhD/Doctorate
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any TPE taken place at UNTL level?	A1. To the best of my knowledge there is no such TPE taken place at UNTL.
Q2. What are the main reason that this TPE has not been done?	A2. I cannot comment on the years that I did not joined UNTL before. Since 2011 up to date I joined UNTL, I feel the need that TPE is really important and we need to do annual review on lecturers because it helps the teachers to improve their quality of service in the area of teaching and learning.
Q3. I heard that at the University there are different types of lecturers teachers such as those who are with educational science background and the others with pure science background and with lack of pedagogical skills and other competences? What is your observation?	A3. I cannot comment on the other faculties, but certainly, I can comment on my Faculty of Medicine, actually we have pedagogical training in collaboration with Gulbenkian to the lecturers where took place in the 2012 for two weeks, containing how to organize the class, how to teach better, how to do a good evaluation on students, and now we have another second training conducted through the same organization Gulbenkian on pedagogical skills and professional skills with the purpose to upgrade the teaching skills of teachers to provide a better service in teaching and learning. Target is mostly targeting for the newly recruited teachers.
Q4. Do you think that this TPE is really important?	A4. Is really very important. The point you made earlier is that those teachers with educational science background and they have pedagogical skills but many faculties who are not coming from educational science like us, we have lack of pedagogical skills. We are professional but we don't have pedagogical skills and the TPE is really imperative. Therefore it is important to have a special training on the pedagogical skills so that we as

Questions	Answers
	teachers can provide a good quality of teaching and guarantee the quality of teaching. Making sure that the students getting what they supposed to getting.
Q5. I heard that, when the university is in the process of establishing itself, annually they are using the performance evaluation that were administered by the public service commission? Any idea?	A5. I know the form of Public Service Commission is applied to normal public servants and public administration staffs but it is not suitable for the lecturers or teachers. But those who are working as teachers or as lecturers at the faculties or department there have been no Teacher Performance Evaluation taken place until today. We are only talking but the implementation is not there. None at all.
Q6. What is the gap that you think of why the UNTL until now did not implement the TPE?	A6. I do not know why UNTL did implement this yet. I am not aware of. I think this must be upper level management decision especially the Rectorate. The administration structure have to know this. They should instituted very soon.
Q7. Do you think that some teachers would be resistant to accept this concept of teacher performance evaluation by the institution?	A7. It might be individual resistant but from organizational view point and academic institution TPE is for all teachers or lecturers and they are obliged to comply and follow it.
Q8. Is there any legislation that is already in place to regulate this issue of teacher performance evaluation?	A8. Well. The career regime is requiring and obliging us teachers to really show a good performance for example teachers are obliged to come up with certain publication of articles and other academic demands, and that are all for the promotion and career development. We are obliged to do a regular performance.
Q9. What is your key recommendation to UNTL in relation to the implementation of TPE?	A9. UNTL should be as soon as possible to set up a committee to be in charge to of Teacher Performance Evaluation to all lecturers working at UNTL. If the UNTL want to achieve excellence then the TPE is really important and imperative
Q10. In order to ensure the professional development of teachers, is there any training plan at your faculty either it is yearly or biannually?	A10. Through this collaboration with the Gulbenkian we have this pedagogical training although it takes a bit long time for the second training, and the training is going on and targeted for those teachers who have not attending the pedagogical training and also targeted recruited teachers. Perhaps after this then we will take next step.
Q11. In terms of facilities that would facilitate the lecturers to better provide a good quality of teaching, what is your observation?	A11. In terms of facilities we are improving, some classes already have projector, we still struggling about the classrooms, I mean the

Questions	Answers
	quantity or the management of the classroom itself that needs to be carefully assessed.
Q12. In terms of the ratio of the students for each class. What is the range?	A12. There is a lot. Some people or teachers have an ideal class. It ranges from 1/40, but some others the ratio is 1/70 or 1/80. We need more space or we need a better management of classroom we already have and don't let some class empty because we still see some class are still empty.
Q13. In terms of Laboratory any observation?	A13. That is the area that we have not improved. We have a lot of discussions but nothing has happened yet.
Q14. In terms of Library is there sufficient books avail for teachers and students to have access?	A14. Library is sufficient at the moment and we need more staff to manage the library.
Q15. In terms of Internet access?	A15. Sometimes in on and off. Because the size of wireless is not that large, the signal is not good and sometimes is difficult to access to journal and articles.
Q16. Any complaints about the punctuality of teachers in the classroom? What is your observation?	A16. There is no close supervision therefore some teachers are still not complying with it. There are a lot teachers no coming on time.
Q17. Do you have experience any direct or indirect observation to teachers teaching in the class?	A17. From internally we should have. I think at the UNTL level already assigned someone to take care of this monitoring and supervision in the class. There are already mechanisms of supervision and teachers were cautious about that.
Q18. Is there any code of conduct to regulate all teachers.	A18. I think there is. By all account all the teachers are public servants, they all need to be present in the campus 8 hours per day.
Q19. What are your means of proving that such a teacher is performing or not? Any indicator whether or not the teacher is performing?	A19. That is only through evaluation. From the Dean of the Department but also evaluation coming from the students on the competences of teacher.
Q20. Do you think that the pedagogical training is important to all teachers?	A20. Yes, Is very important and imperative.

INTERVIEWEE 15. – Interviewed on 19th of July 2016 Time: 10:00 – 10:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Male
1.2 Age	47
2. Professional Data	
2.1 Years of work as Teacher	16
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	PhD/Doctorate
3.2 Others academic background	None
Questions	
Q1. What is your experience on how TPE (Teacher Performance Evaluation) implemented at UNTL (Universidade Nacional de Timor Lorosae)?	A1. The performance evaluation of teachers and public servant's staffs at UNTL is based on the standard form of public service commission of Timor Leste, since all the teachers are public servants therefore all of teachers follow the criterion of Public service commission of Timor Leste. There are several points underlined in the format such as on punctuality, honesty, it is about 8 points, however, we have never established TPE for our teachers at UNTL, only by now we think to assign teacher Mr. Miguel Maia and teacher Angelo dos santos to be in charge to establish TPE of UNTL and semester performance evaluation.
Q2. What are the gaps or problems that the TPE is only start by now?	A2. We suppose must already have the TPE since the institution established in the very beginning, however, the previous governance of this institution did not provide venues to do such establishment of TPE. That is the reason TPE is instrument for improvement of the conditions that is being evaluated. For example a teacher progress is good or not will only be known through TPE, if the teacher performance is not good, come into class without punctuality, do not teach using teaching methodology, have no lesson planning, no syllabus, how to deal with failed students, we construct SOP which is specifics or appropriate for the teachers and also appropriate for Public Service Commission. Punctuality, relevance of the subject, evaluate whether the method is student centered or teacher centered methodology, how to evaluate the students' teachers provide task to the students, home work for teachers did the

Questions	Answers
	teacher evaluate or not. During this time we use the performance evaluation based on the commission of public service. Where the public service commission performance evaluation form is not suitable to teachers.
Q3. During the establishment of the institution did you observe those teachers who teaches at UNTL do they obtained pedagogical skills and other competences before teaching at UNTL?	A3. We know that those teachers who have no background with pedagogical skills that means they have no pedagogical skills background and then they have to go through a pedagogical training for teachers. For those teachers with educational science background I think they automatically have sufficient skills in the area of pedagogical skills and the other teachers with non-educational science background no doubt they have scientific skills but we have to find out either or not they have acquired the pedagogical competences or not. I don't like to be skeptic whether or not they have pedagogical skills, I think we need to do a data collection and find out either or not they have pedagogical skills. If not then they still need to go through a pedagogical training or continuous formation to improve their ability in provide better quality of services especially in teaching and learning.
Q4. Do you think the evaluation form from Public Service Commission is appropriate for the teachers at UNTL?	A4. The public service commission evaluation form is not really appropriate to measure the teachers performance. It is too general.
Q5. Is the code of conduct of UNTL can also be used as measure to measure teacher's performance to identify whether or not we have to keep the good teachers and dismiss the bad teachers?	A5. We have to formulate the evaluation form and do a survey to our teachers, where we can evaluate the teachers, and from there we better not to be skeptic to dismiss the bad teachers, we have to find ways and if the diagnostic test found out that the teacher still has deficiency in the performance then we try to provide mechanisms for the teacher to go through a training and then after the training we will try to see whether or not the teacher is already improving in their performance or not. Normally after the training there will be changed because each teacher has their own internal Dynamics that would contribute to change.
Q6. If after the training there is no change at all in the part of teacher in their performance what measures would be taken?	A6. If after the training they are still not performing better, then the institution should decide upon either provide further training or decide if the teacher will be asked to do other general job other than specific job such as teaching. At UNTL we have one section at the

Questions	Answers
	<p>university the so called Pro Rector for Internal Quality Control that will also help in doing monitoring the performance of our teachers.</p>
<p>Q7. Is there any legislation that deals with the TPE?</p>	<p>A7. There is a special career regime to go up from one scale to another scale then there must be a mechanism to evaluate the performance of the teachers and find out either or not the teachers have come up with scientific articles and other academic research because that will be used as the bases for measuring performance and for career development and promotion. We must also have the internal bylaws that can regulate the performance of teachers.</p>
<p>Q8. I heard that some lecturers or teachers are resistant to TPE during the TPE survey 2014? What is your perspective?</p>	<p>A8. In our special carrier regime, there is no teacher would escape from the TPE, there is no teacher that escape from the observation, teacher have no right to be resistant. If the teacher is not at ease with the rule of law that means that he or she do not want to become a teacher. TPE does not apply the concept of either one is senior or junior.</p>
<p>Q9. What is your key recommendation to better implement the TPE?</p>	<p>A9. Construct an evaluation form that is appropriate to measure the performance of teachers based on the indicators such as pedagogical skills, professional skills, personality and social competences. Most specifically, their lesson planning, syllabus, methods of teaching, methods of evaluation and etc.</p> <p>Teacher is an eternal student. Should always learn until death</p>

INTERVIEWEE 16- Interviewed on 19th of July 2016 Time: 14:00 – 14:45 Venue: UNTL

1. Personal Data	
1.1 Gender	Female
1.2 Age	35
2. Professional Data	
2.1 Years of work as Teacher	5
2.2 Experience of class observation prior to the process of performance evaluation	None
3. Academic Formation	
3.1 Educational Background	Masters
3.2 Others academic background	None
Questions	
Answers	
Q1. Is there any instance where TPE conducted at UNTL?	A1. There is none. And even if there is it will be difficult.
Q2. Why, and what is the the reason?	A2. Firstly, because of the language which is Portuguese, and the teachers will probably not understand the questions in Portuguese regarding the questionnaires. Then is the process which is more based on ideas, opinions of what we have on the teachers and is not based on the evidences, so, that is the big negative point.
Q3. What was the point of Ministry of Education doing such instruments or criterion on TPE without consulting a public higher institution like UNTL.	A3. Well, they did this presentation, but then the Minister of Education changed, the people in charge of the process they changed. And they just kept what had been done before and they did not do any follow up, so, they just presented, so, here is the final result. I think because they did it because based on the other countries experience, maybe the Philippines or Indonesia, is based on their evaluation performance assessment. So they assumed that is probably is enough, they did not consult anybody at UNTL. And I think private Universities the same.
Q4. Is there instrument designed at UNTL to measure the TPE?	A4. I think at UNTL there is no such measuring instrument, I believe teachers are evaluated by Public Service Commission evaluation measures.
Q5. Do you think that the content of the measurement of Public Service Commission is suitable to evaluate teachers?	A5. No. Is not suitable for teachers. An instrument and a frame of assessment needs to be designed specifically for teachers, but the one that has been designed by the Ministry of Education and that we are going to implement, I think it lacks a lot of adjustments.

Questions	Answers
Q6. Are you referring to those competences measures instrument?	A6. The competences and then, even the questionnaires. The questionnaires is the same for all parts, we have students, we have peers, we have chiefs, and the questionnaire is the same, so, how can students evaluate teacher with the same questions as peers, or colleagues evaluate. So, this has not been thought of.
Q7. Do you think that those evaluation questionnaire instrument designed by Ministry of Education to measure TPE, do you think the content itself is ok. There are 4 (four) points covering the pedagogical competence, professional competence and others?	A7. Yes, I think those areas are OK.. But I think because students they are not used to evaluating, they have never done this before at high school, and then we they get here I think they will evaluate the teachers as in 4 and 5 which is good and very good. We have had that experience, we asked the students to evaluate UNTL and teachers as well and Unidades Curriculares Transversais, and it was all good and very good, they think the libraries are very good, they think classrooms are very good, so they are not used to evaluating so I think they will evaluate teachers as putting it all good and very good.
Q8. What was the major issue in carrying out that pilot study in relation to the teacher performance evaluation in your team?	A8. Yes, there was a report. It was me. Prof, Joao Cancio and Armindo 's Team, and we have a final report. As I have previously said. It was the average which good and very good but we need to revisit the questionnaires, and it has to be suitable to teachers context.
Q9. What are your key recommendations in relation to the better implementation of TPE in the future at UNTL? Any suggestions?	A9. Firstly, the Ministry of Education needs to be more side by side with universities, to implemented the system, evaluating is not an easy process especially when peers have to evaluating each other, the teacher will be evaluating each other and that might cause not very good environment of work. For example: I have a Master's Degree and you have a PhD and I am going to evaluate you, so, and you might think why are you evaluating me, you only have Masters and I am a PhD Degree holder, this has not been clarified by Ministry of Education so who is evaluating who? we don't know. We need more guidelines from Ministry of Education. For example who is going to evaluate the pro Rector we don't know, who is going to evaluate Rector we don ' t know. Do qualifications count for this? We don't know. So, the Ministry of Education needs to help

Questions	Answers
	<p>universities to clarify the procedures, otherwise, I think the end of this, the result would be all teachers will have good and very good. Another thing, there should be numerous clauses as in for very good, otherwise everybody will have very good. That is how they are going to evaluate. And then we don't have evidences, so the Director of Department is going to evaluate 10 teachers, how does he know what they are doing? Is very difficult for him. He has to evaluate them as the head of the department in the classroom activities. But in fact, The Director of Department does not have to go into the classroom and observe the teacher teaching, so, how he can evaluate the teacher if he does not see him teaching, is all based on ideas. How can you evaluate on pedagogical competences, if you have never seen teacher teaching in the class? There should observation in the class. UNTL task to develop those evidences, or indicators. If you say that this teacher is very good in the class, how can you prove that, if you have never seen that person teaching. Otherwise all teachers will be evaluating each other with good and very good.</p>
<p>Q10. Is there any rules and regulations regulating the TPE? For example Career Regime, what is your observation?</p>	<p>A10. Career Regime will have an impact from the result of the evaluation if you have a score of very good then you can proceed, if you have good then you can also proceed. But if you have not very good then you cannot proceed. Nobody is going to have bad or not very good.</p>
<p>Q11. What is your observation on the competences of teacher in the area of scientific research?</p>	<p>A11. That is one of the dimensions of the scientific research. Not everybody here is ready to be a scientific researcher, so I wonder how they are going to be evaluated. Nobody does research here. Only those that are working with SNIC (Center for Investigation) because they consider monographic as research, but that is not research, nobody is doing research, nobody is writing articles, web siting them, very few teachers are doing that in fact. I wonder how they are going to be evaluated that dimension. That is going to be a challenge for UNTL.</p>
<p>Q12. Do you think there would be any reluctance or resistance from the teachers?</p>	<p>A12. Teachers will question why am assigned here, the answer to that question because you</p>

Questions	Answers
	are not doing research. And they will say I don't do the research because we don't have sufficient library and laboratory to do research, how come I can conduct research. That is really true that we don't have sufficient library and laboratory. Libraries are very poor, We don't have scientific magazines, no access to scientific magazines
Q13. Do you think the establishment of information technology database can bridge the relationship where teachers and students so that they have regular communication in sharing knowledge?	A13. It will facilitate the process of teaching and learning easier. Teachers here don't usually accessing their emails, very difficult, is very cultural when you send email somebody and they will not answer, and not effective. Only few teachers access their emails.
Q14. What type of legislation that being established by the Ministry of Education on TPE?	A14. There is not specifically, not that am aware of. We have Career Regime Progression, We have accreditation process in here working apart. We should be working together because when we evaluate the programs then we need to evaluate the teacher as well. It has to be doing separately. For example, Accreditation section is evaluating the programs, they have teachers section and they are evaluating teachers on the number of subjects teacher teach but in fact legislation only requires teacher to teach from 6 to 12 hours, it does not have a specific number of subjects. It is about numbers of teaching hours not numbers of subjects. But the accreditation section is asking about how many subjects a teacher teach but that is not legislated.
Q15. Is there any structure being set up to do TPE?	A15. There is no structure. Who is going to evaluate a Rector and he has a PhD and at the same time work as teacher. Do the teacher with PhD will accept an evaluation done by the head pf department who only holds Master's Degree. Will the PhD teachers will accept the evaluation, maybe yes or no.
Q16. At the moment this practice happens?	A16 It happens because we don't have any guidelines. We will have people who holds Bachelor's Degree and Master's Degree evaluating PhD Degree holder. I think the PhD teachers will not be very happy. I think the Ministry of Education should come to universities and should talk a little about the impact of the evaluation on the career. Teachers has been evaluated but there is no

Questions	Answers
	such impact, they need support from Ministry of Education on how I am going to evaluate my colleagues.
Q17. Do you think the continuous formation for the evaluators and the teachers is on how to do TPE is crucial?	A17. Yes. They need support to understand how to do a good evaluation, need support, guidelines and trainings